



Artificial Intelligence (AI) and the Criminal Justice System

AI to predict crimes and as a criminal intelligence tool

ONLINE CONFERENCE, 5-6 November 2020 (CET)

**UP
GRADE**
YOUR LEGAL
EXPERTISE

**Criminal
Law**



Speakers (invited) and Chairs

Sophia Adams Bhatti, Head of Strategy and Policy, Wavelength; former Director of Legal and Regulatory Policy, The Law Society, London

Laviero Buono, Head of Section, European Criminal Law, ERA, Trier

Francien Dechesne, Assistant Professor, Center for Law and Digital Technologies (eLaw), Leiden University

Dick Heimans, Deputy Head of Unit, General Criminal Law, DG Justice, European Commission, Brussels

Robert Kroplewski, Member of the European Commission's High-Level Expert Group on Artificial Intelligence, Brussels; Solicitor; Plenipotentiary of Minister of Digital Affairs for Information Society, AIDA NASK, Warsaw

Cornelia Kutterer, Senior Director, EU Government Affairs, AI, Privacy and Digital Policies, Microsoft, Brussels

James Luke, Technical Executive, Global Business Services, IBM, Winchester

Anjali Mazumder, AI, Justice and Human Rights Theme Lead, The Alan Turing Institute, London

Sigrit Siht, Director of Data Policy, Ministry of Economic Affairs and Communications, Tallinn

Philipp Slusallek, Member of the European Commission's High-Level Expert Group on Artificial Intelligence, Brussels; Scientific Director, German Research Centre for Artificial Intelligence (DFKI), Saarbrücken

Matthias Spielkamp, Co-Founder and Executive Director, Algorithmwatch, Berlin

Simone van der Hof, Professor, Law and Information Society, Leiden University

Ian Walden, Professor, Head of the Institute of Computer and Communications Law, Queen Mary University of London

Key topics

- AI and its implications for the criminal justice system
- Using AI in criminal intelligence
- AI to predict crimes
- The internet industry approach to AI
- AI, ethics and privacy

Language

English

Event number

320SD61

Organisers

ERA (Laviero Buono) in cooperation with Queen Mary University of London



With the support of the Erasmus+
programme of the European Union

Artificial Intelligence (AI) and the Criminal Justice System

Thursday, 5 November 2020

08:30 Connecting to the videoconference platform

09:00 **Welcome and introduction**
Laviero Buono

I. ARTIFICIAL INTELLIGENCE AND ITS IMPLICATIONS FOR THE CRIMINAL JUSTICE SYSTEM

Chair: Laviero Buono

09:15 **Introducing the technical aspects of artificial intelligence and its implications for privacy-preserving big data technologies**
Francien Dechesne

10:00 Discussion

10:15 **Putting forward a European approach to artificial intelligence and robotics: work carried out by the European Commission**
Dick Heimans

11:00 Discussion

11:15 Short break

Chair: Ian Walden

11:30 **Algorithms in justice and justice in algorithms: are there legal decisions that can really be delegated to algorithms?**
Anjali Mazumder

12:15 Discussion

12:30 Break

II. USING ARTIFICIAL INTELLIGENCE TO PREDICT CRIMES: FORECASTING CRIMINAL BEHAVIOUR AND FUTURE DANGEROUSNESS

Chair: Laviero Buono

13:00 **Automated decision-making for law enforcement purposes: risk of rubber-stamping and other concerns**
Sophia Adams Bhatti

13:45 Discussion

Chair: Laviero Buono

14:00 **A unified approach for a wider implementation of AI: work carried out in Estonia**
Sigrit Siht

14:45 Discussion

15:00 Short break

15:15 **"Sweetie 2.0": an automated chat function to track, identify and deter individuals using the internet to sexually abuse children**
Simone van der Hof

16:00 Discussion

16:15 End of first online conference day

Objective

The rapid progress of artificial intelligence (AI) and machine-learning raises key questions regarding benefits and risks for the criminal justice system. This online conference will focus on how AI could be used to predict crimes and as a criminal intelligence tool, enabling the shift from a traditionally reactionary approach to crime-fighting to one that is more proactive and preventative in character. It is aimed at judges, lawyers, prosecutors, police investigators and civil servants with a view to highlighting the aspects of artificial intelligence already in use today as well as the changes and opportunities on the horizon.

Who should attend?

Defence lawyers, in-house counsel, judges, prosecutors, law enforcement authorities, civil servants and policymakers active in the field of criminal law.

Interactive online conference

The online conference will be hosted on ERA's own online training platform. You will be able to interact immediately and directly with our top-quality speakers and other participants. We will make the most of the technical tools available to deliver an intensive, interactive training experience. As the platform is hosted on our own server, the highest security settings will be applied to ensure that you can participate safely in a high-quality online training environment.

Your contact persons



Laviero Buono
Head of Section
E-Mail: LBuono@era.int



Liz Greenwood
Assistant
Tel.: +49(0)651 9 37 37 322
E-Mail: egreenwood@era.int

CPD

The Academy of European Law's conferences and seminars meet the standard requirements for Continuing Professional Development (CPD) recognition. Participation in this event corresponds to **9.5 CPD hours**.

Friday, 6 November 2020

III. ARTIFICIAL INTELLIGENCE AND THE INTERNET INDUSTRY'S APPROACH TO SERVE THE CRIMINAL JUSTICE SYSTEM BETTER

Chair: Ian Walden

- 08:45 Connecting to the videoconference platform
- 09:00 **Responsible AI: developing codes of conduct in designing and using algorithms to help law enforcers and judicial authorities better**
- *Cornelia Kutterer, Microsoft*
 - *James Luke, IBM*
- 09:45 Discussion
- 10:00 **Evaluating and shedding light on algorithmic decision-making processes with social relevance: a view from the coalface**
Matthias Spielkamp
- 10:30 **Trust can be sold: a set of legal rules for the proper design, development and implementation of AI**
Robert Kroplewski
- 11:00 Discussion
- 11:15 Short break

IV. ARTIFICIAL INTELLIGENCE, PREDICTIVE POLICING AND THE INTERSECTION BETWEEN ETHICS AND PRIVACY

Chair: Laviero Buono

- 11:30 **Ethics as an essential element in developing computer systems able to simulate human-like intelligence**
Philipp Slusallek
- 12:00 Discussion
- 12:10 **Artificial Intelligence, big data and the General Data Protection Regulation (GDPR)**
Ian Walden
- 12:40 Discussion
- 12:50 End of online conference

For programme updates: www.era.int
Programme may be subject to amendment.

Save the date

Online Annual Conference on European Border Management 2020
22-23 October 2020

Annual Conference on European Criminal Justice 2020
Event will be live streamed
Trier, 12-13 November 2020

Sanctions in the EU's External Relations
Trier, 3-4 December 2020

Annual Conference on Countering Terrorism in the EU 2020
Trier, 10-11 December 2020

ERA e-Learning courses

The Case Law of the CJEU for Criminal Law Practitioners: 10 Key Questions
Specialised e-learning course
This e-learning course answers 10 key questions on the general principles of EU law as applied by the CJEU to criminal justice and its interpretation of concepts of substantive and procedural criminal law.

The European Arrest Warrant: 10 Key Questions for Defence Counsel
Specialised e-learning course
This e-learning course answers 10 key questions on the role of the defence lawyer in European Arrest Warrant (EAW) proceedings.
Each answer provides the reader with a structured overview of the topic as well as detailed information regarding relevant instruments and cases.

www.era.int/elearning

Table of Contents

320SD61

TABLE OF CONTENTS

I GENERAL INFORMATION ABOUT THE SEMINAR

II SPEAKERS' CONTRIBUTIONS

III BACKGROUND DOCUMENTATION

General Information

Artificial Intelligence (AI) and the Criminal Justice System

5. - 6.11.2020

320SD61

List of speakers

Sophia Adams Bhatti
Head of Strategy and Policy
Simmons Wavelength Ltd
50-60 Station Road
UK-CAMBRIDGE CB1 2JH

Francien Dechesne
Assistant Professor
Leiden University
PO Box 9520
NL-2300 RA LEIDEN

Dick Heimans
Deputy Head of Unit
Directorate B Unit 2
BE-1049 BRUSSELS

Robert Kroplewski
Solicitor
Robert Kroplewski Law Office
ulica Orężna 47
PL-02-937 WARSZAWA

Cornelia Kutterer
Senior Manager Regulatory Policy
Microsoft
Rue Montoyer 51
BE-1000 BRUSSELS

James Luke
Distinguished Engineer
IBM Academy
Master Inventor
Hursley Park
UK-WINCHESTER SO21 2JN

Anjali Mazumder
Justice and Human Rights Theme Lead
British Library
The Alan Turing Institute
96 Euston Road
UK-LONDON NW1 2DB

Sigrit Siht
Director of Data Policy
Ministry of Economic Affairs and Communications
Department of State Information Systems
Suur-Ameerika 1
EE-10122 TALLINN

Professor Dr Philipp Slusallek
Executive Director
German Research Centre for Artificial Intelligence
Trippstadter Strasse 122
DE-67663 KAISERSLAUTERN

Matthias Spielkamp
Co-Founder and Executive Director
AW AlgorithmWatch gGmbH
Linienstrasse 13
DE-10178 BERLIN

Europäische Rechtsakademie Trier
Academy of European Law Trier
Académie de Droit Européen de Trèves

Artificial Intelligence (AI) and the Criminal Justice System

5. - 6.11.2020

320SD61

List of speakers

Simone van der Hof
Professor
Leiden University, Faculty of Law
Steenschuur 25
NL-2311 ES LEIDEN

Prof. Ian Walden
Centre for Commercial Law Studies
Queen Mary, University of London
67-69 Lincoln's Inn Fields
UK-LONDON WC2A 3 JB

Speakers' Contributions

Francien Dechesne

Francien Dechesne is assistant professor at the eLaw Center for Law and Digital Technologies within the Leiden Law School.

She studied mathematics, philosophy, and computer science next to a short professional excursion in IT Management. She holds a PhD in formal logic. She has postdoctoral research experience in the field of formal verification of computer security, and in Ethics of Information Technology. The challenge of formalising non-functional requirements like privacy brought her interest to frameworks of design for values, and understanding the broader societal implications of the expanding use of digital technologies in our lives. In 2015, she joined the eLaw-Center in the Leiden Law School for a research project on responsible innovation with data-analytics.

She was appointed as assistant professor on Ethics and Digital Technologies in 2018. Her current research focuses on methods for interdisciplinary translation in technical and regulatory developments around AI, in particular on the topics of fairness, accountability and transparency in algorithmic systems. Francien was project leader of a project on Ethics and AI at the Dutch Police in 2019.

Dick Heimans

Dick Heimans is Deputy Head of the General Criminal Law Unit in the Directorate-General for Justice and Consumers of the European Commission. He mainly works on establishing the European Public Prosecutor's Office, Eurojust and ECRIS, the European system for criminal records exchange. He studied international and criminal law at Leiden University before joining the Dutch Ministry of Justice in 1989, where he worked on international law enforcement issues. In 1995, he became the first legal adviser to Europol, and in 1999 Head of the Legal Affairs unit at that organisation. In 2004, he joined the European Commission, where he has worked in policy areas such as cybercrime, counter-terrorism and protection of critical infrastructure before accepting his current position.

Putting forward a European approach to artificial intelligence and robotics: work carried out by the European Commission

Dick HEIMANS
DG Justice & Consumers
European Commission

ERA event – 5-6 November

1

The questions

- What is AI?
- What European policies on AI already exist?
- How is AI relevant for criminal law in the EU?
- What policy should be recommended?

2

Artificial Intelligence for Europe

- Boost technology, industrial capacity, uptake across the economy
- Prepare for socio-economic changes
- Ensure appropriate ethical and legal framework

Coordinated plan on Artificial Intelligence

- World-leading region in cutting-edge, ethical and secure AI
- Promoting human-centric approach
- Ethics by design
- Building on Charter of Fundamental Rights
- Assessing whether legal framework is fit-for purpose

Building Trust in Human-Centric Artificial Intelligence

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

5

AI White Paper - 1

- Ecosystem of excellence, ecosystem of trust
- Repeats seven requirements for trustworthy AI
- Regulatory framework needed - risk-based, not overly prescriptive
- Two main risk areas:
 - fundamental rights/data protection/non-discrimination
 - Safety/effective functioning of liability regime
- Tackling high-risk applications in particular, per sector and within sector

6

AI White Paper - 2

- Applicable requirements
 - Training data
 - Data and record-keeping
 - Information to be provided
 - Robustness and accuracy requirements
 - Human oversight
 - Specific requirements per AI application
- Governance mechanism
 - Ex ante/ex post? Both?
 - MS level or EU level?
 - Coordination challenge?
- Consultation period
- Legislative proposal Q1 2021

7

Interim Conclusions

- Developing policy – ethical and legal approaches prominent
- Ethical frameworks need to be tested and further developed
- Existing frameworks offer protection, but is it enough?
- No policy work yet specific to the criminal justice area at EU / international level

8

AI and criminal justice - issues

- Current and future usage: prevention, detection, investigation, prosecution, court practice, sentencing
- Risk assessment methodology – what are high risk use cases/applications?
- AI enabled crime – crime typology
- Criminal liability
- Expert meeting AI and criminal law October 2020 – no policy agreed yet!

9

AI and criminal justice - prevention

- Predictive policing
- Cybersecurity
- No-fly lists
- Sex trafficking

10

AI and criminal justice - detection

- Facial recognition
- Combating fake news
- Scene understanding
- Gunshot detection and analysis
- DNA analysis
- Administrative data – fraud and money laundering
- Robot patrol systems

11

AI and criminal justice - investigation

- Localisation of the crime
- Data lakes analysis
- Identification of victims
- Determination of the content of the crime
- Identification of child pornography
- Statement taking machines
- Criminal analysis

12

AI and criminal justice - prosecution

- Directing law enforcement
- Providing explanations in Court on AI usage
- Judicial authorisations of AI usage?
- Decisions on pre-trial detention
- Sentencing guidelines vs AI advice
- Legal expert systems
- Automated fines, speech-enabled AI

13

AI and criminal justice - Courts

- Fair criminal trials
- "Robo-judges"?
- Independence of the judiciary
- Forensic expertise – expert witnesses
- AI-enabled "facial expression detector"
- AI manipulated evidence
- Legal expert systems

14

AI and criminal justice - Sentence

- AI Risk and threat assessment
- AI detection technology in prisons
- Automated guard systems

15

AI enabled crime - 1

- Expansion of existing threats
 - Making crime easier
- Introduction of new threats
 - Tasks which would be impossible for humans
- Change to the typical nature of threats
 - More effective, finely targeted, difficult to attribute

16

AI enabled crime - 2

- Digital security breaches
- Drones / automated vehicles
- Political crimes – suppress dissent, deepfakes
- Market manipulation, collusion
- Trafficking / sales banned substances
- Harassment, torture
- Theft/fraud/spearphishing

17

Criminal liability - the AI did it?

- Free will, consciousness, moral sense
- Criminal intentions
- Will there be "general AI"?
- How to punish an AI...
- What about automated vehicles?

18

The way forward?

- AI will become increasingly relevant to criminal justice
- General norms on AI need to be translated into frameworks for specific applications
- How do we get there for criminal justice?
- Do we need regulation? If yes, at what level?

19

Issues to be looked at - 1

- AI assisted legal analysis for criminal courts
- AI assisted judging in criminal courts
- Juries in criminal trials and AI
- AI assisted sentencing
- AI generated evidence
- Detection of fake evidence (audio/video)
- AI assisted forensics
- AI-enabled crime and causality/criminal liability
- Review of existing legal offences for completeness with respect to AI-enabled crimes

20

Issues to be looked at - 2

- Standardisation/Certification/Licensing of AI as “ethical by design” or as “embedding legal norms in the design” for criminal justice purposes?
- Further developing the current ethical framework for usage in the criminal justice field?
- Training on AI for legal practitioners
- Investment/funding for development of AI to support criminal justice needs
- Research on AI for criminal justice

21

Issues to be looked at - 3

- Explainability of law enforcement AI applications in criminal courts
- Cross-border admissibility of AI-generated evidence
- AI to support the execution of mutual legal assistance requests
- Lining up criminal justice AI initiatives with initiatives on AI for security/migration – joined-up approach

22

Thank you for your
attention!

Anjali Mazumder

Sophia Adams Bhatti

Sophia is a public policy expert with 20 years' experience working across a number of sectors including the law, financial services, competition, consumer affairs, health care and immigration and asylum. She is currently the head of strategy and policy at Simmons Wavelength where she leads on the intersection of law, technology and policy. Previously, she was the director of policy at the Law Society of England and Wales, where she led, amongst other things, all law reform programs and the Law Society's ground-breaking work on the use of AI in the Criminal Justice sector. She speaks regularly across the world on policy, regulation and technology, and is a member of a number of advisory groups on related issues.

Sigrit Siht

My bio (in third person): Sigrít Siht is the Director of Data Policy. Her main focus is open data but she is an advocate of the use of AI to make the government more efficient and save money. Currently she is also writing her master's thesis on smart city development.

AI in Governmental Activities

Sigrit Siht

Director of Data Policy

GCIO Office

Ministry of Economic Affairs and Communications of Estonia

05.11.2020

© Karl Anker Adams

The scope

1

Overview of the Estonian
mindset

2

#KrattAI

3

AI use cases

Situation in Estonia

- 1 328 976 people living in Estonia in 2020.
 - 1/3 living in the capital city of Tallinn.
 - 1 179 667 prognosis for 2080.
- E-friendly and e-trusting society:
 - E-voting.
 - E-tax refunds.
- IT programmes are very popular in universities and secondary schools.

Government AI Readiness Index 2020

- 17th place in overall.
 - Objective analysis.
 - Time and place to make necessary changes.
- 1st place in the Responsible AI Sub-Index.
 - Only some countries were in the race.

#KrattAI

- Estonian mythology.
- 50 AI solutions (kratt) in the public sector by the end of 2020.
- Problem: people don't know which public agency to contact.
- Opportunity: improvements in AI enable to solve this problem significantly better than before.
- Background: mobile devices are hugely popular.
- Aim: make digital public services easier to use.

#KrattAI roadmap for 2020:	Mar '20	Jun '20	Sep '20	Dec '20	Mar '21	Jun '21
Chatbot and messaging room PoC						
Analysis and PoC on the use of alternative channels						
Classification of citizen messages analysis and PoC						
Analysis and PoC of eesti.ee national mobile application						
Information extraction						
Speech synthesis						
Consent management platform						
Roadmap with companies						
Roadmap with Finland						

*PoC = Proof of Concept

Chatbots

- Many public sector agencies have their chatbots:
 - Iti.
 - Suve.
 - Vesta.
- Unified standard training sets:
 - Profanity.
 - General questions.
 - What time is it?
 - What's the weather like?
- No vendor lock:
 - The training data is public sector owned.

SATIKAS

- SATIKAS is an automated satellite-based information system.
- Detects all mowing on agricultural grasslands in Estonia.
 - Automate the EU's agricultural subsidy checks.
- Help to decrease and replace the inspectors' field visits.
- A reminder to the farmers who have not yet fulfilled the mowing requirements.
- 65 000 € personnel expenses and 600 000 € in correct grants annually.

SATIKAS – how it works

- Sudden changes in the biomass levels in the time series parameters of the satellite data.
- The original satellite pictures are converted into several processed data layers.
- The average value is calculated of each date's picture and each grassland.
- A deep learning algorithm is used on the time series for detecting the mowing information.
- The information received from SATIKAS will then move into the information systems of the Estonian Agricultural Registers and Information Board.

HANS

- Transcribing the Estonian Parliament's (Riigikogu) sessions.
- Publish the verbatim report within one hour from the end of the sitting.
- At least 93-95% correctness for plenary sittings.
- Training data: 1 500 hours of recordings and transcripts.

Automated keywording

- National Library of Estonia's employees manually keyword different kinds of texts.
- Estonian Subject Thesaurus contains 40 000 keywords.
- Training data available only consisted of 8 000 keywords.
- Vector calculations.
- Analysis + PoC.

An aerial photograph of a dense forest. The trees are mostly green, but many have turned yellow, indicating autumn. The perspective is from directly above, looking down on the canopy.

aitäh + questions

Sigrit Siht
sigrit.siht@mkm.ee
LinkedIn Sigrit Siht

Simone van der Hof

Professor Simone van der Hof is the head of the Center for Law and Digital technologies (eLaw) at Leiden Law School, Leiden University. She is also the programme director of the Master of Laws Advanced Studies programme in Law and Digital Technologies and one of the directors of the research profile area Interaction Between Legal Systems at Leiden Law School. Her academic interest is in the field of children's rights in the digital world with a particular focus on privacy, data protection and economic exploitation. Simone has over 20 years of academic teaching experience. She coordinates and/or teaches several courses at master level: Digital Child Rights (Advanced Master Law and Digital Technologies); Regulating online child safety (Master Youth Law), The Rights of the Child in the Digital World (Advanced Master International Children's Rights).

Sweetie: using artificial intelligence to fight cybercrime

Professor Simone van der Hof

Center for Law and Digital Technologies



Universiteit
Leiden
The Netherlands



Discover the world at Leiden University

Persistent
poverty

Increasing
internet use

trends

Live online sexual
child abuse

Discover the world at Leiden University

Live online sexual child abuse

- Children performing sex shows before a webcam
- Form of child abuse that grows exponentially
⇒ 750,000 men online at any given moment looking for opportunities to sexually abuse children online
- Not a new phenomenon (engaging children in sexual performances) but new is abusing children remotely
- Live streams (instantaneous transmission)
⇒ no evidence (unless recorded)
⇒ (hardly) no prosecutions

Discover the world at Leiden University

Article 19

UN CRC

Article 34

Lanzarote Convention

Article 21

protection against

Optional Protocol to the Convention on the Rights of the Child on the sale of
children,
child prostitution and child pornography

sexual abuse

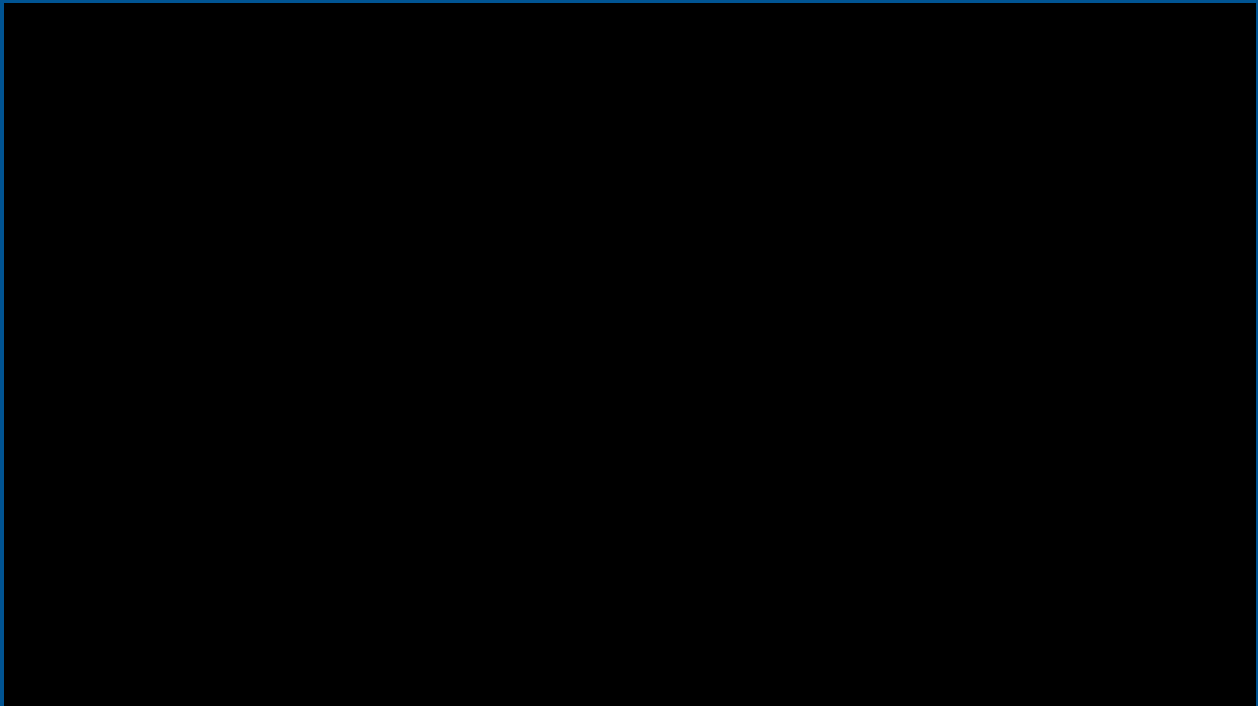
Discover the world at Leiden University

right to

Live online sexual
child abuse?

protection against sexual abuse

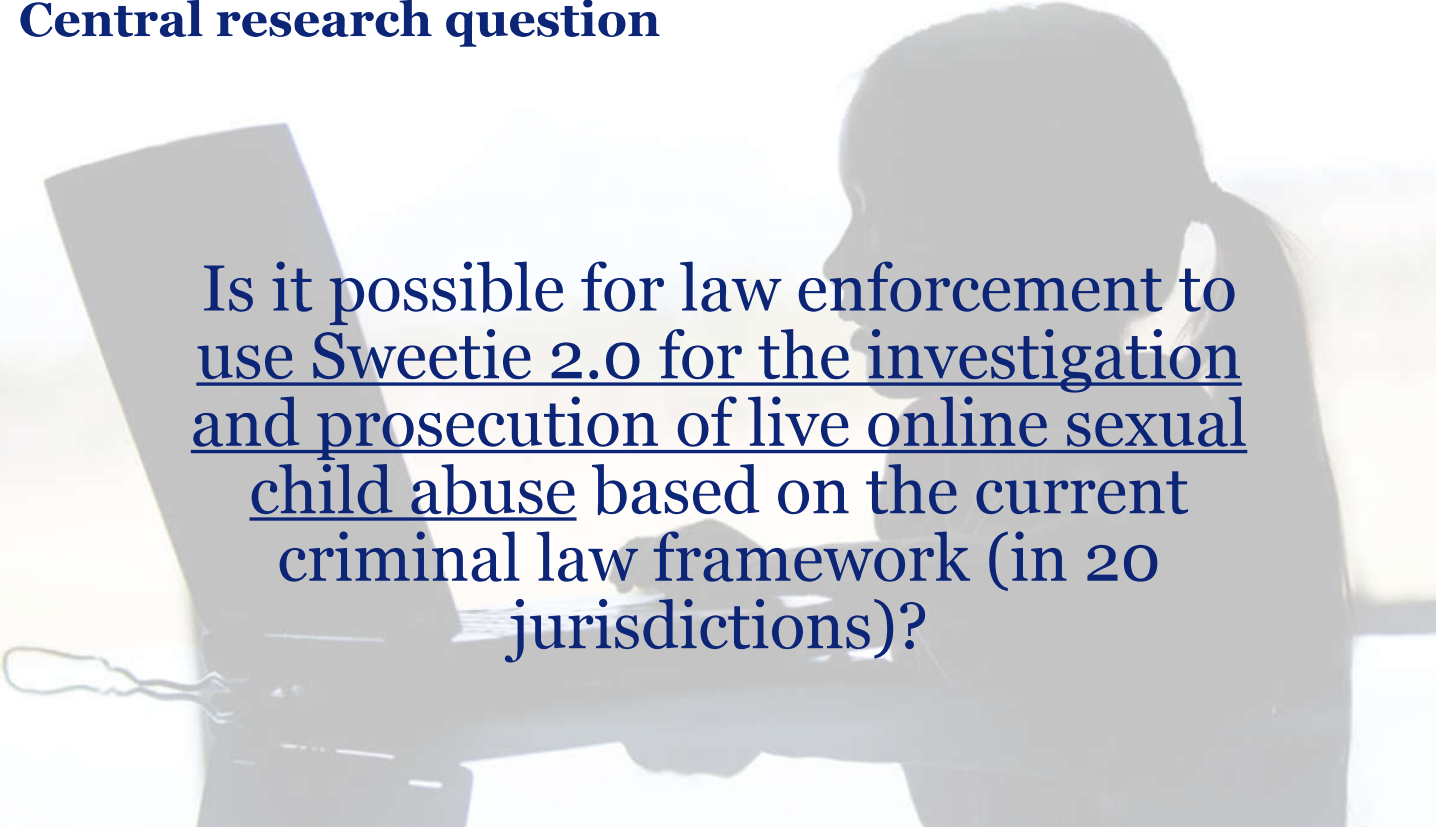
Discover the world at Leiden University



Sweetie 2.0

Discover the world at Leiden University

Central research question



Is it possible for law enforcement to use Sweetie 2.0 for the investigation and prosecution of live online sexual child abuse based on the current criminal law framework (in 20 jurisdictions)?

Discover the world at Leiden University



Subquestions

- How is live online sexual child abuse criminalised?
- Do crime descriptions apply to virtual 'victims'?
- Is the (proactive) investigation of live online sexual child abuse using AI tools like Sweetie 2.0 allowed?
- Are there specific limitations in criminal procedure when it comes to entrapment and what are the consequences of this for using Sweetie 2.0?

The technology

1. 3D imagery ⇒ realistic representation of a child

2. Chatbot facility ⇒ eliminate human intervention

- based on the experiences, work instructions and chat logs from the initial Sweetie project
- the conversation model will simulate as realistically as possible a fictitious 10/11-year-old child

3. Software framework

- Store chats
- Process identifiable material
- Recognise repeating chat partners, indecent proposals/ or explicit materials
- Confront potentially offending chat partners with their own behaviour and chat phrases



Legal challenges



Substantive
criminal law

Complicating factor 1: Virtual victim (1)

Substantive criminal law

- Most crime descriptions contain the element of a victim
- This victim is a real person
- Impossible to fulfil all constituent elements of the crime

Discover the world at Leiden University

Complicating factor 1: Virtual victim (2)

Substantive criminal law

But...

- Some criminal systems contain provisions on 'fictitious persons' (mainly grooming)
- Virtual child pornography is also criminalised in some jurisdictions
- Interacting with Sweetie might present probable cause for law enforcement

Discover the world at Leiden University

Complicating factor 2: No sexually explicit behaviour

Substantive criminal law

Virtual child pornography is also criminalised in some jurisdictions but...

- Sweetie does not undress
- Or engage in sexually explicit behaviour

Discover the world at Leiden University

Criminalisation of attempt: Another viable avenue?

Substantive criminal law

Sweetie = an inadequate object $\Rightarrow$ inadequate attempt?

- Relative inadequacy of the attempt (factual impossibility)
- Absolute inadequacy of the attempt (legal impossibility)



Discover the world at Leiden University

Sweetie — relative or absolute inadequate object?

Discover the world at Leiden University

Criminalisation of attempt: Another viable avenue?

Substantive criminal law

Sweetie = relative or absolute inadequate object?

⇒ depends on emphasis of the criminal system:
objective or subjective elements?

Discover the world at Leiden University

Criminal procedure

Investigative tool

Sweetie is ...

- a lure
- an undercover agent
- a recording device

👉 No explicit rules on use of AI agents for criminal investigations

👉 Requires specific authorization if interfering with suspect's rights

Potential interference with the suspect's rights: privacy

Relevant AI operations:

- Observing behaviour in chat room
- Recording information in chat room
- Engaging in one-on-one conversations
- Recording one-on-one conversations

Criminal procedure

Discover the world at Leiden University

Potential interference with the suspect's rights: privacy

Relevant AI operations:

- Observing behaviour in chat room
- Recording information in chat room
- Engaging in one-on-one conversations
- Recording one-on-one conversations

Criminal procedure

Public sphere: lower expectations of privacy for suspects

Level of protection: differs per jurisdiction

Discover the world at Leiden University

Potential interference with the suspect's rights: privacy

Relevant AI operations:

- Observing behaviour in chat room
- Recording information in chat room
- Engaging in one-on-one conversations
- Recording one-on-one conversations

Criminal procedure

Purpose

How **severe** is the interference?

Is it in **accordance with the law** (existing investigatory powers)?

Discover the world at Leiden University

Potential interference with the suspect's rights: fair trial

Permissible versus non-permissible entrapment...

- Targeted entrapment (particular suspect)
- Non-targeted entrapment (general suspicion regarding particular place or location)

Criminal procedure

Discover the world at Leiden University

Targeted entrapment

Criminal procedure

‘*essentially* passive’ standard (ECHR)

- not pressure the suspect to commit crimes, e.g. by pro-actively contacting him
- not prompt the suspect to continue after initial refusal
- not make an offer that is too hard to refuse
- appeal to the suspects compassion

Discover the world at Leiden University

Non-targeted entrapment

Criminal procedure

Different criteria amongst jurisdictions

- Same criteria as targeted entrapment
- Reasonable suspicion that criminal activities are occurring there
- No substantial change of the original situation

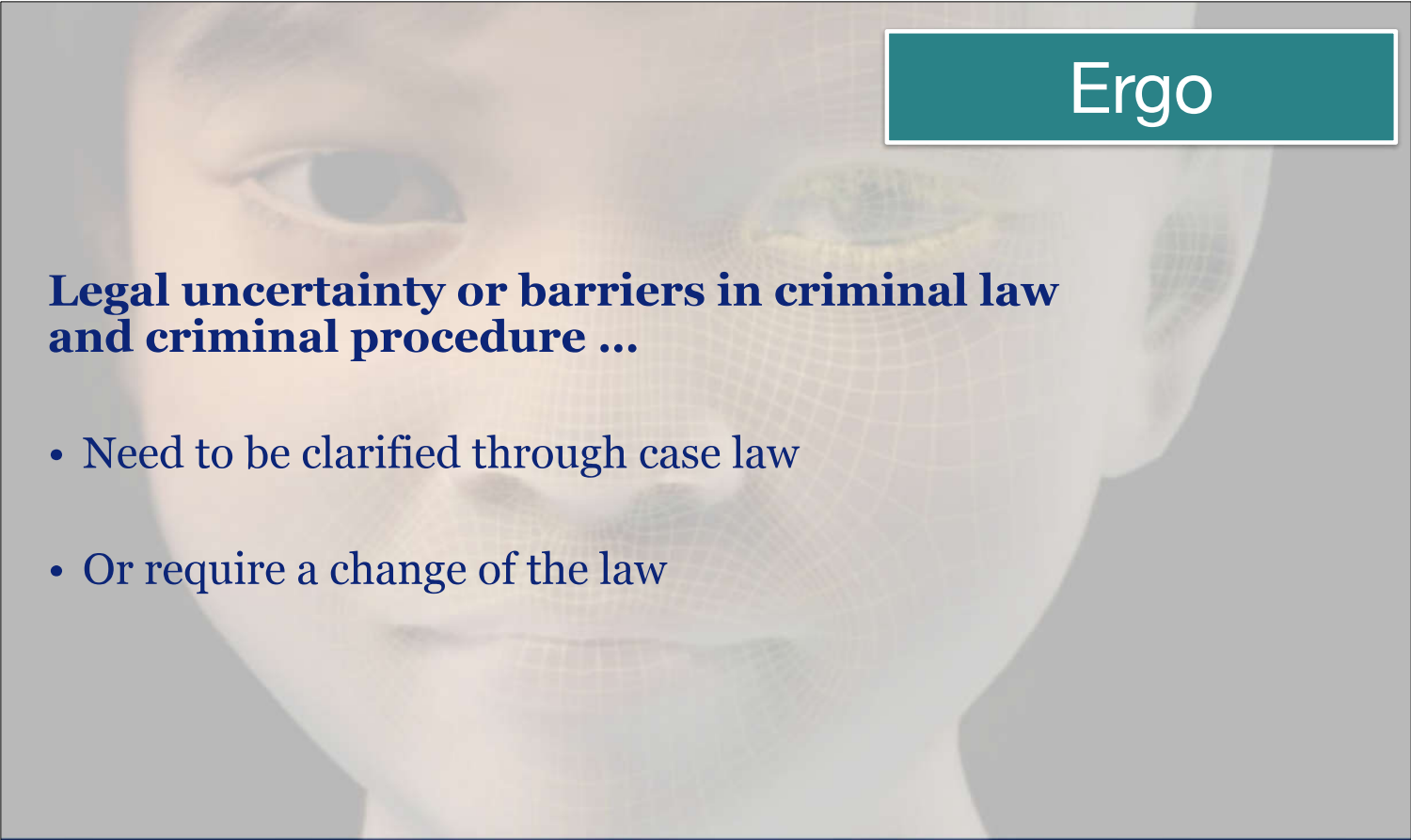
Discover the world at Leiden University

Conclusions

Ergo

Sweetie ...

- Is essentially passive
- Does not solicit or steer the conversation
- Leaves room for the suspect to retreat
- Does not significantly alter the situation



Ergo

Legal uncertainty or barriers in criminal law and criminal procedure ...

- Need to be clarified through case law
- Or require a change of the law

Discover the world at Leiden University

Thank you for your attention



**Universiteit
Leiden**
The Netherlands

Discover the world at Leiden University

Cornelia Kutterer



Cornelia Kutterer
Senior Director, Rule of Law & Responsible Tech
European Government Affairs, CELA
Microsoft

Cornelia leads Microsoft's European Rule of Law & Responsible Tech team which focuses on the impact of new technologies and regulatory frameworks that meet expectation of society. Her team covers policies such as responsible/ethical/trustworthy AI, digital safety and content regulation, privacy, lawful access, human rights and competition. In her role, she works hand in hand with Microsoft's Office of Responsible AI, Microsoft Research and members of the AI, Ethics, and Effects in Engineering and Research (Aether) Committee. She regularly engages with leading European scholars in these fields to advance academic thinking. Cornelia has long standing experience in Information Society & Internet policies and speaks regularly at regional and international conferences. Before joining Microsoft, she headed the legal department of BEUC, the European Consumer Organisation. She has also gained experience in a top 10 law firm and started her professional career in the European Parliament as a political advisor to an MEP. Cornelia is a qualified German lawyer, and holds a master's degree in information technology and telecommunication laws. She studied law at the Universities of Passau, Porto, Hamburg and Glasgow/Strathclyde.

James Luke

Matthias Spielkamp

Matthias Spielkamp is co-founder and executive director of AlgorithmWatch (Theodor Heuss Medal 2018, Grimme Online Nominee 2019). He testified before committees of the Council of Europe, the European Parliament, the German Bundestag and other institutions on AI and automation. Matthias serves on the governing board of the German section of Reporters Without Borders and the advisory councils of Stiftung Warentest and the Whistleblower Network. He was a fellow of ZEIT Stiftung, Stiftung Mercator and the American Council on Germany. Matthias received several awards (Grimme Online Award, Lead Award, Klicksafe Award, .info Award) for the magazine iRights.info he co-founded and led as managing editor and publisher for more than ten years. He has written and edited books on the automation of society, digital journalism and Internet governance and holds master's degrees in Journalism from the University of Colorado in Boulder and in Philosophy from the Free University of Berlin.

Robert Kroplewski



The logo features a circular emblem with a blue arc at the top and a ring of yellow stars. Below the emblem, the text 'ERA' is written in large blue letters, and 'Academy of European' is written in smaller blue letters below it.

Trust can be sold: a set of legal rules for the proper design, development and implementation of AI

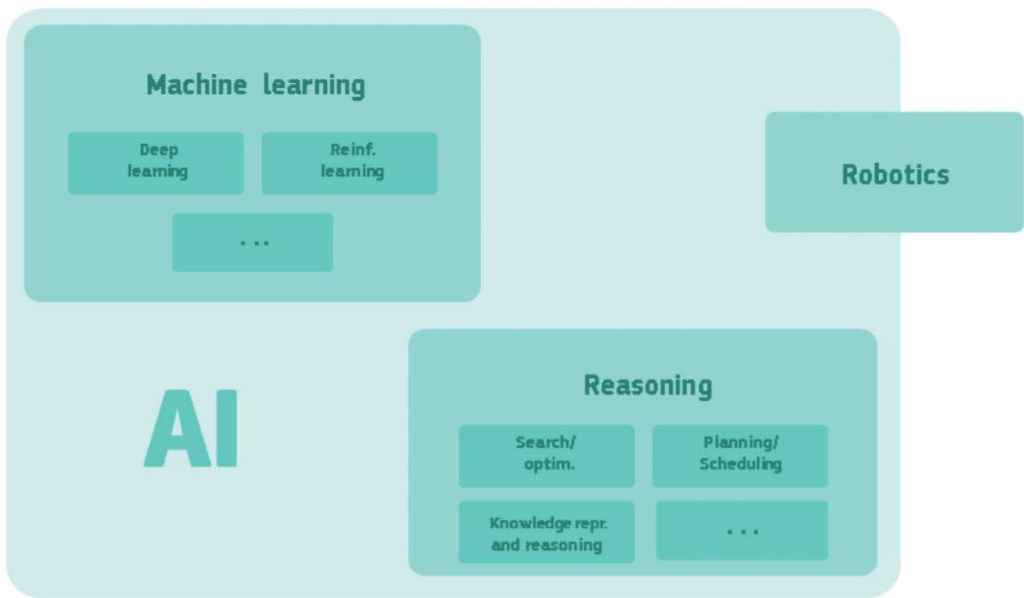
Artificial Intelligence (AI) and the Criminal Justice System
AI to predict crimes and as a criminal intelligence tool

Conference, Nov. 2020

www.kroplewski.com

1

Unclosed set of AI technologies

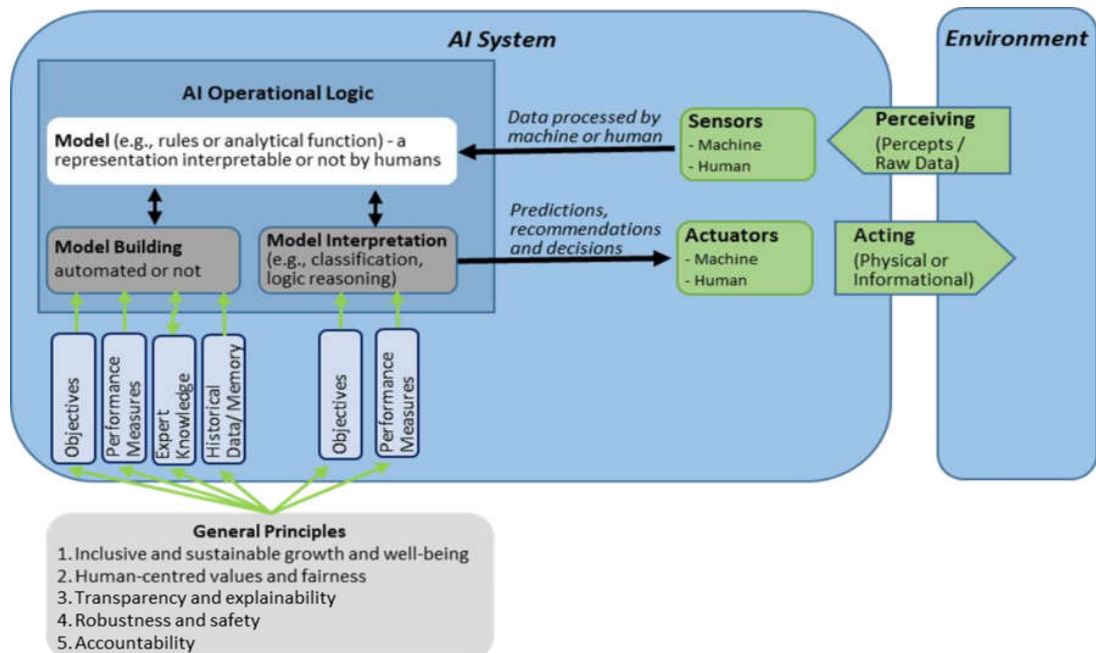


The diagram illustrates the relationship between various AI sub-disciplines. A large light blue rounded rectangle is labeled 'AI' in the bottom left corner. Inside this rectangle, there are two smaller rounded rectangles. The top-left one is labeled 'Machine learning' and contains three smaller boxes: 'Deep learning', 'Reinf. learning', and an ellipsis '...'. The bottom-right one is labeled 'Reasoning' and contains three smaller boxes: 'Search/optim.', 'Planning/Scheduling', and 'Knowledge repr. and reasoning', followed by an ellipsis '...'. To the right of the 'AI' rectangle, outside of it, is a separate rounded rectangle labeled 'Robotics'.

Figure 2: A simplified overview of AI's sub-disciplines and their relationship. Both machine learning and reasoning include many other techniques, and robotics includes techniques that are outside AI. The whole of AI falls within the computer science discipline.

2

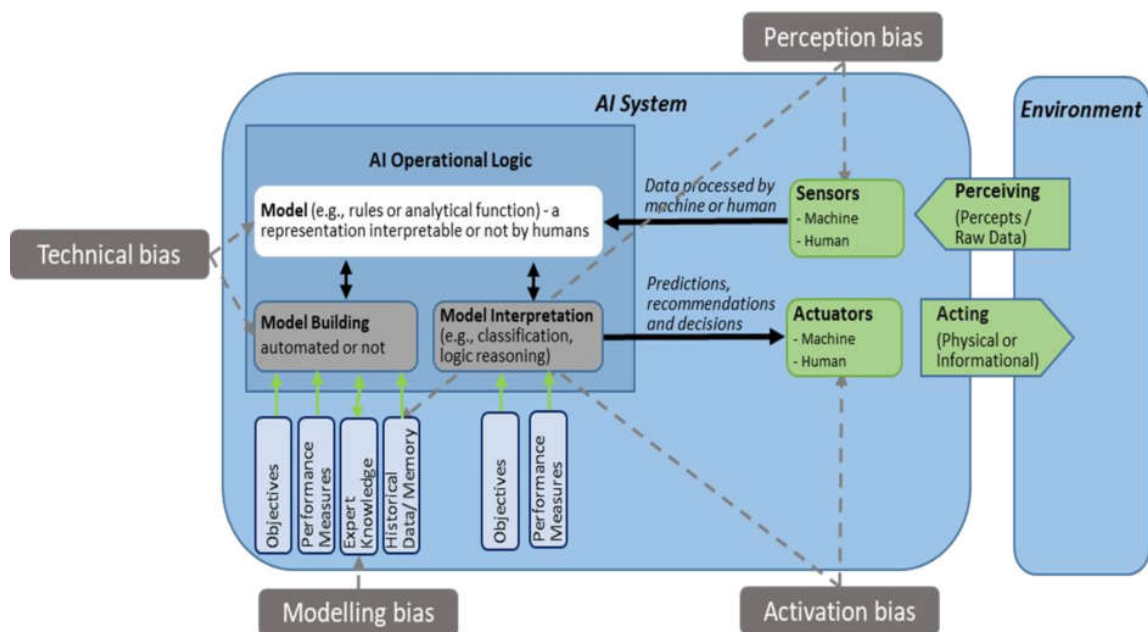
Values modeling in AI system



www.kroplewski.com

3

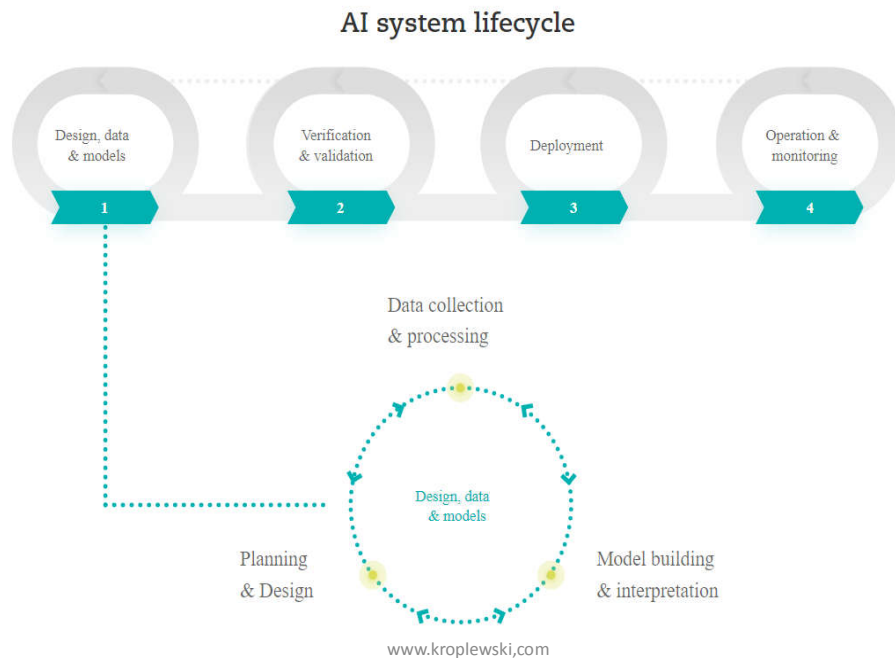
Roads of risk of bias embedding



www.kroplewski.com

4

AI System lifecycle



5

Black-box AI and explainability

- Some machine learning techniques, although very successful from the accuracy point of view, are very opaque in terms of understanding how they make decisions. The notion of black-box AI refers to such scenarios, where it is not possible to trace back to the reason for certain decisions.
- Explainability is a property of those AI systems that instead can provide a form of explanation for their actions

Goal-directed AI

- Current AI systems are goal-directed, meaning that they receive the specification of a goal to achieve from a human being and use some techniques to achieve such goal.
- They do not define their own goals. However, some AI systems (such as those based on certain machine learning techniques) can have more freedom to decide which path to take to achieve the given goal.

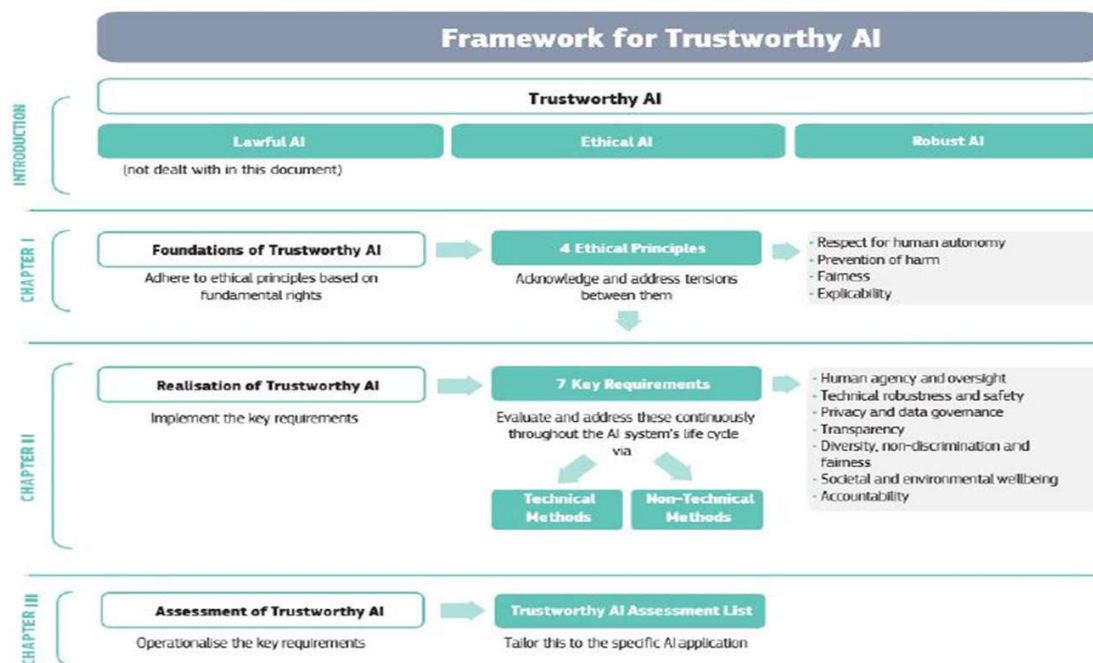


Figure 1: The Guidelines as a framework for Trustworthy AI

D4 of AIHLEG's Observations for Public Sector: Justice and law-enforcement

- In the sectoral workshop on public services, specific attention was devoted to justice and law enforcement. AI-enabled solutions have proven to be at once potentially beneficial and controversial, due to the possibility that seemingly efficient solutions amplify (or fail to address) bias and discrimination.
- This in turn could lead to cases of lack of (access to) justice
- Secure individual and societal well-being
- Promote interoperability of frameworks that enable efficient and secure communication between jurisdictions. Develop and deploy interfaces that facilitate and promote cross border and other stakeholders.
- Promote best data (evidence) practice

New additions (1/2)

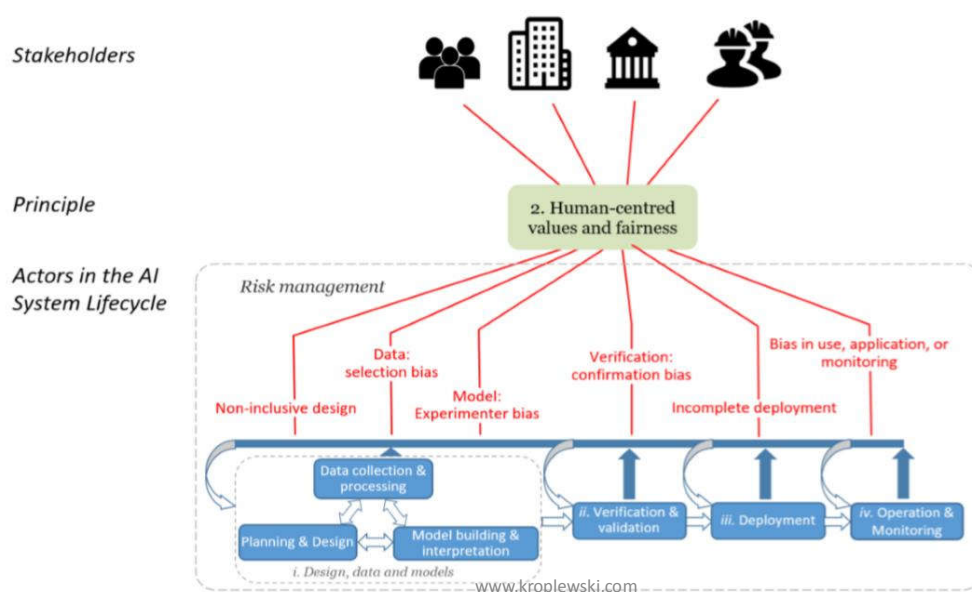
- ***Support the development of AI applications and interfaces that improve citizens and legal entities' access to justice.*** AI can lead to more effective access to justice *i.a.* by offering guidance and helping individuals understand their rights and their specific legal situation. Several initiatives around the world have been launched to facilitate access to justice through AI solutions: emphasis on trustworthy AI is particularly important to ensure that these solutions are fully respectful of fundamental rights.
- ***Consider promoting the development of strong natural language processing (NLP) competence that includes domain knowledge and common sense reasoning.*** Justice is a very challenging domain of application for NLP because it often requires understanding and correlating items that are of very different sorts (such as legal texts, phone conversations, etc.). The digitisation of proceedings is a precondition for the training of NLP applications in this domain.
- ***Consider ensuring the right to a fair trial by securing equitable access to solutions such as personal AI assistants.*** The right to a fair trial requires that parties involved in litigation benefit from equal access to AI assistants as other parties. This right must also be guaranteed whenever a court uses an AI assistant to support decision-making.

New additions (2/2)

- **Allow free open and AI-assisted access to case records to help identify and understand biases and errors.** Professionals and experts will increasingly use AI tools to check biases and errors in judicial decisions, which may help detect the existence of unfair and discriminatory prejudice. To protect privacy, access to these data should occur under strict non-disclosure agreements, and data should be anonymised whenever possible and appropriate.
- **Launch a reflection on the implications of AI-assisted and AI-enabled decision-making systems for law enforcement and justice decisions.** Though AI-assisted and AI-enabled decision-making systems are already used in specific law enforcement and justice contexts, their deployment at greater scale generates risks and opportunities that are not yet fully understood. Much more research, scrutiny and deliberation are needed prior to formulating legal, ethical and policy guidance in relation to the use of AI-assisted decision-making systems in law enforcement and justice.
- **Provide explicit rules for the admissibility of evidence or decisions** produced by AI systems in law enforcement and justice proceedings

Human centred values & fairness in validation of AI

Figure 8. A view of fairness considerations by AI actors within the AI system lifecycle



Compass of trade-offs

Human dignity, well-being and prevention of harm

Second Draft of the Globac Cod of Ethics by AHEG – UNESCO

See: <https://en.unesco.org/artificial-intelligence/ethics>

Values Foundation of the EU - TrustworthAI

The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities⁴. These values are common to the societies of all Member States in which pluralism, non-discrimination, tolerance, justice, solidarity and equality prevail. In addition, the EU Charter of Fundamental Rights brings together – in a single text – the personal, civic, political, economic and social rights enjoyed by people within the EU.

European commission's work in progres in justice

- The European Commission for the Efficiency of Justice (CEPEJ) adopted in December 2018 the European Ethical Charter for the use of artificial intelligence in judicial systems which sets **5 key principles (respect of fundamental rights, non-discrimination, quality and security, transparency, impartiality and fairness, "under the control" of the user)** for the use of AI systems in this field. The CEPEJ is currently studying the advisability and feasibility of a certification or labelling framework for artificial intelligence products used in judicial systems. The European Committee on Legal Co-operation (CDCJ) is preparing guidelines to ensure the compatibility of these mechanisms with Articles 6 and 13 of the Convention on Human Rights. The European Committee on Crime Problems (CDPC) is currently studying the topic of AI and criminal law and may propose the creation of a new specialised legal instrument.

[CEPEJ, European Ethical Charter on the use of artificial intelligence in judicial systems and their environment - CEPEJ\(2018\)14, December 2018](#)

See CDPC(2020)3Rev, [Feasibility Study on a future Council of Europe instrument on a future Council of Europe instrument on AI and criminal law.](#)

Context of AI Legal Regulations

• OECD/AIGO		• <u>Stewardship of Trustworthy AI</u> (<u>governance</u> , <u>data</u> , <u>knowledge</u> , <u>AI's</u> <u>models</u>)
• EC/AIHLEG		• <u>Guidelines of Trustworthy AI and Policy and Investment of T-AI</u>
• UNESCO		• <u>Global Ethical Code for competition of ethics of AI</u>
• WIPO		• <u>Intellectual Property Framework for AI</u>
• <u>Council of Europe</u>		• <u>Fremwork of fundalemtal rights, rule of law and democracy</u>

De lege ferenda

- Ex-ante test and long live cycle audit of AI System adopted in justice and in criminal trial and proceedings
- Modeling AI System for justice with respectful implementation every legal rules of trial and criminal proceedings
- High-risk application under mutual recognised certification system
- Human oversight but in hand of judge oversight
- Making difference of liability of AI referring to operators and end-users
- Judge as end-user responsible for trial harms and lack of access to justice
- Outputs came from AI System are only source of knowledge for judges but not evidences needed to be established as part of factual findings under trial rules



Minister of
Digital Affairs
Republic of Poland



Thank you very much for attention,
see you in AI's complexity and reality

Robert Kroplewski

solicitor,
plenipotentiary of Minister of Digital
Affairs in Poland resp. for Information
Society
CEO of Future Industry Digital Platform
expert #AIHLEG (EC)
expert #AIGO (OECD)
expert #CAHAI (CoE)

robert@kroplewski.com

www.kroplewski.com

TT@bobo.kroplewski

Philipp Slusallek

Ian Walden

Ian Walden is Professor of Information and Communications Law and Director of the [Centre for Commercial Law Studies](#), Queen Mary, University of London. His publications include *Media Law and Practice* (2009), *Free and Open Source Software* (2013), *Computer Crimes and Digital Investigations* (2nd ed., 2016) and *Telecommunications Law and Regulation* (5th ed., 2018). Ian has been a visiting professor at the universities of Texas, Melbourne and KU Leuven. Ian has been involved in law reform projects for the World Bank, European Commission, Council of Europe, Commonwealth and UNCTAD, as well as numerous individual states. Ian was a ‘expert nationaux détaché’ to the European Commission (1995-96); Board Member and Trustee of the Internet Watch Foundation (2004-09); on the Executive Board of the UK Council for Child Internet Safety (2010-12); the Press Complaints Commission (2009-14), a member of the RUSI Independent Surveillance Review (2014-15) and a member of the Code Adjudication Panel at the Phone-paid Services Authority (2016-). Ian is a member of the Commission’s [Expert Group](#) to support the application of the GDPR. Ian is a solicitor and Of Counsel to [Baker McKenzie](#). Ian leads Queen Mary’s [qLegal initiative](#) and is a principal investigator on the [Cloud Legal Project](#).

Artificial Intelligence, Big Data and the GDPR:

Professor Ian Walden



Introductory remarks

- Big (personal) data
 - Personal data, sensitive personal data, pseudonymised personal data, anonymised data
- Data protection regimes
 - Directive 95/46/EC
 - Directive 02/58/EC
 - Traffic and location data
 - Charter of Fundamental Rights, Art. 8
 - GDPR
- ‘Profiling’
 - Evaluate to analyse (past) or predict (future) behaviours

Data protection principles

- Lawfulness, fairness and transparency
 - Positive ('at least one' criterion) and negative (not in breach of a legal obligation)
- Purpose limitation
 - Compatible secondary purposes (art. 6(4))
- Data minimisation
 - Data protection by design and default (art. 25)
- Accuracy
 - Right to rectification (art. 16)
- Storage limitation
 - Archiving, scientific or historical research or statistical purposes (art. 89)
- Integrity and confidentiality
 - Accidental, unauthorised and unlawful (art. 3—34)

Data protection principles

- Accountability
 - “The controller shall be responsible for and able to demonstrate compliance with, paragraph 1” (the other principles)
 - organization must not only to be compliant with Regulation but it has also must be able to show its compliance
 - Document everything!
 - Records of processing (art. 30)
 - Data Protection Impact Assessments

AI decision making

- Profiling
 - Identifying patterns, making evaluations and predictions
 - Processing personal data & the 'personalisation paradox'
 - Human bias, prejudice, stereotyping...
 - Input bias, technical bias and emergent bias
 - ML/AI: ability to identify & disregard bias?

The Right

- GDPR, Article 22
 - right not to be subject to a decision based...
 - *solely* on automated processing (including profiling)
 - Requiring a human in the loop....but how much?
 - which produces legal effects / significantly affects individual
 - *Except* (depending on the lawful basis):
 - Performance of a contract
 - Authorised by law
 - Explicit consent
 - Special categories of data (22(4))
 - Guidelines (2017)
 - Art. 29 WP endorsed by EDPB

AI and individual autonomy

- Individual autonomy (self-determination) as a societal value
 - Embedded in commercial law & fundamental rights
 - e.g. privacy intrusions diminish autonomy
- Autonomy & legal fictions
 - Human fictions
 - Consent: “freely given, specific, informed and unambiguous indication”
 - Terms of service
 - ML/AI (as autonomous systems) fictions?
 - Liability, e.g. European Parliament Resolution (February 2017): “creating a specific legal status for robots”
 - Creativity: From *Next Rembrandt/Mozart’s 42nd Symphony* to *AIVA/No Man’s Sky*

Fairness

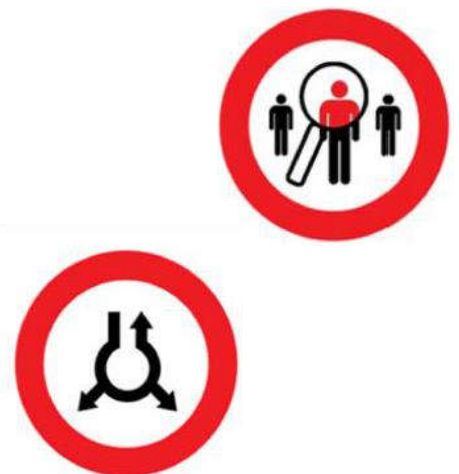
- Discrimination ≠ unfair treatment
 - Unlawful correct decisions
 - *Association Belge des Consommateurs Test-Achats ASBL v Conseil des Ministres* (Case C-236/09, 1 March 2011)
 - Positive discrimination to address systemic bias
 - Discrimination as personalisation, but...
 - e.g. insurance as socialised risk
- Training data sets / input data
 - Quality, quantity & labelling
- Correlation ≠ causation

Fairness

- Ethics & fairness
 - When is it unethical not to extract information from data?
- Justice & fairness
 - 'Code as code' (Lessig)
 - AI objectivity: good (removes bias) and bad (lacks empathy/discretion)
 - Appealing decisions
 - To whom?

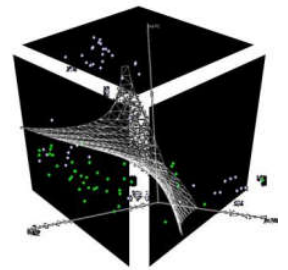
Transparency

- Notification
 - From data subject or a 3rd party
 - Existence of automated decision making, significance and consequences for data subject
 - Icons
- Access rights
 - Personal data
 - Meta-data



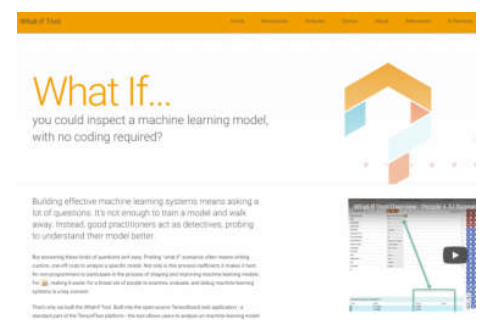
Explaining the 'black box'

- Why?
 - Protect individual rights, e.g. 'automated decision-making'
 - Protect competition (e.g. self-preferencing)
 - Regulation 2019/1150 'promoting fairness and transparency', Art. 5(6)
 - Determine liability, e.g. as evidence in litigation
- What?
 - 'meaning information about the logic involved' (GDPR, art. 13(2)(f))
 - The criteria / factors / parameters / variables used
 - The data (sets) fed into the system
- When?
 - Ex ante (as a condition of supply)
 - Obligation to enable rule extraction
 - Ex post (on appeal)



Getting it right

- Counterfactuals & abnormalities
 - Google's 'What-If' tool
- Knowing decisions have been made
 - Identifying the 'nudge'
- To whom?
 - Selective: relevance depends on audience
 - Transparency (means) ≠ intelligibility (ends)
- Protecting intellectual property
 - Trade secrets & open data
 - Investment climate: 'chilling effect'



Concluding remarks.....

Background Documentation

III BACKGROUND DOCUMENTATION

EU documents on Artificial Intelligence

1	White Paper on Artificial Intelligence - A European approach to excellence and trust (<i>Brussels, 19.2.2020 COM(2020) 65 final</i>)	1
2	Study requested by the LIBE Committee of the EP on Artificial Intelligence and the Law Enforcement (<i>July 2020</i>)	28
3	Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Building Trust in Human-Centric Artificial Intelligence (<i>Brussels, 8.4.2019 COM(2019) 168 final</i>)	120
4	European Commission's High-Level Expert Group on Artificial Intelligence - Ethics guidelines for trustworthy AI (<i>Brussels, 8 April 2019</i>)	132
5	Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Coordinated Plan on Artificial Intelligence (<i>COM(2018) 795 final, 7 December 2018</i>)	174
6	Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Artificial Intelligence for Europe (<i>SWD(2018) 137 final, 25 April 2018</i>)	198

CoE documents on Artificial Intelligence (PRINTED)

7	European Commission for the Efficiency of Justice (CEPEJ) - European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment (<i>Adopted at the 31st plenary meeting of the CEPEJ, Strasbourg, 3-4 December 2018</i>)	218

Other EU criminal justice documents with hyperlinks

A) The institutional framework for criminal justice in the EU

A1) Main treaties and conventions

A1-01	Protocol (No 36) on Transitional Provisions
A1-02	Statewatch Analysis, "The Third Pillar acquis" after the Treaty of Lisbon enters into force, Professor Steve Peers, University of Essex, Second Version, 1 December 2009
A1-03	Consolidated version of the Treaty on the functioning of the European Union, art. 82-86 (<i>OJ C 326/47; 26.10.2012</i>)
A1-04	Consolidated Version of the Treaty on the European Union, art. 9-20 (<i>OJ C326/13; 26.10.2012</i>)
A1-05	Charter of fundamental rights of the European Union (<i>OJ. C 364/1; 18.12.2000</i>)
A1-06	Explanations relating to the Charter of Fundamental Rights (<i>2007/C 303/02</i>)
A1-07	Convention implementing the Schengen Agreement of 14 June 1985 (<i>OJ L 239; 22.9.2000, P. 19</i>)

A2) Court of Justice of the European Union

A2-01	Consolidated Version of the Statute of the Court of Justice of the European Union (01 August 2016)
A2-02	Consolidated version of the Rules of Procedure of the Court of Justice (25 September 2012)

A3) European Convention on Human Rights (ECHR)

A3-01	Convention for the Protection of Human Rights and Fundamental Freedoms as amended by Protocols No. 11 and No. 14 together with additional protocols No. 4, 6, 7, 12 and 13, Council of Europe
A3-02	Case of Mihalache v. Romania [GC] (Application no. 54012/10), Strasbourg, 08 July 2019
A3-03	Case of Altay v. Turkey (no. 2) (Application no. 11236/09), Strasbourg, 09 April 2019
A3-04	Case Beuze v. Belgium (Application no. 71409/10), Strasbourg, 09 November 2018
A3-05	Case of Vizgirda v. Slovenia (Application no. 59868/08), Strasbourg, 28 August 2018
A3-06	Case of Şahin Alpay v. Turkey (Application no. 16538/17), Strasbourg, 20 March 2018
A3-07	Grand Chamber Hearing, Beuze v. Belgium [GC] (Application no. 71409/10), Strasbourg, 20 December 2017
A3-08	Case of Blokhin v. Russia (Application no. 47152/06), Judgment European Court of Human Rights, Strasbourg, 23 March 2016
A3-09	Case of A.T. v. Luxembourg (Application no. 30460/13), Judgment European Court of Human Rights, Strasbourg, 09 April 2015
A3-10	Case of Blaj v. Romania (Application no. 36259/04), Judgment European Court of Human Rights, Strasbourg, 08 April 2014
A3-11	Case of Boz v. Turkey (Application no. 7906/05), Judgment European Court of Human Rights, Strasbourg, 01 October 2013 (FR)
A3-12	Case of Pishchalnikov v. Russia (Application no. 7025/04), Judgment European Court of Human Rights, Strasbourg, 24 October 2009
A3-13	Case of Salduz v. Turkey (Application no. 36391/02), Judgment, European Court of Human Rights, Strasbourg, 27 November 2008

A4) Brexit

A4-01	Draft text of the Agreement on the New Partnership between the European Union and the United Kingdom (UKTF 2020-14), 18 March 2020
A4-02	Draft Working Text for an Agreement on Law enforcement and Judicial Cooperation in Criminal Matters
A4-03	The Law Enforcement and Security (Amendment) (EU Exit) Regulations 2019 (2019/742), 28th March 2019
A4-04	Brexit next steps: The European Arrest Warrant, House of Commons, 20 February 2020
A4-05	Brexit next steps: The Court of Justice of the EU and the UK, House of Commons, 7 February 2020
A4-06	The Law Society, "Brexit no deal: Criminal Justice Cooperation", London, September 2019
A4-07	European Commission, Factsheet, „A „No-deal“-Brexit: Police and judicial cooperation”, April 2019

A4-08	CEPS: Criminal Justice and Police Cooperation between the EU and the UK after Brexit: Towards a principled and trust-based partnership, 29 August 2018
A4-09	Policy paper: The future relationship between the United Kingdom and the European Union, 12 July 2018
A4-10	House of Lords, Library Briefing, Proposed UK-EU Security Treaty, London, 23 May 2018
A4-11	HM Government, Technical Note: Security, Law Enforcement and Criminal Justice, May 2018
A4-12	LSE-Blog, Why Britain's habit of cherry-picking criminal justice policy cannot survive Brexit, Auke Williams, London School of Economics and Political Science, 29 March 2018
A4-13	House of Commons, Home Affairs Committee, UK-EU Security Cooperation after Brexit, Fourth Report of Session 2017-19, London, 21 March 2018
A4-14	HM Government, Security, Law Enforcement and Criminal Justice, A future partnership paper
A4-15	European Criminal Law after Brexit, Queen Mary University London, Valsamis Mitsilegas, 2017
A4-16	House of Lords, European Union Committee, Brexit: Judicial oversight of the European Arrest Warrant, 6 th Report of Session 2017-19, London, 27 July 2017
A4-17	House of Commons, Brexit: implications for policing and criminal justice cooperation (24 February 2017)
A4-18	Scottish Parliament Information Centre, Briefing, Brexit: Impact on the Justice System in Scotland, Edinburgh, 27 October 2016

B) Mutual legal assistance

B1) Legal framework

B1-01	Council Act of 16 October 2001 establishing in accordance with Article 34 of the Treaty on European Union, the Protocol to the Convention on Mutual Assistance in Criminal Matters between the Member States of the European Union (2001/C 326/01), (OJ C 326/01; 21.11.2001,P. 1)
B1-02	Council Act of 29 May 2000 establishing in accordance with Article 34 of the Treaty on European Union the Convention on Mutual Assistance in Criminal Matters between the Member States of the European Union (OJ C 197/1; 12.7.2000, P. 1)
B1-03	Agreement between the European Union and the Republic of Iceland and the Kingdom of Norway on the surrender procedure between the Member States of the European Union and Iceland and Norway (OJ L 292, 21.10.2006, p. 2–19)
B1-04	Second Additional Protocol to the European Convention on Mutual Assistance in Criminal Matters (Strasbourg, 8.XI.2001)
B1-05	Additional Protocol to the European Convention on Mutual Assistance in Criminal Matters (Strasbourg, 17.III.1978)
B1-06	European Convention on Mutual Assistance in Criminal Matters (Strasbourg, 20.IV.1959)
B1-07	Third Additional Protocol to the European Convention on Extradition (Strasbourg, 10.XI.2010)
B1-08	Second Additional Protocol to the European Convention on Extradition (Strasbourg, 17.III.1978)
B1-09	Additional Protocol to the European Convention on Extradition (Strasbourg, 15.X.1975)

B1-10	European Convention on Extradition (<i>Strasbourg, 13.XII.1957</i>)

B2) Mutual recognition: the European Arrest Warrant

B2-01	Council Framework Decision 2009/299/JHA of 26 February 2009 amending Framework Decisions 2002/584/JHA, 2005/214/JHA, 2006/783/JHA, 2008/909/JHA and 2008/947/JHA, thereby enhancing the procedural rights of persons and fostering the application of the principle of mutual recognition to decisions rendered in the absence of the person concerned at the trial (<i>OJ L 81/24; 27.3.2009</i>)
B2-02	Council Framework Decision of 13 June 2002 on the European arrest warrant and the surrender procedures between Member States (<i>OJ L 190/1; 18.7.2002, P. 1</i>)
B2-03	Case law by the Court of Justice of the European Union on the European Arrest Warrant – Overview, Eurojust, 15 March 2020
B2-04	Case C-717/18, X (European arrest warrant – Double criminality) Judgement of the Court of 3 March 2020
B2-05	Case C-314/18, SF Judgement of the Court of 1 March 2020
B2-06	Joined Cases C-566/19 PPU (JR) and C-626/19 PPU (YC), Opinion of AG Campos Sánchez-Bordona, 26 November 2019
B2-07	Case C-489/19 PPU (NJ), Judgement of the Court (Second Chamber) of 09 October 2019
B2-08	Case 509/18 (PF), Judgement of the Court (Grand Chamber), 27 May 2019
B2-09	Joined Cases C-508/18 (OG) and C-82/19 PPU (PI), Judgement of the Court (Grand Chamber), 24 May 2019
B2-10	The Guardian Press Release: Dutch court blocks extradition of man to 'inhumane' UK prisons, 10 May 2019
B2-11	Case 551/18, IK, Judgement of the Court of 06 December 2018 (First Chamber)
B2-12	CJEU Press Release No 141/18, Judgement in Case C-207/16, Ministerio Fiscal, 2 October 2018
B2-13	CJEU Press Release No 135/18, Judgement in Case C-327/18 PPU RO, 19 September 2019
B2-14	Case C-268/17, AY, Judgement of the Court of 25 July 2018 (Fifth Chamber)
B2-15	Case C-220/18 PPU, ML, Judgement of the Court of 25 July 2018 (First Chamber)
B2-16	Case C-216/18 PPU, LM, Judgement of the Court of 25 July 2018 (Grand Chamber)
B2-17	InAbsentEAW, Background Report on the European Arrest Warrant - The Republic of Poland, Magdalena Jacyna, 01 July 2018
B2-18	Case C-571/17 PPU, Samet Ardic, Judgment of the court of 22 December 2017
B2-19	C-270/17 PPU, Tupikas, Judgment of the Court of 10 August 2017 (Fifth Chamber)
B2-20	Case C-271/17 PPU, Zdziasek, Judgment of the Court of 10 August 2017 (Fifth Chamber)
B2-21	Case C-579/15, Popławski, Judgement of the Court (Fifth Chamber), 29 June 2017
B2-22	Case C-640/15, Vilkas, Judgement of the Court (Third Chamber), 25 January 2017
B2-23	Case C-477/16 PPU, Kovalkovas, Judgement of the Court (Fourth Chamber), 10 November 2016
B2-24	Case C-452/16 PPU, Poltorak, Judgement of the Court (Fourth chamber), 10 November 2016

B2-25	Case C-453/16 PPU, Özçelik, Judgement of the Court (Fourth Chamber), 10 November 2016
B2-26	Case C-294/16 PPU, JZ v Śródmieście, Judgement of the Court (Fourth Chamber), 28 July 2016
B2-27	Case C241/15 Bob-Dogi, Judgment of the Court (Second Chamber) of 1 June 2016
B2-28	C-108/16 PPU Paweł Dworzecki, Judgment of the Court (Fourth Chamber) of 24 May 2016
B2-29	Cases C-404/15 Pál Aranyosi and C-659/15 PPU Robert Căldăraru, Judgment of 5 April 2016
B2-30	Case C-237/15 PPU Lanigan, Judgment of 16 July 2015 (Grand Chamber)
B2-31	Case C-168/13 PPU <i>Jeremy F / Premier ministre</i> , Judgment of the court (Second Chamber), 30 May 2013
B2-32	Case C-399/11 <i>Stefano Melloni v Ministerio Fiscal</i> , Judgment of 26 February 2013
B2-33	Case C-396/11 Ciprian Vasile Radu, Judgment of 29 January 2013
B2-34	C-261/09 Mantello, Judgment of 16 November 2010
B2-35	C-123/08 Wolzenburg, Judgment of 6 October 2009
B2-36	C-388/08 Leymann and Pustovarov, Judgment of 1 December 2008
B2-37	C-296/08 Goicoechea, Judgment of 12 August 2008
B2-38	C-66/08 Szymon Kozłowski, Judgment of 17 July 2008

B3) Mutual recognition: freezing and confiscation and asset recovery

B3-01	FATF, COVID-19-related Money Laundering and Terrorist Financing Risk and Policy Responses, Paris, 4 May 2020
B3-02	Money-Laundering and COVID-19: Profit and Loss, Vienna, 14 April 2020
B3-03	FATF President Statement – COVID-19 and measures to combat illicit financing, Paris 1 April 2020
B3-04	Moneyval Plenary Meeting report, Strasbourg, 31 January 2020
B3-05	Directive (EU) 2019/1153 of the European Parliament and of the Council of 20 June 2019, laying down rules facilitating the use of financial and other information for the prevention, detection, investigation or prosecution of certain criminal offences, and repealing Council Decision 2000/642/JHA
B3-06	Commission Delegated Regulation (EU) .../... of 13.2.2019 supplementing Directive (EU) 2015/849 of the European Parliament and of the Council by identifying high-risk third countries with strategic deficiencies, C(2019) 1326 final
B3-07	Regulation 2018/1805 of the European Parliament and of the Council on the mutual recognition of freezing and confiscation orders, L 303/1, Brussels, 14 November 2018
B3-08	Directive (EU) 2018/1673 of the European Parliament and of the Council of 23 October 2018 on combating money laundering by criminal law, L 284/22
B3-09	Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU (Text with EEA relevance), PE/72/2017/REV/1 OJ L 156, p. 43–74, 19 June 2018

B3-10	Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA
B3-11	Regulation (EU) 2016/1675 of 14 July 2016 supplementing Directive (EU) 2015/849 of the European Parliament and of the Council by identifying high-risk third countries with strategic deficiencies (Text with EEA relevance)
B3-12	Directive (EU) 2015/849 of the European Parliament and of the Council of 20 May 2015 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, and repealing Directive 2005/60/EC of the European Parliament and of the Council and Commission Directive 2006/70/EC (Text with EEA relevance)
B3-13	Regulation (EU) 2015/847 of the European Parliament and of the Council of 20 May 2015 on information accompanying transfers of funds and repealing Regulation (EC) No 1781/2006 (Text with EEA relevance)
B3-14	Regulation (EC) No 1889/2005 of the European Parliament and of the Council of 26 October 2005 on controls of cash entering or leaving the Community
B3-15	Council Framework Decision of 26 June 2001 on money laundering, the identification, tracing, freezing, seizing and confiscation of instrumentalities and the proceeds of crime (2001/500/JHA)
B3-16	Council Decision of 17 October 2000 concerning arrangements for cooperation between financial intelligence units of the Member States in respect of exchanging information (2000/642/JHA)

B4) Mutual recognition: Convictions

B4-01	Council Framework Decision 2009/829/JHA of 23 October 2009 on the application, between Member States of the European Union, of the principle of mutual recognition to decisions on supervision measures as an alternative to provisional detention (<i>OJ L 294/20; 11.11.2009</i>)
B4-02	Council Framework Decision 2008/947/JHA on the application of the principle of mutual recognition to judgments and probation decisions with a view to the supervision of probation measures and alternative sanctions (<i>OJ L 337/102; 16.12.2008</i>)
B4-03	Council Framework Decision 2008/909/JHA of 27 November 2008 on the application of the principle of mutual recognition to judgments in criminal matters imposing custodial sentences or measures involving deprivation of liberty for the purpose of their enforcement in the European Union (<i>OJ L 327/27; 5.12.2008</i>)
B4-04	Council Framework Decision 2008/675/JHA of 24 July 2008 on taking account of convictions in the Member States of the European Union in the course of new criminal proceedings (<i>OJ L 220/32; 15.08.2008</i>)
B4-05	Case C-234/18, Judgment of 20 March 2020
B4-06	Case C-390/16, Dániel Bertold Lada, Opinion of AG Bot, delivered on 06 February 2018
B4-07	Case C-171/16, Trayan Beshkov, Judgement of the Court (Fifth Chamber), 21 September 2017
B4-08	Case C-528/15, Policie ČR, Krajské ředitelství policie Ústeckého kraje, odbor cizinecké policie v Salah Al Chodor, Ajlin Al Chodor, Ajvar Al Chodor, Judgement of the Court (Second Chamber), 15 March 2017
B4-09	Case C-554/14, Ognyanov, Judgement of the Court (Grand Chamber), 8 November 2016
B4-10	Case C-439/16 PPU, Milev, Judgement of the Court (Fourth Chamber), 27 October 2016
B4-11	C-294/16 PPU, JZ v Śródmieście, Judgement of the Court (Fourth Chamber), 28 July 2016
B4-12	C-601/15 PPU, J. N. v Staatssecretaris voor Veiligheid en Justitie, Judgement of the Court (Grand Chamber), 15 February 2016
B4-13	C-474/13, Thi Ly Pham v Stadt Schweinfurt, Amt für Meldewesen und Statistik, Judgement of the Court (Grand Chamber), 17 July 2014
B4-14	Joined Cases C-473/13 and C-514/13, Bero and Bouzalmate, Judgement of the Court (Grand Chamber), 17 July 2014
B4-15	C-146/14 PPU, Bashir Mohamed Ali Mahdi, Judgement of the Court (Third Chamber), 5 June 2014
B4-16	Case C-383/13 PPU, M. G., N. R., Judgement of the Court (Second Chamber), 10 September 2013

B5) Mutual recognition in practice: evidence and e-evidence

B5-01	The European Law Blog, „E-Evidence: The way forward. Summary of a Workshop held in Brussels on 25 September 2019, Theodore Christakis, 06 November 2019
B5-02	Joint Note of Eurojust and the European Judicial Network on the Practical Application of the European Investigation Order, June 2019
B5-03	European Commission, Press Release, „Security Union: Commission recommends negotiating international rules for obtaining electronic evidence”, Brussels, 05 February 2019
B5-04	EURCRIM, “The European Commission’s Proposal on Cross Border Access to e-Evidence – Overview and Critical Remarks” by Stanislaw Tosza, Issue 4/2018, pp. 212-219
B5-05	Recommendation for a Council Decision authorising the opening of negotiations in view of an agreement between the European Union and the United States of America on cross-border access to electronic evidence for judicial cooperation in criminal matters, COM(2019) 70 final, Brussels, 05 February 2019
B5-06	Annex to the Recommendation for a Council Decision authorising the opening of negotiations in view of an agreement between the European Union and the United States of America on cross-border access to electronic evidence for judicial cooperation in criminal matters, COM(2019) 70 final, Brussels, 05 February 2019
B5-07	Fair Trials, Policy Brief, „The impact on the procedural rights of defendants of cross-border access to electronic data through judicial cooperation in criminal matters”, October 2018
B5-08	ECBA Opinion on European Commission Proposals for: (1) A Regulation on European Production and Preservation Orders for electronic evidence & (2) a Directive for harmonised rules on the appointment of legal representatives for the purpose of gathering evidence in criminal proceedings, Rapporteurs: Stefanie Schott (Germany), Julian Hayes (United Kingdom)
B5-09	Proposal for a Directive of the European Parliament and of the Council laying down harmonised rules on the appointment of legal representatives for the purpose of gathering evidence in criminal proceedings, COM(2018) 226 final, Strasbourg, 17 April 2018
B5-10	Proposal for a Regulation of the European Parliament and of the Council on European Production and Preservation Orders for electronic evidence in criminal matters, COM(2018) 225 final, Strasbourg, 17 April 2018
B5-11	Non-paper from the Commission services: Improving cross-border access to electronic evidence: Findings from the expert process and suggested way forward (8 June 2017)
B5-12	Non-paper: Progress Report following the Conclusions of the Council of the European Union on Improving Criminal Justice in Cyberspace (7 December 2016)
B5-13	ENISA 2014 - Electronic evidence - a basic guide for First Responders (Good practice material for CERT first responders)
B5-14	Directive 2014/41/EU of 3 April 2014 regarding the European Investigation Order in criminal matters (OJ L 130/1; 1.5.2014)
B5-15	Guidelines on Digital Forensic Procedures for OLAF Staff” (Ref. Ares(2013)3769761 - 19/12/2013, 1 January 2014
B5-16	ACPO Good Practice Guide for Digital Evidence (March 2012)
B5-17	Council Framework Decision 2008/978/JHA of 18 December 2008 on the European evidence warrant for the purpose of obtaining objects, documents

	and data for use in proceedings in criminal matters (<i>OJ L, 350/72, 30.12.2008</i>)
B5-18	Council Framework Decision 2003/577/JHA of 22 July 2003 on the execution in the European Union of orders freezing property or evidence (<i>OJ L 196/45; 2.8.2003</i>)
B5-19	Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market (Directive on electronic commerce) (<i>Official Journal L 178/1, 17.7.2000</i>)
B5-20	Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions ensuring security and trust in electronic communication - Towards a European Framework for Digital Signatures and Encryption (<i>COM (97) 503</i>), October 1997

B6) Criminal records, Interoperability

B6-01	Regulation (EU) 2019/816 of the European Parliament and of the Council of 17 April 2019 establishing a centralised system for the identification of Member States holding conviction information on third-country nationals and stateless persons (ECRIS-TCN) to supplement the European Criminal Records Information System and amending Regulation (EU) 2018/1726) (<i>OJ L135/85, 22.05.2019</i>)
B6-02	Regulation (EU) 2019/818 of the European Parliament and of the Council of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of police and judicial cooperation, asylum and migration and amending Regulations (EU) 2018/1726, (EU) 2018/1862 and (EU) 2019/816 (<i>OJ L 135/85, 22.05.2019</i>)
B6-03	Regulation (EU) 2019/817 of the European Parliament and of the Council of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of borders and visa and amending Regulations (EC) No 767/2008, (EU) 2016/399, (EU) 2017/2226, (EU) 2018/1240, (EU) 2018/1726 and (EU) 2018/1861 of the European Parliament and of the Council and Council Decisions 2004/512/EC and 2008/633/JHA (<i>OJ L 135/27, 22.05.2019</i>)
B6-04	Directive of the European Parliament and of the Council amending Council Framework Decision 2009/315/JHA, as regards the exchange of information on third-country nationals and as regards the European Criminal Records Information System (ECRIS), and replacing Council Decision 2009/316/JHA, PE-CONS 87/1/18, Strasbourg, 17 April 2019
B6-05	Council Framework Decision 2009/315/JHA of 26 February 2009 on the organisation and content of the exchange of information extracted from the criminal record between Member States (<i>OJ L 93/23; 07.4.2009</i>)
B6-06	Council Decision on the exchange of information extracted from criminal records – Manual of Procedure (<i>6397/5/06 REV 5; 15.1.2007</i>)
B6-07	Council Decision 2005/876/JHA of 21 November 2005 on the exchange of information extracted from the criminal record (<i>OJ L 322/33; 9.12.2005</i>)

B7) Conflicts of jurisdiction – *Ne bis in idem*

B7-01	Case law by the Court of Justice of the European Union on the principle of ne bis in idem in criminal matters, Eurojust, April 2020
B7-02	Council Framework Decision 2009/948/JHA of 30 November 2009 on prevention and settlement of conflicts of exercise of jurisdiction in criminal proceedings (OJ L 328/42; 15.12.2009, P.42)
B7-03	European Convention on the Transfer of Proceedings in Criminal Matters (Strasbourg, 15.V.1972)

C) Procedural guarantees in the EU

C-01	Directive (EU) 2016/1919 of the European Parliament and of the Council of 26 October 2016 on legal aid for suspects and accused persons in criminal proceedings and for requested persons in European arrest warrant proceedings (OJ L 297/1, 4.11.2016)
C-02	Directive (EU) 2016/800 of the European Parliament and of the Council of 11 May 2016 on procedural safeguards for children who are suspects or accused persons in criminal proceedings (OJ L 132 1; 21.5.2016)
C-03	Directive 2016/343 of 9 March 2016 on the strengthening of certain aspects of the presumption of innocence and of the right to be present at the trial in criminal proceedings (11.3.2016; OJ L 65/1)
C-04	Directive 2013/48/EU of 22 October 2013 on the right of access to a lawyer in criminal proceedings and in European arrest warrant proceedings, and on the right to have a third party informed upon deprivation of liberty and to communicate with third persons and with consular authorities while deprived of liberty (OJ L 294/1; 6.11.2013)
C-05	Directive 2012/13/EU of the European Parliament and of the Council of 22 May 2012 on the right to information in criminal proceedings (1.6.2012; OJ L 142/1)
C-06	Directive 2010/64/EU of the European Parliament and of the Council of 20 October 2010 on the right to interpretation and translation in criminal proceedings (OJ L 280/1; 26.10.2010)
C-07	Case C-659/18, Judgement of the Court of 2 March 2020
C-08	Case C-688/18, Judgement of the Court of 3 February 2020
C-09	Case C-467/18, Rayonna prokuratura Lom, Judgment of the Court of 19 September 2019
C-10	Case C-467/18 on directive 2013/48/EU on the right of access to a lawyer in criminal proceedings, EP, Judgement of the court (Third Chamber), 19. September 2019
C-11	Case C-377/18, AH a. o., Judgment of the Court of 05 September 2019
C-12	Case C-646/17 on directive 2012/13/EU on the right to information in criminal proceedings, Gianluca Moro, Judgement of the Court (First Chamber), 13 June 2019
C-13	Case C-8/19 PPU, criminal proceedings against RH (presumption of innocence), Decision of the Court (First Chamber), 12. February 2019
C-14	Case C-646/17, Gianluca Moro, Opinion of the AG Bobek, 05 February 2019
C-15	Case C-551/18 PPU, IK, Judgment of the Court (First Chamber), 6 December 2018
C-16	Case C-327/18 PPU, RO, Judgment of 19 September 2018 (First Chamber)
C-17	Case C-268/17, AY, Judgment of the Court (Fifth Chamber), 25 July 2018
C-18	Case C-216/18 PPU, LM, Judgment of 25 July 2018 (Grand Chamber)

C-19	Joined Cases C-124/16, C-188/16 and C-213/16 on Directive 2012/13/EU on the right to information in criminal proceedings Ianos Tranca, Tanja Reiter and Ionel Opria, Judgment of 22 March 2017 (Fifth Chamber)
C-20	Case C-439/16 PPU, Emil Milev (presumption of innocence), Judgment of the Court (Fourth Chamber), 27 October 2016
C-21	Case C-278/16 Frank Sleutjes ("essential document" under Article 3 of Directive 2010/64), Judgment of 12 October 2017 (Fifth Chamber)
C-22	C-25/15, István Balogh, Judgment of 9 June 2016 (Fifth Chamber)
C-23	Opinion of Advocate General Sharpston, delivered on 10 March 2016, Case C-543/14
C-24	C-216/14 Covaci, Judgment of 15 October 2015 (First Chamber)

D) Approximating criminal law and Victims' Rights

D1) Terrorism

D1-01	Terrorism Situation and Trend Report (TE-SAT) 2019
D1-02	Communication from the Commission to the European Parliament, the European Council and the Council, Twentieth Progress Report towards an effective and genuine Security Union, COM(2019) 552 final, Brussels, 30 October 2019
D1-03	Communication from the Commission to the European Parliament, and the Council, Towards better Implementation of the EU's anti-money laundering and countering the financing of terrorism framework, COM(2019) 360 final, Brussels, 24 July 2019
D1-04	Directive (EU) 2019/713 of the European Parliament and of the Council of 17 April 2019 on combating fraud and counterfeiting of non-cash means of payment and replacing Council Framework Decision 2001/413/JHA, L 123/18
D1-05	Commission Delegated Regulation (EU) 2019/758 of 31 January 2019 amending Directive (EU) 2015/849 of the European Parliament and of the Council with regard to regulatory technical standards for the minimum action and the type of additional measures credit and financial institutions must take to mitigate money laundering and terrorist financing risk in certain third countries, L 125/4 (Text with EEA relevance)
D1-06	Council Decision (CFSP) 2019/25 of 08 January 2019 updating the list of persons, groups and entities subject to Articles 2, 3 and 4 of Common Position 2001/931/CFSP on the application of specific measures to combat terrorism and repealing Decision (CFSP) 2016/1136, Brussels, 08 January 2019
D1-07	Proposal for a Regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, Brussels, 12.9.2018, COM(2018) 640 final
D1-08	Regulation (EU) 2017/2226 of the European Parliament and of the Council of 30 November 2017 establishing an Entry/Exit System (EES) to register entry and exit data and refusal of entry data of third-country nationals crossing the external borders of the Member States and determining the conditions for access to the EES for law enforcement purposes, and amending the Convention implementing the Schengen Agreement and Regulations (EC) No 767/2008 and (EU) No 1077/2011 (OJ L 327/20; 9.12.2017)
D1-09	Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework

	Decision 2002/475/JHA and amending Council Decision 2005/671/JHA (OJ L 88/6)
D1-10	Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime (OJ L 119/132; 4.5.2016)

D2) Trafficking in Human Beings, Migrant Smuggling and Sexual Exploitation of Children

D2-01	Regulation of the European Parliament and of the Council amending Regulation (EC) No 810/2009 establishing a Community Code on Visas (Visa Code), PE-CONS 29/19, Brussels, 15 May 2019
D2-02	European Migrant Smuggling Centre – 4th Annual Activity Report, The Hague, 15 May 2020
D2-03	Report from the European Commission to the European Parliament and the Council, Second report on the progress made in the fight against trafficking in human beings (2018) as required under Article 20 of Directive 2011/36/EU on preventing and combating trafficking in human beings and protecting its victims, COM(2018) 777 final, Brussels, 03 December 2018
D2-04	UNODC – Global Study on Smuggling of Migrants 2018, Vienna/New York, June 2018
D2-05	Council Conclusions on setting the EU's priorities for the fight against organised and serious international crime between 2018 and 2021, Brussels, 9450/17, 19 May 2017
D2-06	Directive 2011/36/EU of 5 April 2011 on preventing and combating trafficking in human beings and protecting its victims, and replacing Council Framework Decision 2002/629/JHA

D3) Cybercrime

D3-01	Internet Organised Crime Threat Assessment (IOCTA) 2019
D3-02	Special Eurobarometer 480, Report, "Europeans' Attitudes towards Internet Security", Brussels, March 2019
D3-03	Directive 2013/40/EU of the European Parliament and of the Council of 12 August 2013 on attacks against information systems and replacing Council Framework Decision 2005/222/JHA (Official Journal L 218/8 of 14.08.2013)
D3-04	Directive of the European Parliament and of the Council on combating the sexual abuse, sexual exploitation of children and child pornography, repealing Framework Decision 2004/68/JHA (OJ L 335; 17.12.2011)
D3-05	Council Framework Decision 2005/222/JHA of 24 February 2005 on attacks against information systems (OJ L 69/67; 16.3.2005)
D3-06	Council Framework Decision 2004/68/JHA of 22 December 2003 on combating the sexual exploitation of children and child pornography (OJ L 13/44; 20.1.2004)
D3-07	Additional Protocol to the Convention on cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems (Strasbourg, 28.I.2003)
D3-08	Convention on Cybercrime (Budapest, 23.XI.2001)

D4) Protecting Victims' Rights

D4-01	European Commission, Executive Summary of the Report on strengthening Victims' Rights: From Compensation to Reparation – For a new EU Victims' Rights Strategy 2020-2025, Report of the Special Adviser Joëlle Milquet to the President of the European Commission, Brussels, 11 March 2019
D4-02	Regulation (EU) No 606/2013 of the European Parliament and of the Council of 12 June 2013 on mutual recognition of protection measures in civil matters
D4-03	European Commission, DG Justice Guidance Document related to the transposition and implementation of Directive 2012/29/EU of the European Parliament and of the Council of 25 October 2012 establishing minimum standards on the rights, support and protection of victims of crime, and replacing Council Framework Decision 2001/220/JHA
D4-04	Directive 2012/29/EU of the European Parliament and of the Council of 25 October 2012 establishing minimum standards on the rights, support and protection of victims of crime, and replacing Council Framework Decision 2001/220/JHA
D4-05	Directive 2011/99/EU of the European Parliament and of the Council of 13 December 2011 on the European protection order
D4-06	Council Directive 2004/80/EC of 29 April 2004 relating to compensation to crime victims
D4-07	Website of the European Union Agency for Fundamental Rights (FRA) – Victims' rights
D4-08	Victim Support Europe

E) Criminal justice bodies and networks

E1) European Judicial Network

E1-01	European Judicial Network, Report on Activities and Management 2017-2018
E1-02	Council Decision 2008/976/JHA of 16 December 2008 on the European Judicial Network (<i>OJ L 348/130, 24.12.2008, P. 130</i>)

E2) Eurojust

E2-01	Eurojust quarterly newsletter
E2-02	Eurojust Guidelines on Jurisdiction
E2-03	Eurojust Annual Report 2019
E2-04	Guidelines for deciding on competing requests for surrender and extradition, October 2019
E2-05	Regulation (EU) 2018/1727 of the European Parliament and of the Council of 14 November 2018 on the European Union Agency for Criminal Justice Cooperation (Eurojust), and replacing and repealing Council Decision 2002/187/JHA

E3) Europol

E3-01	Europol Report – Beyond the Pandemic – How COVID-19 will shape the serious and organised crime landscape in the EU, 30 April 2020
E3-02	Regulation (EU) 2015/2219 of the European Parliament and of the Council of 25 November 2015 on the European Union Agency for Law Enforcement Training (CEPOL) and replacing and repealing Council Decision 2005/681/JHA

E4) European Public Prosecutor's Office

E4-01	Decision 2019/1798 of the European Parliament and of the Council of 14 October 2019 appointing the European Chief Prosecutor of the European Public Prosecutor's Office (<i>OJ L 274/1, 28.10.2019</i>)
E4-02	Opinion on the proposal for a regulation of the European Parliament and of the Council amending Regulation (EU, Euratom) No 883/2013 concerning investigations conducted by the European Anti-Fraud Office (OLAF) as regards cooperation with the European Public Prosecutor's Office and the effectiveness of OLAF investigations Committee on Civil Liberties, Justice and Home Affairs, Rapporteur for opinion: Monica Macovei, 11.1.2019
E4-03	German Judges' Association: Opinion on the European Commission's initiative to extend the jurisdiction of the European Public Prosecutor's Office to include cross-border terrorist offences, December 2018 (only available in German)
E4-04	Communication from the Commission to the European Parliament and the European Council: A Europe that protects: an initiative to extend the competences of the European Public Prosecutor's Office to cross-border terrorist crimes, Brussels, 12.9.2018, COM(2018) 641 final
E4-05	Annex to the Communication from the Commission to the European Parliament and the European Council: A Europe that protects: an initiative to extend the competences of the European Public Prosecutor's Office to cross-border terrorist crimes, Brussels, 12.9.2018, COM (2018) 641 final
E4-06	Council Implementing Decision (EU) 2018/1696 of 13 July 2018 on the operating rules of the selection panel provided for in Article 14(3) of Regulation (EU) 2017/1939 implementing Enhanced cooperation on the establishment of the European Public Prosecutor's Office ('the EPPO')
E4-07	Annex to the Proposal for a Council Implementing Decision on the operating rules of the selection panel provided for in Article 14(3) of Regulation (EU) 2017/1939 implementing enhanced cooperation on the establishment of the European Public Prosecutor's Office ("the EPPO"), Brussels, 25.5.2018, COM(2018) 318 final)
E4-08	Council Regulation (EU) 2017/1939 of 12 October 2017 implementing enhanced cooperation on the establishment of the European Public Prosecutor's Office ('the EPPO')

F) Data Protection

F-01	Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA (4.5.2016; OJ L 119/89)

G) Police Cooperation in the EU

G1) General

G1-01	European Commission, Press Release, „Commission marks ten years of judicial and police cooperation between between Member States of the European Union“, 01 December 2019
G1-02	Regulation of the European Parliament and of the Council on establishing a framework of interoperability between EU information systems in the field of police and judicial cooperation, asylum and migration and amending Regulations (EU) 2018/1726 and (EU) 2018/1862 and (EU) 2019/816 [the ECRIS-TCN Regulation], PE-CONS 31/19, Brussels, 2 May 2019
G1-03	Regulation (EU) 2018/1862 of the European Parliament and of the Council of 28 November 2018 on the establishment, operation and use of the Schengen Information System (SIS) in the field of police cooperation and judicial cooperation in criminal matters, amending and repealing Council Decision 2007/533/JHA, and repealing Regulation (EC) No 1986/2006 of the European Parliament and of the Council and Commission Decision 2010/261/EU
G1-04	Council Decision 2008/616/JHA of 23 June 2008 on the implementation of Decision 2008/615/JHA on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime (OJ L 210/12; 06.08.2008)
G1-05	Council Decision 2008/615/JHA of 23 June 2008 on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime (OJ L 210/1; 06.08.2008)
G1-06	Council Framework Decision of 18 December 2006 on simplifying the exchange of information and intelligence between law enforcement authorities of the Member States of the European Union (OJ L 386/89; 29.12.2006, P. 89)
G1-07	Convention on the stepping up of cross-border cooperation, particularly in combating terrorism, cross-border crime and illegal migration of 27. May 2005 (10900/05; 27.5.2005)

G2) Joint Investigation Teams (JITs)

G2-01	Eurojust Information on JITs
G2-02	Third JIT Evaluation Report, Eurojust, March 2020
G2-03	Joint Investigation Teams Practical Guide (Brussels, 14 February 2017; 6128/1/17)
G2-04	Council Resolution on a Model Agreement for Setting up a Joint Investigation Team (JIT) – 2017/C18/01, Strasbourg, 19 January 2017
G2-05	Council Framework Decision of 13 June 2002 on joint investigation teams (OJ L 162/1; 20.6.2002)



Brussels, 19.2.2020
COM(2020) 65 final

WHITE PAPER

On Artificial Intelligence - A European approach to excellence and trust

White Paper on Artificial Intelligence

A European approach to excellence and trust

Artificial Intelligence is developing fast. It will change our lives by improving healthcare (e.g. making diagnosis more precise, enabling better prevention of diseases), increasing the efficiency of farming, contributing to climate change mitigation and adaptation, improving the efficiency of production systems through predictive maintenance, increasing the security of Europeans, and in many other ways that we can only begin to imagine. At the same time, Artificial Intelligence (AI) entails a number of potential risks, such as opaque decision-making, gender-based or other kinds of discrimination, intrusion in our private lives or being used for criminal purposes.

Against a background of fierce global competition, a solid European approach is needed, building on the European strategy for AI presented in April 2018¹. To address the opportunities and challenges of AI, the EU must act as one and define its own way, based on European values, to promote the development and deployment of AI.

The Commission is committed to enabling scientific breakthrough, to preserving the EU's technological leadership and to ensuring that new technologies are at the service of all Europeans – improving their lives while respecting their rights.

Commission President Ursula von der Leyen announced in her political Guidelines² a coordinated European approach on the human and ethical implications of AI as well as a reflection on the better use of big data for innovation.

Thus, the Commission supports a regulatory and investment oriented approach with the twin objective of promoting the uptake of AI and of addressing the risks associated with certain uses of this new technology. The purpose of this White Paper is to set out policy options on how to achieve these objectives. It does not address the development and use of AI for military purposes. The Commission invites Member States, other European institutions, and all stakeholders, including industry, social partners, civil society organisations, researchers, the public in general and any interested party, to react to the options below and to contribute to the Commission's future decision-making in this domain.

1. INTRODUCTION

As digital technology becomes an ever more central part of every aspect of people's lives, people should be able to trust it. Trustworthiness is also a prerequisite for its uptake. This is a chance for Europe, given its strong attachment to values and the rule of law as well as its proven capacity to build safe, reliable and sophisticated products and services from aeronautics to energy, automotive and medical equipment.

Europe's current and future sustainable economic growth and societal wellbeing increasingly draws on value created by data. AI is one of the most important applications of the data economy. Today most data are related to consumers and are stored and processed on central cloud-based infrastructure. By contrast a large share of tomorrow's far more abundant data will come from industry, business and the public sector, and will be stored on a variety of systems, notably on computing devices working at the edge of the network. This opens up new opportunities for Europe, which has a strong position in

¹ AI for Europe, COM/2018/237 final

² https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf.

digitised industry and business-to-business applications, but a relatively weak position in consumer platforms.

Simply put, AI is a collection of technologies that combine data, algorithms and computing power. Advances in computing and the increasing availability of data are therefore key drivers of the current upsurge of AI. Europe can combine its technological and industrial strengths with a high-quality digital infrastructure and a regulatory framework based on its fundamental values to **become a global leader in innovation in the data economy and its applications** as set out in the European data strategy³. On that basis, it can develop an AI ecosystem that brings the benefits of the technology to the whole of European society and economy:

- for **citizens** to reap new benefits for example improved health care, fewer breakdowns of household machinery, safer and cleaner transport systems, better public services;
- for **business** development, for example a new generation of products and services in areas where Europe is particularly strong (machinery, transport, cybersecurity, farming, the green and circular economy, healthcare and high-value added sectors like fashion and tourism); and
- for services of **public interest**, for example by reducing the costs of providing services (transport, education, energy and waste management), by improving the sustainability of products⁴ and by equipping law enforcement authorities with appropriate tools to ensure the security of citizens⁵, with proper safeguards to respect their rights and freedoms.

Given the major impact that AI can have on our society and the need to build trust, it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection.

Furthermore, the impact of AI systems should be considered not only from an individual perspective, but also from the perspective of society as a whole. The use of AI systems can have a significant role in achieving the Sustainable Development Goals, and in supporting the democratic process and social rights. With its recent proposals on the European Green Deal⁶, Europe is leading the way in tackling climate and environmental-related challenges. Digital technologies such as AI are a critical enabler for attaining the goals of the Green Deal. Given the increasing importance of AI, the environmental impact of AI systems needs to be duly considered throughout their lifecycle and across the entire supply chain, e.g. as regards resource usage for the training of algorithms and the storage of data.

A common European approach to AI is necessary to reach sufficient scale and avoid the fragmentation of the single market. The introduction of national initiatives risks to endanger legal certainty, to weaken citizens' trust and to prevent the emergence of a dynamic European industry.

This White Paper presents policy options to enable a trustworthy and secure development of AI in Europe, in full respect of the values and rights of EU citizens. The main building blocks of this White Paper are:

³ COM(2020) 66 final.

⁴ AI and digitalisation in general are critical enablers of Europe's Green deal ambitions. However, the current environmental footprint of the ICT sector is estimated at more than 2% of all global emissions. The European digital strategy accompanying this White Paper proposes green transformation measures for digital.

⁵ AI tools can provide an opportunity for better protecting EU citizens from crime and acts of terrorism. Such tools could, for example, help identify online terrorist propaganda, discover suspicious transactions in the sales of dangerous products, identify dangerous hidden objects or illicit substances or products, offer assistance to citizens in emergencies and help guide first responders.

⁶ COM(2019) 640 final.

- The policy framework setting out measures to align efforts at European, national and regional level. In partnership between the private and the public sector, the aim of the framework is to mobilise resources to achieve an **‘ecosystem of excellence’** along the entire value chain, starting in research and innovation, and to create the right incentives to accelerate the adoption of solutions based on AI, including by small and medium-sized enterprises (SMEs).
- The key elements of a future regulatory framework for AI in Europe that will create a unique **‘ecosystem of trust’**. To do so, it must ensure compliance with EU rules, including the rules protecting fundamental rights and consumers’ rights, in particular for AI systems operated in the EU that pose a high risk⁷. Building an ecosystem of trust is a policy objective in itself, and should give citizens the confidence to take up AI applications and give companies and public organisations the legal certainty to innovate using AI. The Commission strongly supports a human-centric approach based on the Communication on Building Trust in Human-Centric AI⁸ and will also take into account the input obtained during the piloting phase of the Ethics Guidelines prepared by the High-Level Expert Group on AI.

The European strategy for data, which accompanies this White Paper, aims to enable Europe to become the most attractive, secure and dynamic data-agile economy in the world – empowering Europe with data to improve decisions and better the lives of all its citizens. The strategy sets out a number of policy measures, including mobilising private and public investments, needed to achieve this goal. Finally, the implications of AI, Internet of Things and other digital technologies for safety and liability legislation are analysed in the Commission Report accompanying this White Paper.

2. CAPITALISING ON STRENGTHS IN INDUSTRIAL AND PROFESSIONAL MARKETS

Europe is well placed to benefit from the potential of AI, not only as a user but also as a creator and a producer of this technology. It has excellent research centres, innovative start-ups, a world-leading position in robotics and competitive manufacturing and services sectors, from automotive to healthcare, energy, financial services and agriculture. Europe has developed a strong computing infrastructure (e.g. high-performance computers), essential to the functioning of AI. Europe also holds large volumes of public and industrial data, the potential of which is currently under-used. It has well-recognised industrial strengths in safe and secure digital systems with low-power consumption that are essential for the further development of AI.

Harnessing the capacity of the EU to invest in next generation technologies and infrastructures, as well as in digital competences like data literacy, will increase Europe’s technological sovereignty in key enabling technologies and infrastructures for the data economy. The infrastructures should support the creation of European data pools enabling trustworthy AI, e.g. AI based on European values and rules.

Europe should leverage its strengths to expand its position in the ecosystems and along the value chain, from certain hardware manufacturing sectors to software all the way to services. This is already happening to an extent. Europe produces more than a quarter of all industrial and professional service robots (e.g. for precision farming, security, health, logistics.), and plays an important role in developing and using software applications for companies and organisations (business-to-business applications such as Enterprise Resource Planning, design and engineering software) as well as applications to support e-government and the "intelligent enterprise".

⁷ Although further arrangements may need to be put in place to prevent and counter misuse of AI for criminal purposes, this is outside the scope of this white paper.

⁸ COM(2019) 168.

Europe leads the way in deploying AI in manufacturing. Over half of the top manufacturers implement at least one instance of AI in manufacturing operations⁹.

One reason for Europe's strong position in terms of research is the EU funding programme that has proven instrumental in pooling action, avoiding duplications, and leveraging public and private investments in the Member States. Over the past three years, EU funding for research and innovation for AI has risen to €1.5 billion, i.e. a 70% increase compared to the previous period.

However, investment in research and innovation in Europe is still a fraction of the public and private investment in other regions of the world. Some €3.2 billion were invested in AI in Europe in 2016, compared to around €12.1 billion in North America and €6.5 billion in Asia¹⁰. In response, Europe needs to increase its investment levels significantly. The Coordinated plan on AI¹¹ developed with Member States is proving to be a good starting point in building closer cooperation on AI in Europe and in creating synergies to maximise investment in the AI value chain.

3. SEIZING THE OPPORTUNITIES AHEAD: THE NEXT DATA WAVE

Although Europe currently is in a weaker position in consumer applications and on online platforms, which results in a competitive disadvantage in data access, major shifts in the value and re-use of data across sectors are underway. The volume of data produced in the world is growing rapidly, from 33 zettabytes in 2018 to an expected 175 zettabytes in 2025¹². Each new wave of data brings opportunities for Europe to position itself in the data-agile economy and to become a world leader in this area. Furthermore, the way in which data are stored and processed will change dramatically over the coming five years. Today 80% of data processing and analysis that takes place in the cloud occurs in data centres and centralised computing facilities, and 20% in smart connected objects, such as cars, home appliances or manufacturing robots, and in computing facilities close to the user ("edge computing"). By 2025 these proportions are set to change markedly¹³.

Europe is a global leader in low-power electronics which is key for the next generation of specialised processors for AI. This market is currently dominated by non-EU players. This could change with the help of initiatives such as the European Processor Initiative, which focuses on developing low-power computing systems for both edge and next generation high-performance computing, and the work of the Key Digital Technology Joint Undertaking, proposed to start in 2021. Europe also leads in neuromorphic solutions¹⁴ that are ideally suited to automating industrial processes (industry 4.0) and transport modes. They can improve energy efficiency by several orders of magnitude.

Recent advances in quantum computing will generate exponential increases in processing capacity¹⁵. Europe can be at the forefront of this technology thanks to its academic strengths in quantum computing, as well as European industry's strong position in quantum simulators and programming environments for quantum computing. European initiatives that aim to increase the availability of quantum testing and experimentation facilities will help apply these new quantum solutions to a number of industrial and academic sectors.

⁹ Followed by Japan (30%) and the US (28%). Source: CapGemini (2019).

¹⁰ 10 imperatives for Europe in the age of AI and automation, McKinsey (2017).

¹¹ COM(2018) 795.

¹² IDC (2019).

¹³ Gartner (2017).

¹⁴ Neuromorphic solutions means any very large-scale system of integrated circuits that mimic neuro-biological architectures present in the nervous system.

¹⁵ Quantum computers will have the capacity to process in less than seconds many fold larger data sets than today's highest performance computers allowing for the development of new AI applications across sectors.

In parallel, Europe will continue to lead progress in the algorithmic foundations of AI, building on its own scientific excellence. There is a need to build bridges between disciplines that currently work separately, such as machine learning and deep learning (characterised by limited interpretability, the need for a large volume of data to train the models and learn through correlations) and symbolic approaches (where rules are created through human intervention). Combining symbolic reasoning with deep neural networks may help us improve explainability of AI outcomes.

4. AN ECOSYSTEM OF EXCELLENCE

To build an ecosystem of excellence that can support the development and uptake of AI across the EU economy and public administration, there is a need to step up action at multiple levels.

A. WORKING WITH MEMBER STATES

Delivering on its strategy on AI adopted in April 2018,¹⁶ in December 2018 the Commission presented a Coordinated Plan - prepared together with the Member States - to foster the development and use of AI in Europe¹⁷.

This plan proposes some 70 joint actions for closer and more efficient cooperation between Member States, and the Commission in key areas, such as research, investment, market uptake, skills and talent, data and international cooperation. The plan is scheduled to run until 2027, with regular monitoring and review.

The aim is to maximise the impact of investment in research, innovation and deployment, assess national AI strategies and build on and extend the Coordinated Plan on AI with Member States:

- *Action 1: The Commission, taking into account the results of the public consultation on the White Paper, will propose to the Member States a revision of the Coordinated Plan to be adopted by end 2020*

EU-level funding in AI should attract and pool investment in areas where the action required goes beyond what any single Member State can achieve. The objective is to attract over €20 billion¹⁸ of total investment in the EU per year in AI over the next decade. To stimulate private and public investment, the EU will make available resources from the Digital Europe Programme, Horizon Europe as well as from the European Structural and Investment Funds to address the needs of less-developed regions as well as rural areas.

The Coordinated Plan could also address societal and environmental well-being as a key principle for AI. AI systems promise to help tackling the most pressing concerns, including climate change and environmental degradation. It is also important that this happens in an environmentally friendly manner. AI can and should itself critically examine resource usage and energy consumption and be trained to make choices that are positive for the environment. The Commission will consider options to encourage and promote AI solutions that do this together with the Member States.

B. FOCUSING THE EFFORTS OF THE RESEARCH AND INNOVATION COMMUNITY

¹⁶ Artificial Intelligence for Europe, COM(2018) 237.

¹⁷ Coordinated Plan on Artificial Intelligence, COM(2018) 795.

¹⁸ COM(2018) 237.

Europe cannot afford to maintain the current fragmented landscape of centres of competence with none reaching the scale necessary to compete with the leading institutes globally. It is imperative to create more synergies and networks between the multiple European research centres on AI and to align their efforts to improve excellence, retain and attract the best researchers and develop the best technology. Europe needs a lighthouse centre of research, innovation and expertise that would coordinate these efforts and be a world reference of excellence in AI and that can attract investments and the best talents in the field.

The centres and the networks should concentrate in sectors where Europe has the potential to become a global champion such as industry, health, transport, finance, agrifood value chains, energy/environment, forestry, earth observation and space. In all these domains, the race for global leadership is ongoing, and Europe offers significant potential, knowledge and expertise¹⁹. Equally important is to create testing and experimentation sites to support the development and subsequent deployment of novel AI applications.

- *Action 2: the Commission will facilitate the creation of excellence and testing centres that can combine European, national and private investments, possibly including a new legal instrument. The Commission has proposed an ambitious and dedicated amount to support world reference testing centres in Europe under the Digital Europe Programme and complemented where appropriate by research and innovation actions of Horizon Europe as part of the Multiannual Financial Framework for 2021 to 2027.*

C. SKILLS

The European approach to AI will need to be underpinned by a strong focus on skills to fill competence shortages.²⁰ The Commission will soon present a reinforcement of the Skills Agenda, which aims to ensure that everyone in Europe can benefit from the green and digital transformations of the EU economy. Initiatives could also include the support of sectoral regulators to enhance their AI skills in order to effectively and efficiently implement relevant rules. The updated Digital Education Action Plan will help make better use of data and AI-based technologies such as learning and predictive analytics with the aim to improve education and training systems and make them fit for the digital age. The Plan will also increase awareness of AI at all levels of education in order to prepare citizens for informed decisions that will be increasingly affected by AI.

Developing the skills necessary to work in AI and upskilling the workforce to become fit for the AI-led transformation will be a priority of the revised Coordinated Plan on AI to be developed with Member States. This could include transforming the assessment list of the ethical guidelines into an indicative “curriculum” for developers of AI that will be made available as a resource for training institutions. Particular efforts should be undertaken to increase the number of women trained and employed in this area.

In addition, a lighthouse centre of research and innovation for AI in Europe would attract talent from all over the world due to the possibilities it could offer. It would also develop and spread excellence in skills that take root and grow across Europe.

¹⁹ The future European Defence Fund and Permanent Structured Cooperation (PESCO) will also provide opportunities for research and development in AI. These projects should be synchronized with the wider EU civilian programmes devoted to AI.

²⁰ <https://ec.europa.eu/jrc/en/publication/academic-offer-and-demand-advanced-profiles-eu>

- *Action 3: Establish and support through the advanced skills pillar of the Digital Europe Programme networks of leading universities and higher education institutes to attract the best professors and scientists and offer world-leading masters programmes in AI.*

Beyond upskilling, workers and employers are directly affected by the design and use of AI systems in the workplace. The involvement of social partners will be a crucial factor in ensuring a human-centred approach to AI at work.

D. FOCUS ON SMEs

It will also be important to ensure that SMEs can access and use AI. To this end, the Digital Innovation Hubs²¹ and the AI-on-demand platform²² should be strengthened further and foster collaboration between SMEs. The Digital Europe Programme will be instrumental in achieving this. While all Digital Innovation Hubs should provide support to SMEs to understand and adopt AI, it will be important that at least one innovation hub per Member State has a high degree of specialisation in AI.

SMEs and start-ups will need access to finance in order to adapt their processes or to innovate using AI. Building on the forthcoming pilot investment fund of €100 million in AI and blockchain, the Commission plans to further scale up access to finance in AI under InvestEU²³. AI is explicitly mentioned among the eligible areas for the use of the InvestEU guarantee.

- *Action 4: the Commission will work with Member States to ensure that at least one digital innovation hub per Member State has a high degree of specialisation on AI. Digital Innovation Hubs can be supported under the Digital Europe Programme.*
- *The Commission and the European Investment Fund will launch a pilot scheme of €100 million in Q1 2020 to provide equity financing for innovative developments in AI. Subject to final agreement on the MFF, the Commission's intention is to scale it up significantly from 2021 through InvestEU.*

E. PARTNERSHIP WITH THE PRIVATE SECTOR

It is also essential to make sure that the private sector is fully involved in setting the research and innovation agenda and provides the necessary level of co-investment. This requires setting up a broad-based public private partnership, and securing the commitment of the top management of companies.

- *Action 5: In the context of Horizon Europe, the Commission will set up a new public private partnership in AI, data and robotics to combine efforts, ensure coordination of research and innovation in AI, collaborate with other public-private partnerships in Horizon Europe and work together with the testing facilities and the Digital Innovation Hubs mentioned above.*

²¹ ec.europa.eu/digital-single-market/en/news/digital-innovation-hubs-helping-companies-across-economy-make-most-digital-opportunities.

²² www.Ai4eu.eu.

²³ Europe.eu/investeu.

F. PROMOTING THE ADOPTION OF AI BY THE PUBLIC SECTOR

It is essential that public administrations, hospitals, utility and transport services, financial supervisors, and other areas of public interest rapidly begin to deploy products and services that rely on AI in their activities. A specific focus will be in the areas of healthcare and transport where technology is mature for large-scale deployment.

- *Action 6: The Commission will initiate open and transparent sector dialogues giving priority to healthcare, rural administrations and public service operators in order to present an action plan to facilitate development, experimentation and adoption. The sector dialogues will be used to prepare a specific 'Adopt AI programme' that will support public procurement of AI systems, and help to transform public procurement processes themselves.*

G. SECURING ACCESS TO DATA AND COMPUTING INFRASTRUCTURES

The areas for action set out in this White Paper are complementary to the plan presented in parallel under the European data strategy. Improving access to and the management of data is fundamental. Without data, the development of AI and other digital applications is not possible. The enormous volume of new data yet to be generated constitutes an opportunity for Europe to position itself at the forefront of the data and AI transformation. Promoting responsible data management practices and compliance of data with the FAIR principles will contribute to build trust and ensure re-usability of data²⁴. Equally important is investment in key computing technologies and infrastructures.

The Commission has proposed more than €4 billion under the Digital Europe Programme to support high-performance and quantum computing, including edge computing and AI, data and cloud infrastructure. The European data strategy develops these priorities further.

H. INTERNATIONAL ASPECTS

Europe is well positioned to exercise global leadership in building alliances around shared values and promoting the ethical use of AI. The EU's work on AI has already influenced international discussions. When developing its ethical guidelines, the High-Level Expert Group involved a number of non-EU organisations and several governmental observers. In parallel, the EU was closely involved in developing the OECD's ethical principles for AI²⁵. The G20 subsequently endorsed these principles in its June 2019 Ministerial Statement on Trade and Digital Economy.

In parallel, the EU recognises that important work on AI is ongoing in other multilateral fora, including the Council of Europe, the United Nations Educational Scientific and Cultural Organization (UNESCO), the Organisation for Economic Co-operation and Development's (OECD), the World Trade Organisation and the International Telecommunications Union (ITU). At the UN, the EU is involved in the follow-up of the report of the High-Level Panel on Digital Cooperation, including its recommendation on AI.

The EU will continue to cooperate with like-minded countries, but also with global players, on AI, based on an approach based on EU rules and values (e.g. supporting upward regulatory convergence, accessing key resources including data, creating a level playing field). The Commission will closely monitor the policies of third countries that limit data flows and will address undue restrictions in

²⁴ Findable, Accessible, Interoperable and Reusable as stated in the Final Report and Action Plan from the Commission Expert Group on FAIR data, 2018, https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf.

²⁵ <https://www.oecd.org/going-digital/ai/principles/>

bilateral trade negotiations and through action in the context of the World Trade Organization. The Commission is convinced that international cooperation on AI matters must be based on an approach that promotes the respect of fundamental rights, including human dignity, pluralism, inclusion, non-discrimination and protection of privacy and personal data²⁶ and it will strive to export its values across the world²⁷. It is also clear that the responsible development and use of AI can be a driving force to achieve the Sustainable Development Goals and advance the 2030 Agenda.

5. AN ECOSYSTEM OF TRUST: REGULATORY FRAMEWORK FOR AI

As with any new technology, the use of AI brings both opportunities and risks. Citizens fear being left powerless in defending their rights and safety when facing the information asymmetries of algorithmic decision-making, and companies are concerned by legal uncertainty. While AI can help protect citizens' security and enable them to enjoy their fundamental rights, citizens also worry that AI can have unintended effects or even be used for malicious purposes. These concerns need to be addressed. Moreover, in addition to a lack of investment and skills, lack of trust is a main factor holding back a broader uptake of AI.

That is why the Commission set out an AI strategy²⁸ on 25 April 2018 addressing the socioeconomic aspects in parallel with an increase in investment in research, innovation and AI-capacity across the EU. It agreed a Coordinated Plan²⁹ with the Member States to align strategies. The Commission also established a High-Level Expert Group that published Guidelines on trustworthy AI in April 2019³⁰.

The Commission published a Communication³¹ welcoming the seven key requirements identified in the Guidelines of the High-Level Expert Group:

- Human agency and oversight,
- Technical robustness and safety,
- Privacy and data governance,
- Transparency,
- Diversity, non-discrimination and fairness,
- Societal and environmental wellbeing, and
- Accountability.

In addition, the Guidelines contain an assessment list for practical use by companies. During the second half of 2019, over 350 organisations have tested this assessment list and sent feedback. The High-Level Group is in the process of revising its guidelines in light of this feedback and will finalise this work by June 2020. A key result of the feedback process is that while a number of the requirements are already reflected in existing legal or regulatory regimes, those regarding transparency, traceability and human oversight are not specifically covered under current legislation in many economic sectors.

On top of this set of non-binding Guidelines of the High-Level Expert Group, and in line with the President's political guidelines, a clear European regulatory framework would build trust among

²⁶ Under the Partnership Instrument, the Commission will finance a €2.5 million project that will facilitate cooperation with like-minded partners, in order to promote the EU AI ethical guidelines and to adopt common principles and operational conclusions.

²⁷ President Von der Leyen, A Union that strives for more – My agenda for Europe, page 17.

²⁸ COM(2018) 237.

²⁹ COM(2018) 795.

³⁰ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

³¹ COM(2019) 168.

consumers and businesses in AI, and therefore speed up the uptake of the technology. Such a regulatory framework should be consistent with other actions to promote Europe's innovation capacity and competitiveness in this field. In addition, it must ensure socially, environmentally and economically optimal outcomes and compliance with EU legislation, principles and values. This is particularly relevant in areas where citizens' rights may be most directly affected, for example in the case of AI applications for law enforcement and the judiciary.

Developers and deployers of AI are already subject to European legislation on fundamental rights (e.g. data protection, privacy, non-discrimination), consumer protection, and product safety and liability rules. Consumers expect the same level of safety and respect of their rights whether or not a product or a system relies on AI. However, some specific features of AI (e.g. opacity) can make the application and enforcement of this legislation more difficult. For this reason, there is a need to examine whether current legislation is able to address the risks of AI and can be effectively enforced, whether adaptations of the legislation are needed, or whether new legislation is needed.

Given how fast AI is evolving, the regulatory framework must leave room to cater for further developments. Any changes should be limited to clearly identified problems for which feasible solutions exist.

Member States are pointing at the current absence of a common European framework. The German Data Ethics Commission has called for a five-level risk-based system of regulation that would go from no regulation for the most innocuous AI systems to a complete ban for the most dangerous ones. Denmark has just launched the prototype of a Data Ethics Seal. Malta has introduced a voluntary certification system for AI. If the EU fails to provide an EU-wide approach, there is a real risk of fragmentation in the internal market, which would undermine the objectives of trust, legal certainty and market uptake.

A solid European regulatory framework for trustworthy AI will protect all European citizens and help create a frictionless internal market for the further development and uptake of AI as well as strengthening Europe's industrial basis in AI.

A. PROBLEM DEFINITION

While AI can do much good, including by making products and processes safer, it can also do harm. This harm might be both material (safety and health of individuals, including loss of life, damage to property) and immaterial (loss of privacy, limitations to the right of freedom of expression, human dignity, discrimination for instance in access to employment), and can relate to a wide variety of risks. A regulatory framework should concentrate on how to minimise the various risks of potential harm, in particular the most significant ones.

The main risks related to the use of AI concern the application of rules designed to protect fundamental rights (including personal data and privacy protection and non-discrimination), as well as safety³² and liability-related issues.

Risks for fundamental rights, including personal data and privacy protection and non-discrimination

³² This includes issues of cybersecurity, issues associated with AI applications in critical infrastructures, or malicious use of AI.

The use of AI can affect the values on which the EU is founded and lead to breaches of fundamental rights³³, including the rights to freedom of expression, freedom of assembly, human dignity, non-discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation, as applicable in certain domains, protection of personal data and private life,³⁴ or the right to an effective judicial remedy and a fair trial, as well as consumer protection. These risks might result from flaws in the overall design of AI systems (including as regards human oversight) or from the use of data without correcting possible bias (e.g. the system is trained using only or mainly data from men leading to suboptimal results in relation to women).

AI can perform many functions that previously could only be done by humans. As a result, citizens and legal entities will increasingly be subject to actions and decisions taken by or with the assistance of AI systems, which may sometimes be difficult to understand and to effectively challenge where necessary. Moreover, AI increases the possibilities to track and analyse the daily habits of people. For example, there is a potential risk that AI may be used, in breach of EU data protection and other rules, by state authorities or other entities for mass surveillance and by employers to observe how their employees behave. By analysing large amounts of data and identifying links among them, AI may also be used to retrace and de-anonymise data about persons, creating new personal data protection risks even in respect to datasets that per se do not include personal data. AI is also used by online intermediaries to prioritise information for their users and to perform content moderation. The processed data, the way applications are designed and the scope for human intervention can affect the rights to free expression, personal data protection, privacy, and political freedoms.

Certain AI algorithms, when exploited for predicting criminal recidivism, can display gender and racial bias, demonstrating different recidivism prediction probability for women vs men or for nationals vs foreigners. Source: Tolan S., Miron M., Gomez E. and Castillo C. "Why Machine Learning May Lead to Unfairness: Evidence from Risk Assessment for Juvenile Justice in Catalonia", Best Paper Award, International Conference on AI and Law, 2019

Certain AI programmes for facial analysis display gender and racial bias, demonstrating low errors for determining the gender of lighter-skinned men but high errors in determining gender for darker-skinned women. Source: Joy Buolamwini, Timnit Gebru; *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, PMLR 81:77-91, 2018.

Bias and discrimination are inherent risks of any societal or economic activity. Human decision-making is not immune to mistakes and biases. However, the same bias when present in AI could have a much larger effect, affecting and discriminating many people without the social control mechanisms that govern human behaviour³⁵. This can also happen when the AI system ‘learns’ while in operation.

³³ Council of Europe research shows that a large number of fundamental rights could be impacted from the use of AI, <https://rm.coe.int/algorithms-and-human-rights-en-rev/16807956b5>.

³⁴ The General Data Protection Regulation and the ePrivacy Directive (new ePrivacy Regulation under negotiation) address these risks but there might be a need to examine whether AI systems pose additional risks. The Commission will be monitoring and assessing the application of the GDPR on a continuous basis.

³⁵ The Commission’s Advisory Committee on Equal Opportunities for Women and Men is currently preparing an “Opinion on Artificial Intelligence” analysing inter alia the impacts of Artificial Intelligence on gender equality which is expected to be adopted by the Committee in early 2020. The EU Gender Equality Strategy 2020-2024 also addresses the link between AI on gender equality; The European Network of Equality Bodies (Equinet) will publish a report (by Robin

In such cases, where the outcome could not have been prevented or anticipated at the design phase, the risks will not stem from a flaw in the original design of the system but rather from the practical impacts of the correlations or patterns that the system identifies in a large dataset.

The specific characteristics of many AI technologies, including opacity ('black box-effect'), complexity, unpredictability and partially autonomous behaviour, may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights. Enforcement authorities and affected persons might lack the means to verify how a given decision made with the involvement of AI was taken and, therefore, whether the relevant rules were respected. Individuals and legal entities may face difficulties with effective access to justice in situations where such decisions may negatively affect them.

Risks for safety and the effective functioning of the liability regime

AI technologies may present new safety risks for users when they are embedded in products and services. For example, as result of a flaw in the object recognition technology, an autonomous car can wrongly identify an object on the road and cause an accident involving injuries and material damage. As with the risks to fundamental rights, these risks can be caused by flaws in the design of the AI technology, be related to problems with the availability and quality of data or to other problems stemming from machine learning. While some of these risks are not limited to products and services that rely on AI, the use of AI may increase or aggravate the risks.

A lack of clear safety provisions tackling these risks may, in addition to risks for the individuals concerned, create legal uncertainty for businesses that are marketing their products involving AI in the EU. Market surveillance and enforcement authorities may find themselves in a situation where they are unclear as to whether they can intervene, because they may not be empowered to act and/or don't have the appropriate technical capabilities for inspecting systems³⁶. Legal uncertainty may therefore reduce overall levels of safety and undermine the competitiveness of European companies.

If the safety risks materialise, the lack of clear requirements and the characteristics of AI technologies mentioned above make it difficult to trace back potentially problematic decisions made with the involvement of AI systems. This in turn may make it difficult for persons having suffered harm to obtain compensation under the current EU and national liability legislation.³⁷

Allen and Dee Masters) on "Regulating AI: the new role for Equality Bodies – Meeting the new challenges to equality and non-discrimination from increased digitalisation and the use of AI", expected early 2020.

³⁶ An example may be the smart watch for children. This product may cause no direct harm to the child wearing it, but lacking a minimum level of security, it can be easily used as a tool to have access to the child. Market surveillance authorities may find it difficult to intervene in cases where the risk is not linked to the product as such.

³⁷ The implications of AI, Internet of Things and other digital technologies for safety and liability legislation are analysed in the Commission Report accompanying this White Paper.

Under the Product Liability Directive, a manufacturer is liable for damage caused by a defective product. However, in the case of an AI based system such as autonomous cars, it may be difficult to prove that there is a defect in the product, the damage that has occurred and the causal link between the two. In addition, there is some uncertainty about how and to what extent the Product Liability Directive applies in the case of certain types of defects, for example if these result from weaknesses in the cybersecurity of the product.

Thus, the difficulty of tracing back potentially problematic decisions taken by AI systems and referred to above in relation to fundamental rights applies equally to safety and liability-related issues. Persons having suffered harm may not have effective access to the evidence that is necessary to build a case in court, for instance, and may have less effective redress possibilities compared to situations where the damage is caused by traditional technologies. These risks will increase as the use of AI becomes more widespread.

B. POSSIBLE ADJUSTMENTS TO EXISTING EU LEGISLATIVE FRAMEWORK RELATING TO AI

An extensive body of existing EU product safety and liability legislation³⁸, including sector-specific rules, further complemented by national legislation, is relevant and potentially applicable to a number of emerging AI applications.

As regards the protection of fundamental rights and consumer rights, the EU legislative framework includes legislation such as the Race Equality Directive³⁹, the Directive on equal treatment in employment and occupation⁴⁰, the Directives on equal treatment between men and women in relation to employment and access to goods and services⁴¹, a number of consumer protection rules⁴², as well as rules on personal data protection and privacy, notably the General Data Protection Regulation and other sectorial legislation covering personal data protection, such as the Data Protection Law Enforcement Directive⁴³. In addition, as from 2025, the rules on accessibility requirements for goods and services, set out in the European Accessibility Act will apply⁴⁴. In addition, fundamental rights need to be respected when implementing other EU legislation, including in the field of financial services, migration or responsibility of online intermediaries.

While the EU legislation remains in principle fully applicable irrespective of the involvement of AI, it is important to assess whether it can be enforced adequately to address the risks that AI systems create, or whether adjustments are needed to specific legal instruments.

³⁸ The EU legal framework for product safety consists of the General Product Safety Directive (Directive 2001/95/EC), as a safety net, and a number of sector-specific rules covering different categories of products ranging from machines, planes and cars to toys and medical devices aiming to provide a high level of health and safety. Product liability law is complemented by different systems of civil liability for damages caused by products or services.

³⁹ Directive 2000/43/EC.

⁴⁰ Directive 2000/78/EC.

⁴¹ Directive 2004/113/EC; Directive 2006/54/EC.

⁴² Such as the Unfair Commercial Practices Directive (Directive 2005/29/EC) and the Consumer Rights Directive (Directive 2011/83/EC).

⁴³ Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data.

⁴⁴ Directive (EU) 2019/882 on the accessibility requirements for products and services.

For example, economic actors remain fully responsible for the compliance of AI to existing rules that protects consumers, any algorithmic exploitation of consumer behaviour in violation of existing rules shall be not permitted and violations shall be accordingly punished.

The Commission is of the opinion that the legislative framework could be improved to address the following risks and situations:

- *Effective application and enforcement of existing EU and national legislation:* the key characteristics of AI create challenges for ensuring the proper application and enforcement of EU and national legislation. The lack of transparency (opaqueness of AI) makes it difficult to identify and prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation. Therefore, in order to ensure an effective application and enforcement, it may be necessary to adjust or clarify existing legislation in certain areas, for example on liability as further detailed in the Report, which accompanies this White Paper.
- *Limitations of scope of existing EU legislation:* an essential focus of EU product safety legislation is on the placing of products on the market. While in EU product safety legislation software, when is part of the final product, must comply with the relevant product safety rules, it is an open question whether stand-alone software is covered by EU product safety legislation, outside some sectors with explicit rules⁴⁵. General EU safety legislation currently in force applies to products and not to services, and therefore in principle not to services based on AI technology either (e.g. health services, financial services, transport services).
- *Changing functionality of AI systems:* the integration of software, including AI, into products can modify the functioning of such products and systems during their lifecycle. This is particularly true for systems that require frequent software updates or which rely on machine learning. These features can give rise to new risks that were not present when the system was placed on the market. These risks are not adequately addressed in the existing legislation which predominantly focuses on safety risks present at the time of placing on the market.
- *Uncertainty as regards the allocation of responsibilities between different economic operators in the supply chain:* in general, EU legislation on product safety allocates the responsibility to the producer of the product placed on the market, including all components e.g. AI systems. But the rules can for example become unclear if AI is added after the product is placed on the market by a party that is not the producer. In addition, EU product liability legislation provides for liability of producers and leaves national liability rules to govern liability of others in the supply chain.
- *Changes to the concept of safety:* the use of AI in products and services can give rise to risks that EU legislation currently does not explicitly address. These risks may be linked to cyber threats, personal security risks (linked for example to new applications of AI such as to home appliances), risks that result from loss of connectivity, etc. These risks may be present at the time of placing products on the market or arise as a result of software updates or self-learning when the product is being used. The EU should make full use of the tools at its disposal to

⁴⁵ For instance software intended by the manufacturer to be used for medical purposes is considered a medical device under the Medical Device Regulation (Regulation (EU) 2017/745).

enhance its evidence base on potential risks linked to AI applications, including using the experience of the EU Cybersecurity Agency (ENISA) for assessing the AI threat landscape.

As indicated earlier, several Member States are already exploring options for national legislation to address the challenges created by AI. This raises the risk that the single market may be fragmented. Divergent national rules are likely to create obstacles for companies that want to sell and operate AI systems in the single market. Ensuring a common approach at EU level would enable European companies to benefit from smooth access to the single market and support their competitiveness on global markets.

Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics

The Report, which accompanies this White Paper, analyses the relevant legal framework. It identifies uncertainties as to the application of this framework with respect to the specific risks posed by AI systems and other digital technologies.

It concludes that the current product safety legislation already supports an extended concept of safety protecting against all kind of risks arising from the product according to its use. However, provisions explicitly covering new risks presented by the emerging digital technologies could be introduced to provide more legal certainty.

- The autonomous behaviour of certain AI systems during its life cycle may entail important product changes having an impact on safety, which may require a new risk assessment. In addition, human oversight from the product design and throughout the lifecycle of the AI products and systems may be needed as a safeguard.
- Explicit obligations for producers could be considered also in respect of mental safety risks of users when appropriate (ex. collaboration with humanoid robots).
- Union product safety legislation could provide for specific requirements addressing the risks to safety of faulty data at the design stage as well as mechanisms to ensure that quality of data is maintained throughout the use of the AI products and systems.
- The opacity of systems based on algorithms could be addressed through transparency requirements.
- Existing rules may need to be adapted and clarified in the case of a stand-alone software placed as it is on the market or downloaded into a product after its placing on the market, when having an impact on safety.
- Given the increasing complexity of supply chains as regards new technologies, provisions specifically requesting cooperation between the economic operators in the supply chain and the users could provide legal certainty.

The characteristics of emerging digital technologies like AI, the IoT and robotics may challenge aspects of the liability frameworks and could reduce their effectiveness. Some of these characteristics could make it hard to trace the damage back to a person, which would be necessary for a fault-based claim in accordance with most national rules. This could significantly increase the costs for victims and means that liability claims against others than producers may be difficult to make or prove.

- Persons having suffered harm caused with the involvement of AI systems need to enjoy the same level of protection as persons having suffered harm caused by other technologies, whilst technological innovation should be allowed to continue to develop.
- All options to ensure this objective should be carefully assessed, including possible amendments to the Product Liability Directive and possible further targeted harmonisation of national liability rules. For example, the Commission is seeking views whether and to what extent it may be needed to mitigate the consequences of complexity by adapting the burden of proof required by national liability rules for damage caused by the operation of AI applications.

From the discussion above, the Commission concludes that – in addition to the possible adjustments to existing legislation – a new legislation specifically on AI may be needed in order to make the EU legal framework fit for the current and anticipated technological and commercial developments.

C. SCOPE OF A FUTURE EU REGULATORY FRAMEWORK

A key issue for the future specific regulatory framework on AI intelligence is to determine the scope of its application. The working assumption is that the regulatory framework would apply to products and services relying on AI. AI should therefore be clearly defined for the purposes of this White Paper, as well as any possible future policy-making initiative.

In its Communication on AI for Europe the Commission provided a first definition of AI⁴⁶. This definition was further refined by the High Level Expert Group⁴⁷.

In any new legal instrument, the definition of AI will need to be sufficiently flexible to accommodate technical progress while being precise enough to provide the necessary legal certainty.

For the purposes of this White Paper, as well as of any possible future discussions on policy initiatives, it seems important to clarify the main elements that compose AI, which are “data” and “algorithms”. AI can be integrated in hardware. In case of machine learning techniques, which constitute a subset of AI, algorithms are trained to infer certain patterns based on a set of data in order to determine the actions needed to achieve a given goal. Algorithms

In autonomous driving for example, the algorithm uses, in real time, the data from the car (speed, engine consumption, shock-absorbers, etc..) and from the sensors scanning the whole environment of the car (road, signs, other vehicles, pedestrians etc..) to derive which direction, acceleration and speed the car should take to reach a certain destination. Based on the data observed, the algorithm adapts to the situation of the road and to the outside conditions, including other drivers’ behaviour, to derive the most comfortable and safest drive.

may continue to learn when in use. While AI-based products can act autonomously by perceiving their environment and without following a pre-determined set of instructions, their behaviour is largely defined and constrained by its developers. Humans determine and programme the goals, which an AI system should optimise for.

The EU has a strict legal framework in place to ensure inter alia consumer protection, to address unfair commercial practices and to protect personal data and privacy. In addition, the *acquis* contains specific rules for certain sectors (e.g. healthcare, transport). These existing provisions of EU law will continue to apply in relation to AI, although certain updates to that framework may be necessary to reflect the digital transformation and the use of AI (see section B). As a consequence, those aspects that are

⁴⁶ COM(2018) 237 final, p. 1: “Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).”

⁴⁷ High Level Expert Group, A definition of AI, p. 8: “Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.”

already addressed by existing horizontal or sectoral legislation (e.g. on medical devices⁴⁸, in transport systems) will continue to be governed by this legislation.

As a matter of principle, the new regulatory framework for AI should be effective to achieve its objectives while not being excessively prescriptive so that it could create a disproportionate burden, especially for SMEs. To strike this balance, the Commission is of the view that it should follow a risk-based approach.

A risk-based approach is important to help ensure that the regulatory intervention is proportionate. However, it requires clear criteria to differentiate between the different AI applications, in particular in relation to the question whether or not they are ‘high-risk’⁴⁹. The determination of what is a high-risk AI application should be clear and easily understandable and applicable for all parties concerned. Nevertheless even if an AI application is not qualified as high-risk, it remains entirely subject to already existing EU-rules.

The Commission is of the opinion that a given AI application should generally be considered high-risk in light of what is at stake, considering whether both the sector and the intended use involve significant risks, in particular from the viewpoint of protection of safety, consumer rights and fundamental rights. More specifically, an AI application should be considered high-risk where it meets the following two cumulative criteria:

- First, the AI application is employed in a sector where, given the characteristics of the activities typically undertaken, significant risks can be expected to occur. This first criterion ensures that the regulatory intervention is targeted on the areas where, generally speaking, risks are deemed most likely to occur. The sectors covered should be specifically and exhaustively listed in the new regulatory framework. For instance, healthcare; transport; energy and parts of the public sector.⁵⁰ The list should be periodically reviewed and amended where necessary in function of relevant developments in practice;
- Second, the AI application in the sector in question is, in addition, used in such a manner that significant risks are likely to arise. This second criterion reflects the acknowledgment that not every use of AI in the selected sectors necessarily involves significant risks. For example, whilst healthcare generally may well be a relevant sector, a flaw in the appointment scheduling system in a hospital will normally not pose risks of such significance as to justify legislative intervention. The assessment of the level of risk of a given use could be based on the impact on the affected parties. For instance, uses of AI applications that produce legal or similarly significant effects for the rights of an individual or a company; that pose risk of injury, death or significant material or immaterial damage; that produce effects that cannot reasonably be avoided by individuals or legal entities.

The application of the two cumulative criteria would ensure that the scope of the regulatory framework is targeted and provides legal certainty. The mandatory requirements contained in the new regulatory framework on AI (see section D below) would in principle apply only to those applications identified as high-risk in accordance with these two cumulative criteria.

⁴⁸ For example, there are different safety considerations and legal implications concerning AI systems that provide specialized medical information to physicians, AI systems providing medical information directly to the patient and AI systems performing medical tasks themselves directly on a patient. The Commission is examining these safety and liability challenges that are distinct to healthcare.

⁴⁹ EU legislation may categorise “risks” differently to what is described here, depending on the area, such as for example, product safety

⁵⁰ The public sector could include areas like asylum, migration, border controls and judiciary, social security and employment services.

Notwithstanding the foregoing, there may also be exceptional instances where, due to the risks at stake, the use of AI applications for certain purposes is to be considered as high-risk as such – that is, irrespective of the sector concerned and where the below requirements would still apply.⁵¹ As an illustration, one could think in particular of the following:

- In light of its significance for individuals and of the EU acquis addressing employment equality, the use of AI applications for recruitment processes as well as in situations impacting workers' rights would always be considered “high-risk” and therefore the below requirements would at all times apply. Further specific applications affecting consumer rights could be considered.
- the use of AI applications for the purposes of remote biometric identification⁵² and other intrusive surveillance technologies, would always be considered “high-risk” and therefore the below requirements would at all times apply.

D. TYPES OF REQUIREMENTS

When designing the future regulatory framework for AI, it will be necessary to decide on the types of mandatory legal requirements to be imposed on the relevant actors. These requirements may be further specified through standards. As noted in section C above and in addition to already existing legislation, those requirements would apply to high-risk AI applications only, thus ensuring that any regulatory intervention is focused and proportionate.

Taking into account the guidelines of the High Level Expert Group and what has been set out in the foregoing, the requirements for high-risk AI applications could consist of the following key features, which are discussed in further detail in the subsections below:

- training data;
- data and record-keeping;
- information to be provided;
- robustness and accuracy;
- human oversight;
- specific requirements for certain particular AI applications, such as those used for purposes of remote biometric identification.

To ensure legal certainty, these requirements will be further specified to provide a clear benchmark for all the actors who need to comply with them.

a) Training data

It is more important than ever to promote, strengthen and defend the EU's values and rules, and in particular the rights that citizens derive from EU law. These efforts undoubtedly also extend to the high-risk AI applications marketed and used in the EU under consideration here.

⁵¹ It is important to highlight that other pieces of EU legislation may also apply. For example, when incorporated into a consumer product, the General Product Safety Directive may apply to the safety of AI applications.

⁵² Remote biometric identification should be distinguished from biometric authentication (the latter is a security process that relies on the unique biological characteristics of an individual to verify that he/she is who he/she says he/she is). Remote biometric identification is when the identities of multiple persons are established with the help of biometric identifiers (fingerprints, facial image, iris, vascular patterns, etc.) at a distance, in a public space and in a continuous or ongoing manner by checking them against data stored in a database.

As discussed earlier, without data, there is no AI. The functioning of many AI systems, and the actions and decisions to which they may lead, very much depend on the data set on which the systems have been trained. The necessary measures should therefore be taken to ensure that, where it comes to the data used to train AI systems, the EU's values and rules are respected, specifically in relation to safety and existing legislative rules for the protection of fundamental rights. The following requirements relating to the data set used to train AI systems could be envisaged:

- Requirements aimed at providing reasonable assurances that the subsequent use of the products or services that the AI system enables is safe, in that it meets the standards set in the applicable EU safety rules (existing as well as possible complementary ones). For instance, requirements ensuring that AI systems are trained on data sets that are sufficiently broad and cover all relevant scenarios needed to avoid dangerous situations.
- Requirements to take reasonable measures aimed at ensuring that such subsequent use of AI systems does not lead to outcomes entailing prohibited discrimination. These requirements could entail in particular obligations to use data sets that are sufficiently representative, especially to ensure that all relevant dimensions of gender, ethnicity and other possible grounds of prohibited discrimination are appropriately reflected in those data sets;
- Requirements aimed at ensuring that privacy and personal data are adequately protected during the use of AI-enabled products and services. For issues falling within their respective scope, the General Data Protection Regulation and the Law Enforcement Directive regulate these matters.

b) Keeping of records and data

Taking into account elements such as the complexity and opacity of many AI systems and the related difficulties that may exist to effectively verify compliance with and enforce the applicable rules, requirements are called for regarding the keeping of records in relation to the programming of the algorithm, the data used to train high-risk AI systems, and, in certain cases, the keeping of the data themselves. These requirements essentially allow potentially problematic actions or decisions by AI systems to be traced back and verified. This should not only facilitate supervision and enforcement; it may also increase the incentives for the economic operators concerned to take account at an early stage of the need to respect those rules.

To this aim, the regulatory framework could prescribe that the following should be kept:

- accurate records regarding the data set used to train and test the AI systems, including a description of the main characteristics and how the data set was selected;
- in certain justified cases, the data sets themselves;
- documentation on the programming⁵³ and training methodologies, processes and techniques used to build, test and validate the AI systems, including where relevant in respect of safety and avoiding bias that could lead to prohibited discrimination.

The records, documentation and, where relevant, data sets would need to be retained during a limited, reasonable time period to ensure effective enforcement of the relevant legislation. Measures should be

⁵³ For instance, documentation on the algorithm including what the model shall optimise for, which weights are designed to certain parameters at the outset etc.

taken to ensure that they are made available upon request, in particular for testing or inspection by competent authorities. Where necessary, arrangements should be made to ensure that confidential information, such as trade secrets, is protected.

c) Information provision

Transparency is required also beyond the record-keeping requirements discussed in point c) above. In order to achieve the objectives pursued – in particular promoting the responsible use of AI, building trust and facilitating redress where needed – it is important that adequate information is provided in a proactive manner about the use of high-risk AI systems.

Accordingly, the following requirements could be considered:

- Ensuring clear information to be provided as to the AI system’s capabilities and limitations, in particular the purpose for which the systems are intended, the conditions under which they can be expected to function as intended and the expected level of accuracy in achieving the specified purpose. This information is important especially for deployers of the systems, but it may also be relevant to competent authorities and affected parties.
- Separately, citizens should be clearly informed when they are interacting with an AI system and not a human being. Whilst EU data protection legislation already contain certain rules of this kind⁵⁴, additional requirements may be called for to achieve the abovementioned objectives. If so, unnecessary burdens should be avoided. Therefore, no such information needs to be provided, for instance, in situations where it is immediately obvious to citizens that they are interacting with AI systems. It is furthermore important that the information provided is objective, concise and easily understandable. The manner in which the information is to be provided should be tailored to the particular context.

d) Robustness and accuracy

AI systems – and certainly high-risk AI applications – must be technically robust and accurate in order to be trustworthy. That means that such systems need to be developed in a responsible manner and with an ex-ante due and proper consideration of the risks that they may generate. Their development and functioning must be such to ensure that AI systems behave reliably as intended. All reasonable measures should be taken to minimise the risk of harm being caused.

Accordingly, the following elements could be considered:

- Requirements ensuring that the AI systems are robust and accurate, or at least correctly reflect their level of accuracy, during all life cycle phases;
- Requirements ensuring that outcomes are reproducible;
- Requirements ensuring that AI systems can adequately deal with errors or inconsistencies during all life cycle phases.

⁵⁴ In particular, pursuant to Art. 13(2)(f) GDPR, controllers must, at the time when the personal data are obtained, provide the data subjects with further information necessary to ensure fair and transparent processing about the existence of automated decision-making and certain additional information.

- Requirements ensuring that AI systems are resilient against both overt attacks and more subtle attempts to manipulate data or algorithms themselves, and that mitigating measures are taken in such cases.

e) Human oversight

Human oversight helps ensuring that an AI system does not undermine human autonomy or cause other adverse effects. The objective of trustworthy, ethical and human-centric AI can only be achieved by ensuring an appropriate involvement by human beings in relation to high-risk AI applications.

Even though the AI applications considered in this White paper for a specific legal regime are all considered high-risk, the appropriate type and degree of human oversight may vary from one case to another. It shall depend in particular on the intended use of the systems and the effects that the use could have for affected citizens and legal entities. It shall also be without prejudice to the legal rights established by the GDPR when the AI system processes personal data. For instance, human oversight could have the following, non-exhaustive, manifestations:

- the output of the AI system does not become effective unless it has been previously reviewed and validated by a human (e.g. the rejection of an application for social security benefits may be taken by a human only);
- the output of the AI system becomes immediately effective, but human intervention is ensured afterwards (e.g. the rejection of an application for a credit card may be processed by an AI system, but human review must be possible afterwards);
- monitoring of the AI system while in operation and the ability to intervene in real time and deactivate (e.g. a stop button or procedure is available in a driverless car when a human determines that car operation is not safe);
- in the design phase, by imposing operational constraints on the AI system (e.g. a driverless car shall stop operating in certain conditions of low visibility when sensors may become less reliable or shall maintain a certain distance in any given condition from the preceding vehicle).

f) Specific requirements for remote biometric identification

The gathering and use of biometric data⁵⁵ for remote identification⁵⁶ purposes, for instance through deployment of facial recognition in public places, carries specific risks for fundamental rights⁵⁷. The

⁵⁵ Biometric data is defined as “personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique authentication or identification of that natural person, such as facial images or dactyloscopic [fingerprint] data.” (Law Enforcement Directive, Art. 3 (13); GDPR, Art. 4 (14); Regulation (EU) 2018/1725, Art. 3 (18)).

⁵⁶ In connection to facial recognition, identification means that the template of a person’s facial image is compared to many other templates stored in a database to find out if his or her image is stored there. Authentication (or verification) on the other hand is often referred to as one-to-one matching. It enables the comparison of two biometric templates, usually assumed to belong to the same individual. Two biometric templates are compared to determine if the person shown on the two images is the same person. Such a procedure is, for example, used at Automated Border Control (ABC) gates used for border checks at airports.

⁵⁷ For example on people’s dignity. Relatedly, the rights to respect for private life and protection of personal data are at the core of fundamental rights concerns when using facial recognition technology. There is also a potential impact on non-discrimination and rights of special groups, such as children, older persons and persons with disabilities. Moreover, freedom of expression, association and assembly must not be undermined by the use of the technology. See: Facial recognition technology: fundamental rights considerations in the context of law enforcement, <https://fra.europa.eu/en/publication/2019/facial-recognition>.

fundamental rights implications of using remote biometric identification AI systems can vary considerably depending on the purpose, context and scope of the use.

EU data protection rules prohibit in principle the processing of biometric data for the purpose of uniquely identifying a natural person, except under specific conditions⁵⁸. Specifically, under the GDPR, such processing can only take place on a limited number of grounds, the main one being for reasons of substantial public interest. In that case, the processing must take place on the basis of EU or national law, subject to the requirements of proportionality, respect for the essence of the right to data protection and appropriate safeguards. Under the Law Enforcement Directive, there must be a strict necessity for such processing, in principle an authorisation by EU or national law as well as appropriate safeguards. As any processing of biometric data for the purpose of uniquely identifying a natural person would relate to an exception to a prohibition laid down in EU law, it would be subject to the Charter of Fundamental Rights of the EU.

It follows that, in accordance with the current EU data protection rules and the Charter of Fundamental Rights, AI can only be used for remote biometric identification purposes where such use is duly justified, proportionate and subject to adequate safeguards.

In order to address possible societal concerns relating to the use of AI for such purposes in public places, and to avoid fragmentation in the internal market, the Commission will launch a broad European debate on the specific circumstances, if any, which might justify such use, and on common safeguards.

E. ADDRESSEES

In relation to the addressees of the legal requirements that would apply in relation to the high-risk AI applications referred to above, there are two main issues to be considered.

First, there is the question how obligations are to be distributed among the economic operators involved. Many actors are involved in the lifecycle of an AI system. These include the developer, the deployer (the person who uses an AI-equipped product or service) and potentially others (producer, distributor or importer, service provider, professional or private user).

It is the Commission's view that, in a future regulatory framework, each obligation should be addressed to the actor(s) who is (are) best placed to address any potential risks. For example, while the developers of AI may be best placed to address risks arising from the development phase, their ability to control risks during the use phase may be more limited. In that case, the deployer should be subject to the relevant obligation. This is without prejudice to the question whether, for the purpose of liability to end-users or other parties suffering harm and ensuring effective access to justice, which party should be liable for any damage caused. Under EU product liability law, liability for defective products is attributed to the producer, without prejudice to national laws which may also allow recovery from other parties.

Second, there is the question about the geographic scope of the legislative intervention. In the view of the Commission, it is paramount that the requirements are applicable to all relevant economic operators providing AI-enabled products or services in the EU, regardless of whether they are established in the EU or not. Otherwise, the objectives of the legislative intervention, mentioned earlier, could not fully be achieved.

⁵⁸ Article 9 GDPR, Article 10 Law Enforcement Directive. See also Article 10 Regulation (EU) 2018/1725 (applicable to the EU institutions and bodies).

F. COMPLIANCE AND ENFORCEMENT

In order to ensure that AI is trustworthy, secure and in respect of European values and rules, the applicable legal requirements need to be complied with in practice and be effectively enforced both by competent national and European authorities and by affected parties. Competent authorities should be in a position to investigate individual cases, but also to assess the impact on society.

In view of the high risk that certain AI applications pose for citizens and our society (see section A above), the Commission considers at this stage that an objective, prior conformity assessment would be necessary to verify and ensure that certain of the above mentioned mandatory requirements applicable to high-risk applications (see section D above) are complied with. The prior conformity assessment could include procedures for testing, inspection or certification⁵⁹. It could include checks of the algorithms and of the data sets used in the development phase.

The conformity assessments for high-risk AI applications should be part of the conformity assessment mechanisms that already exist for a large number of products being placed on the EU's internal market. Where no such existing mechanisms can be relied on, similar mechanisms may need to be established, drawing on best practice and possible input of stakeholders and European standards organisations. Any such new mechanism should be proportionate and non-discriminatory and use transparent and objective criteria in compliance with international obligations.

When designing and implementing a system relying on prior conformity assessments, particular account should be taken of the following:

- Not all requirements outlined above may be suitable to be verified through a prior conformity assessment. For instance, the requirement about information to be provided generally does not lend itself well for verification through such an assessment.
- Particular account should be taken of the possibility that certain AI systems evolve and learn from experience, which may require repeated assessments over the life-time of the AI systems in question.
- The need to verify the data used for training and the relevant programming and training methodologies, processes and techniques used to build, test and validate AI systems.
- In case the conformity assessment shows that an AI system does not meet the requirements for example relating to the data used to train it, the identified shortcomings will need to be remedied, for instance by re-training the system in the EU in such a way as to ensure that all applicable requirements are met.

The conformity assessments would be mandatory for all economic operators addressed by the requirements, regardless of their place of establishment⁶⁰. In order to limit the burden on SMEs, some support structure might be envisaged including through the Digital Innovation Hubs. In addition, standards as well as dedicated online tools could facilitate compliance.

⁵⁹ The system would be based on conformity assessment procedures in the EU, see Decision 768/2008/EC or on Regulation (EU) 2019/881 (Cybersecurity Act), taking into account the specificities of AI. See the Blue Guide on the Implementation of EU product rules, 2014.

⁶⁰ As regards the relevant governance structure, including the bodies designated to carry out the conformity assessments, see section H below.

Any prior conformity assessment should be without prejudice to monitoring compliance and ex post enforcement by competent national authorities. That holds true in respect of high-risk AI applications, but also in respect of other AI applications subject to legal requirements, although the high-risk nature of the applications at issue may be reason for the competent national authorities to give particular attention to the former. Ex-post controls should be enabled by adequate documentation of the relevant AI application (see section E above) and, where appropriate, a possibility for third parties such as competent authorities to test such applications. This may be especially important where risks to fundamental rights arise, which are context dependent. Such monitoring of compliance should be part of a continuous market surveillance scheme. Governance-related aspects are further discussed in section H below.

Moreover, both for high-risk AI applications and for other AI applications, effective judicial redress for parties negatively affected by AI systems should be ensured. Issues related to liability are further discussed in the Report on the safety and liability framework accompanying this White Paper.

G. VOLUNTARY LABELLING FOR NO-HIGH RISK AI APPLICATIONS

For AI applications that do not qualify as ‘high-risk’ (see section C above) and that are therefore not subject to the mandatory requirements discussed above (see sections D, E and F above), an option would be, in addition to applicable legislation, to establish a voluntary labelling scheme.

Under the scheme, interested economic operators that are not covered by the mandatory requirements could decide to make themselves subject, on a voluntary basis, either to those requirements or to a specific set of similar requirements especially established for the purposes of the voluntary scheme. The economic operators concerned would then be awarded a quality label for their AI applications.

The voluntary label would allow the economic operators concerned to signal that their AI-enabled products and services are trustworthy. It would allow users to easily recognise that the products and services in question are in compliance with certain objective and standardised EU-wide benchmarks, going beyond the normally applicable legal obligations. This would help enhance the trust of users in AI systems and promote the overall uptake of the technology.

This option would entail the creation of a new legal instrument that sets out the voluntary labelling framework for developers and/or deployers of AI systems that are not be considered as high-risk. While participation in the labelling scheme would be voluntary, once the developer or the deployer opted to use the label, the requirements would be binding. The combination of *ex ante* and *ex post* enforcement would need to ensure that all requirements are complied with.

H. GOVERNANCE

A European governance structure on AI in the form of a framework for cooperation of national competent authorities is necessary to avoid fragmentation of responsibilities, increase capacity in Member States, and make sure that Europe equips itself progressively with the capacity needed for testing and certification of AI-enabled products and services. In this context, it would be beneficial to support competent national authorities to enable them to fulfil their mandate where AI is used.

A European governance structure could have a variety of tasks, as a forum for a regular exchange of information and best practice, identifying emerging trends, advising on standardisation activity as well as on certification. It should also play a key role in facilitating the implementation of the legal framework, such as through issuing guidance, opinions and expertise. To that effect, it should rely on a network of national authorities, as well as sectorial networks and regulatory authorities, at national and EU level. Moreover, a committee of experts could provide assistance to the Commission.

The governance structure should guarantee maximum stakeholders participation. Stakeholders – consumer organisation and social partners, businesses, researchers, and civil society organisations – should be consulted on the implementation and the further development of the framework.

Given already existing structures such as in finance, pharmaceuticals, aviation, medical devices, consumer protection, data protection, the proposed governance structure should not duplicate existing functions. It should instead establish close links with other EU and national competent authorities in the various sectors to complement existing expertise and help existing authorities in monitoring and the oversight of the activities of economic operators involving AI systems and AI-enabled products and services.

Finally, if this option is pursued, the carrying out of conformity assessments could be entrusted to notified bodies designated by Member States. Testing centres should enable the independent audit and assessment of AI-systems in accordance with the requirements outlined above. Independent assessment will increase trust and ensures objectivity. It could also facilitate the work of relevant competent authorities.

The EU enjoys excellent testing and assessment centres and should develop its capacity also in the area of AI. Economic operators established in third countries wanting to enter the internal market could either make use of designated bodies established in the EU or, subject to mutual recognition agreements with third countries, have recourse to third-country bodies designated to carry out such assessment.

The governance structure relating to AI and the possible conformity assessments at issue here would leave the powers and responsibilities under existing EU law of the relevant competent authorities in specific sectors or on specific issues (finance, pharmaceuticals, aviation, medical devices, consumer protection, data protection, etc.) unaffected.

6. CONCLUSION

AI is a strategic technology that offers many benefits for citizens, companies and society as a whole, provided it is human-centric, ethical, sustainable and respects fundamental rights and values. AI offers important efficiency and productivity gains that can strengthen the competitiveness of European industry and improve the wellbeing of citizens. It can also contribute to finding solutions to some of the most pressing societal challenges, including the fight against climate change and environmental degradation, the challenges linked to sustainability and demographic changes, and the protection of our democracies and, where necessary and proportionate, the fight against crime.

For Europe to seize fully the opportunities that AI offers, it must develop and reinforce the necessary industrial and technological capacities. As set out in the accompanying European strategy for data, this also requires measures that will enable the EU to become a global hub for data.

The European approach for AI aims to promote Europe's innovation capacity in the area of AI while supporting the development and uptake of ethical and trustworthy AI across the EU economy. AI should work for people and be a force for good in society.

With this White Paper and the accompanying Report on the safety and liability framework, the Commission launches a broad consultation of Member States civil society, industry and academics, of concrete proposals for a European approach to AI. These include both policy means to boost investments in research and innovation, enhance the development of skills and support the uptake of AI by SMEs, and proposals for key elements of a future regulatory framework. This consultation will

allow a comprehensive dialogue with all concerned parties that will inform the next steps of the Commission.

The Commission invites for comments on the proposals set out in the White Paper through an open public consultation available at https://ec.europa.eu/info/consultations_en. The consultation is open for comments until 19 May 2020.

It is standard practice for the Commission to publish submissions received in response to a public consultation. However, it is possible to request that submissions, or parts thereof, remain confidential. Should this be the case, please indicate clearly on the front page of your submission that it should not be made public and also send a non-confidential version of your submission to the Commission for publication.

Artificial Intelligence and Law Enforcement

Impact on Fundamental Rights



Policy Department for Citizens' Rights and Constitutional Affairs
Directorate-General for Internal Policies
PE 656.295 – July 2020

EN

Artificial Intelligence and Law Enforcement

Impact on Fundamental Rights

Abstract

This study, commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs at the request of the LIBE Committee, examines the impact on fundamental rights of Artificial Intelligence in the field of law enforcement and criminal justice, from a European Union perspective. It presents the applicable legal framework (notably in relation to data protection), and analyses major trends and key policy discussions. The study also considers developments following the Covid-19 outbreak. It argues that the seriousness and scale of challenges may require intervention at EU level, based on the acknowledgement of the area's specificities.

This document was requested by the European Parliament's Committee on Civil Liberties, Justice and Home Affairs.

AUTHOR

Prof. Dr. Gloria GONZÁLEZ FUSTER, Vrije Universiteit Brussel (VUB)

ADMINISTRATOR RESPONSIBLE

Alessandro DAVOLI

EDITORIAL ASSISTANT

Ginka TSONEVA

LINGUISTIC VERSIONS

Original: EN

ABOUT THE EDITOR

Policy departments provide in-house and external expertise to support EP committees and other parliamentary bodies in shaping legislation and exercising democratic scrutiny over EU internal policies.

To contact the Policy Department or to subscribe for updates, please write to:

Policy Department for Citizens' Rights and Constitutional Affairs

European Parliament

B-1047 Brussels

Email: poldep-citizens@europarl.europa.eu

Manuscript completed in July 2020

© European Union, 2020

This document is available on the internet at:

<http://www.europarl.europa.eu/supporting-analyses>

DISCLAIMER AND COPYRIGHT

The opinions expressed in this document are the sole responsibility of the authors and do not necessarily represent the official position of the European Parliament.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the European Parliament is given prior notice and sent a copy

CONTENTS

LIST OF ABBREVIATIONS	5
LIST OF BOXES	7
LIST OF FIGURES	7
EXECUTIVE SUMMARY	8
1. GENERAL INFORMATION	10
1.1. Objectives and structure	11
1.2. Methodology	12
2. EU LEGAL FRAMEWORK	13
2.1. Primary law	13
2.2. Secondary law	14
2.2.1. Rules on the processing of personal data	14
2.2.2. Rules on the processing of non-personal data	19
3. AI IN LAW ENFORCEMENT AND CRIMINAL JUSTICE	21
3.1. 'Predictive policing'	22
3.2. Facial recognition	24
3.3. AI and criminal justice	27
3.4. AI and borders	29
3.4.1. General trends	29
3.4.2. PNR	31
3.4.3. ETIAS	32
4. IMPACT ON FUNDAMENTAL RIGHTS	37
4.1. Privacy and data protection	38
4.2. Freedom of expression and information, and freedom of assembly and association	40
4.3. Non-discrimination	40
4.4. Right to an effective remedy and to a fair trial	43
4.5. Rights of the child	43
5. POLICY DEVELOPMENTS	45
5.1. Introduction	47
5.2. European Commission	48
5.3. European Parliament	54
5.4. The role of ethics	54

5.5. EU-funded support of AI	57
6. COVID-19 OUTBREAK RESPONSE	61
6.1. Data-driven responses	63
6.2. Law enforcement initiatives	66
7. POLICY RECOMMENDATIONS	67
7.1. Assess fully the relevance and gaps of the existing legal framework	67
7.2. Adapt initiatives to the specificities of the field	68
7.3. Consider the need for special rules for certain practices	68
7.4. Clarify the role of ethics	69
7.5. Reinforce trust in EU-funded AI efforts	69
7.6. Screen developments related to the Covid-19 outbreak	69
REFERENCES	71
ANNEX – A COMPARISON BETWEEN ‘ALGORITHMIC DATA RIGHTS’ IN THE GDPR AND THE LED	83

LIST OF ABBREVIATIONS

ACLU	American Civil Liberties Union
ADM	Automated Decision Making
AI	Artificial Intelligence
CJEU	Court of Justice of the European Union
DAPIX	Working Party on Information Exchange and Data Protection
DPA	Data Protection Authority
EASO	European Asylum Support Office
EC	European Commission
ECHR	European Convention on Human Rights
ECRIS-TCN	European Criminal Records Information System for Third-Country Nationals
ECtHR	European Court of Human Rights
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EDRI	European Digital Rights Initiative
EES	Entry/Exit System
EIO	European Investigation Order
EPIC	Electronic Privacy Information Center
EPPO	European Public Prosecutor's Office
EU	European Union
FAT	Fairness, Accountability and Transparency
FIU	Financial Intelligence Unit

GDPR	General Data Protection Regulation
GSMA	Global System for Mobile Communications association
GVM	Gangs Violence Matrix
ICO	Information Commissioner's Office
IT	Information Technology
JRC	Joint Research Centre
LED	Law Enforcement Directive
LFR	Live Facial Recognition
OECD	Organisation for Economic Co-operation and Development
PNR	Passenger Name Record
RBI	Remote Biometric Identification
SIS	Schengen Information System
UNICRI	United Nations Interregional Crime and Justice Research Institute
VIS	Visa Information System

LIST OF BOXES

Example 1: Real-time surveillance of public space	59
Example 2: Autonomous robots at the borders	59
Example 3: Pre-occurring crime prediction and prevention	60
Example 4: Intelligent dialogue-empowered bots for early terrorist detection	60

LIST OF FIGURES

Figure 1: EDPB work on police and justice	19
Figure 2: AI HLEG Framework for Trustworthy AI	50

EXECUTIVE SUMMARY

Background

Artificial Intelligence (AI) is high on the agenda of the European Union (EU). Discussions on its possible regulation have been particularly prominent since Ursula von der Leyen announced, already before her nomination as President of the European Commission (EC), a strong will to situate AI among the Commission's top priorities. The endorsement of AI as an EU-level policy priority has been accompanied by a reflection on how to promote trust in AI technologies, and how to guarantee that AI systems do not compromise, during their development or deployment, EU fundamental rights. This specific reflection is actually not completely novel, but entrenched in prior legal and policy debates on fundamental rights issues connected notably to 'big data', as well as, more generally, related to the regulation of the processing of both personal and non-personal data.

Aim

This study aims at analysing the **impact on EU fundamental rights of AI in the field of law enforcement and criminal justice**. It approaches the subject from an EU law and policy perspective. It first succinctly presents the main features of the applicable legal framework. Second, it introduces recent and ongoing developments in the field, focusing on 'predictive policing', facial recognition, the use of AI in criminal justice, and AI and borders. It discusses the impact on fundamental rights of these trends, and reviews relevant policy discussions around AI regulation. After such exploration, the study puts forward some concrete recommendations.

Findings

The field of law enforcement and criminal justice is **particularly sensitive**, as it touches upon core issues of the relation between the individual and the State. There is a broad consensus on the fact that reliance on AI systems in this area can substantially impact EU fundamental rights, and more generally democracy itself. The issue thus deserves special consideration in the context of any discussion on the future of trustworthy AI.

This study:

- Shows that the advent of AI in the field of law enforcement and criminal justice is already a reality, as **AI systems are increasingly being adopted or considered**; such systems might take many different shapes, and occur at the cross-roads of different sustained trends to increase the use of data, algorithms, and computational power; these developments are connected to pre-existing trends, some of which had been previously broadly framed under other notions or policy priorities ('big data');
- Documents that such developments have already generated **significant controversies**, notably in relation to 'predictive policing', facial recognition, AI and criminal justice, and AI and borders (including a reflection on the European Travel Information and Authorization System, ETIAS), for instance in litigation and calls from civil society to better prevent or mitigate associated risks, both in the EU and beyond;
- Points out that discussions around AI regulation in the EU have been deeply embedded in the **EU Digital Single Market agenda**, and that generally speaking, although they might refer to the need to consider law enforcement and criminal justice specificities, such policy discussions are most often not based on a detailed review and taking into account of concrete applicable rules (and, especially, of applicable restrictions and derogations);

- Warns that such policy discussions suffer from **persistent ambivalences on the role of 'ethics'** in relation to the safeguarding of fundamental rights, generating not only lack of conceptual precision but most importantly uncertainty as to whether effective protection of fundamental rights will be delivered;
- Emphasises that **the current EU data protection legal framework shall not be assumed to offer enough solid safeguards for individuals** in light of the increased uses of automated decision-making and profiling for law enforcement and criminal justice purposes, as:
 - the general safeguards provided by the General Data Protection Regulation (GDPR) do not necessarily apply when the processing is for such purposes, as restrictions and derogations might be applicable;
 - the Law Enforcement Data Protection Directive (LED Directive), which might be the applicable relevant instruments, provides for safeguards that are similar to those of the GDPR, but nevertheless not exactly equivalent, and equally provides for possible restrictions and derogations.
- Reviews a variety of fundamental rights considerations connected to AI in the field of law enforcement and criminal justice, illustrating the fact that they include privacy and data protection issues but also other challenges, notably related to **non-discrimination**;
- Describes a **problematic lack of minimum transparency standards in relation to the support of AI research with EU funds**, affecting notably the possibility to assess how funded research complies with the respect of EU fundamental rights and EU law in general, but also further scientific work towards trustworthy AI solutions;
- Cautions that **responses to the Covid-19 outbreak have led to the rapid proliferation of technological measures and data-driven initiatives to be carefully assessed**, including initiatives which have the ambition of sustaining an unprecedented widespread generalised collection of data about individuals ('contact tracing apps'), and initiatives that build on the fragile distinction between 'personal' and 'anonymised' data to facilitate extensive data processing;
- Recommends that due consideration be given to **the need for a legislative intervention to guarantee EU fundamental rights in the field of law enforcement and criminal justice**; that any possible role granted to 'ethics' in such intervention shall be clarified and justified in detail; that the possible gaps and inefficiencies of the EU data protection law, as well as national laws implementing EU data protection law, shall be duly examined specifically in the light of the law enforcement and criminal justice context; that EU-funded AI research must be better framed, and finally, that developments related to the Covid-19 outbreak shall be monitored with extreme attention.

1. GENERAL INFORMATION

KEY FINDINGS

Law enforcement and criminal justice can benefit from AI developments, and there are many AI-related systems and technologies being adopted and developed for law enforcement and criminal justice purposes in the European Union (EU). Some of these solutions, however, raise important questions in terms of their compatibility with EU fundamental rights.

EU-level policy discussions around the regulation of AI have been marked until recently by their strong embedding in the development of the Digital Single Market. In this policy context, the European Commission (EC) is notably advocating embracing the advent of AI while at the same time pursuing the establishment of an 'ecosystem of trust' – trust in AI – as a policy objective in itself. Ethical considerations have been given much attention from this perspective.

The magnitude and seriousness of challenges triggered by AI in the field of law enforcement and criminal justice, however, do not appear to be conveniently addressed by ongoing reflections. In this sense, future steps should notably consider the legal specificities of the area, and most notably the complexities and limitations of the EU data protection legal framework, especially insofar as personal data processing for law enforcement and criminal justice purposes is concerned, and more generally in relation to data processing in the Area of Freedom, Security and Justice (AFSJ).

EU-funded AI research related to law enforcement and criminal justice also deserves further attention. Important investments have already taken place, and more substantial funding is planned, but the framework accompanying the selection of projects and their implementation presents problematic transparency gaps.

The Covid-19 outbreak has led to data-driven and AI solutions being more present than ever, potentially exponentially multiplying the personal data to be eventually available for law enforcement purposes. Whereas fundamental rights safeguards, including data protection safeguards, appear more important now than ever, there have also been cases in which these safeguards appear to be threatened by exceptional measures taken invoking needs related to the crisis. This generates a particularly delicate situation for EU fundamental rights, calling for great vigilance.

Artificial Intelligence (AI) can be of substantial use **in the field of law enforcement and criminal justice**. This idea is broadly accepted and endorsed at European Union (EU) level,¹ and the trend towards using automated processing techniques and algorithms in crime prevention and the criminal justice systems has been described as generally growing for already a number of years (MSI-NET, 2018, 10). AI, however, is also generally perceived as being, at least potentially, in tension with certain **fundamental rights** recognised as such in the EU, and constituting a particularly high risk to such rights when used in the field of law enforcement and criminal justice.

Paradoxically, despite the wide acknowledgement of these tensions, there are relatively limited studies or even policy discussions on the specific needs and challenges of regulation at European Union (EU) level of AI in law enforcement and criminal justice. These possible specific needs and challenges often

¹ 'AI can ... help to detect fraud and cybersecurity threats, and enables law enforcement authorities to fight crime more efficiently' (COM(2019) 168 final, 1).

appear as a concluding thought, if not an afterthought, amidst broader reflections related to the regulation of AI in general. Almost as an exception to this rule, the European Parliament is currently discussing a ***Draft Report on Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminal matters*** (LIBE, 2020/2016(INI)).

At EU level, deliberations on a possible need for an AI regulatory framework have primarily surfaced and been developed until now in a **Digital Single Market** context. This is not without consequences on the nature and substance of the debate, and has notably tended to leave out of the picture the singularity of law enforcement and criminal justice, both regarding the specific risks in the field and the peculiarities of the EU legal framework in the **Area of Freedom, Security and Justice (AFSJ)**.

If EU-level policy debates have only touched upon the challenges of AI in law enforcement and criminal justice in a limited manner, many of these challenges are nevertheless already being visibly surfacing, in an often-contentious mode, in EU territory and globally. This is most probably connected to the fact that law enforcement and criminal justice have a bearing on foundational aspects of the relations between individuals and the State, relations that are thus particularly sensitive to the deep transformations potentially conveyed by AI systems. It might also be, however, that the manner in which certain related technologies have been deployed was not the best suited manner, and that different approaches – and possibly, indeed, a revised regulatory framework – would contribute to mitigate fundamental rights risks, ease critical concerns, and increase trust in these technologies.

The European Commission noted already in 2014 that although digitisation of public services opened up new opportunities to optimise data analysis, the *'reported use of similar technologies for surveillance purposes, by public or private actors, is liable to feed concern and reduce trust in the digital economy among individuals and organisations'*, deserving thus special attention (COM(2014) 442 final, 3). The amplification of widespread surveillance is, precisely, one of the impacts of the advent of AI in certain countries such as the United States and China, according to some observers (AI Now Institute, 2018).

Much has been said about 'the race for AI dominance', or the 'race of AI', typically referring to the capacity of specific actors to take economic advantage of AI developments.² In the context of law enforcement and criminal justice, the most important ongoing race might be the one between legislative efforts to appropriately frame and counter the risks triggered by AI, on the one hand, and judicial challenges and other 'bottom up' developments that bring to the fore a variety of fundamental issues that appear to be requiring, if not further protection by the law, at least, urgently, more clarity.³

1.1. Objectives and structure

This study aims at critically assessing ongoing developments in relation to **AI in the field of law enforcement and criminal justice**, in order to review their **impact on EU fundamental rights** and, on the basis of such an assessment, put forward policy recommendations. For this, it first briefly introduces the relevant legal framework, and then reviews the main trends and controversies in the area, focusing in particular in 'predictive policing', facial recognition, AI and criminal justice, and AI and border control. It then analyses their impact on EU fundamental rights, before moving to pertinent policy discussions around the regulation of AI, in a section also examining the support of AI in law enforcement and criminal justice through EU funding. The study also introduces some of the main relevant developments that have taken place following the outbreak of the Covid-19 virus. It concludes by putting forward a series of policy recommendations.

² See, for instance: Castro, McLaughlin & Chivot, 2019.

³ In relation to this 'competition' and the regulation of facial recognition, see, for instance: Lynch, 2020.

The notion of AI is here **broadly understood**, in order to offer a wide view of the relevant debates and challenges. Legal and policy discussions often tend to soften the boundaries between certain concomitant notions, and issues which are of interest for the purposes of a discussion of AI in the field of law enforcement and criminal justice have sometimes manifested themselves relying on different terminology (for instance in debates around 'algorithmic discrimination', 'accountable algorithms', 'algorithmic transparency', etc.).

The study does not have the ambition of exhaustively covering all recent and ongoing developments in this area, which are numerous. The EU Agency for Fundamental Rights has, in this sense, compiled a collection of almost 300 policy initiatives in EU Member States and beyond for the period 2016-2020 (EU FRA, April 2020). Researchers working on AI regulation have also stressed the difficulty of having a full overview of the many initiatives concerning the ethical, social and legal aspects of AI (Boddington, 2017, 3).

The study also notably does not cover issues such as the misuse of AI for criminal purposes. When AI is used maliciously (Brundage, 2018), additional challenges for fundamental rights can be triggered.

1.2. Methodology

The study is based on desk-research and the review of existing available data, studies and analyses from sources and documents from EU, national and international institutions, as well as Data Protection Authorities (DPAs), experts, academia, industry, and civil society. It principally encompasses legal instruments and policy documents of EU institutions, bodies and agencies, but also from the Council of Europe. The analysis of EU primary and secondary legislation constitutes an essential part of the research. Various reports have also been used to the extent they are useful to illustrate recent trends and controversies.

2. EU LEGAL FRAMEWORK

Before considering the impact on fundamental rights of AI in the field of law enforcement and criminal justice it is useful to briefly recall which are the applicable standards in the EU. This section succinctly examines the main elements of the relevant **EU legal framework**, touching upon both primary and secondary law.

There are actually **many legal instruments** that might be regarded as directly or indirectly applicable to AI-related developments.⁴ For the purposes of this study, special attention shall be given here to EU fundamental rights obligations, but also to rules on data and its free movement, in light of the importance of data as both an enabler and a product of AI solutions.

It is important to remember however that EU law obligations coexist with other fundamental rights and data protection obligations, such as those derived from national legal orders. In addition, Council of Europe instruments also deserve a special mention, most notably the **European Convention on Human Rights (ECHR)** and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (**'Convention 108'**),⁵ which was recently reviewed leading to a modernised version, known as **'Convention 108+'**.⁶

These **Council of Europe** instruments are accompanied by soft-law instruments such as the *Guidelines on Artificial Intelligence and Data protection* published by the Consultative Committee of Convention 108 in 2019.⁷ Previously, in 2018, the Consultative Committee of Convention 108 had also adopted *The Practical Guide on the Use of Personal Data in the Police Sector*, offering guidance on how to legally prevent and combat crime, including through the use of personal data (Consultative Committee of Convention 108, 2018). The Guide warns that there are '*considerable risks*' linked to using big data technologies and analysis techniques to assist crime detection, and identifies a series of requirements to which data controllers are advised to '*pay additional attention*' when processing personal data for such purposes, such as special awareness requirements (ibid., 10-11).

2.1. Primary law

As established in Article 2 of the Treaty of the EU, the EU '*is founded on the values of respect for **human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities***'. Article 6 of the Treaty of the EU specifies which are the fundamental rights recognised as such by EU law, referring to the future accession to the European Convention on Human Rights (ECHR), the fundamental rights which are part of the general principles of EU law, and the fundamental rights recognised in the **Charter of Fundamental Rights of the EU** – such as human dignity (Article 1), the right to life (Article 2), the right to the integrity of the person (Article 3), the right to liberty and security (Article 6), the right to respect for private and family life (Article 7), the right to the protection of personal data (Article 8), freedom of expression and information (Article 11), freedom of assembly and of association (Article 12), equality before the law (Article 20), non-discrimination

⁴ Such as, for instance, EU cybersecurity law (Hamon, Junklewitz & Sánchez, 2020, 4).

⁵ Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS No.108.

⁶ See Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, CETS No.223.

⁷ Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108), *Guidelines on Artificial Intelligence and Data Protection*, T-PD(2019)01.

(Article 21), the rights of the child (Article 24), the right to an effective remedy and to a fair trial (Article 47), and the presumption of innocence and right of defence (Article 48).

All these fundamental rights are applicable also in relation to law enforcement and criminal justice; fundamental rights requirements derived from primary law apply to all areas of EU law. In relation to the right to the protection of personal data of Article 8, it is worth noting that a Declaration on the protection of personal data in the fields of judicial cooperation in criminal matters and police cooperation attached to the Lisbon Treaty observed that *'specific rules on the protection of personal data and the free movement of such data in the fields of judicial cooperation in criminal matters and police cooperation based on Article 16 of the Treaty on the Functioning of the European Union may prove necessary because of the specific nature of these fields'*.⁸

2.2. Secondary law

In this section are presented the most salient elements of EU secondary law applying to **the processing of both personal data and non-personal data**. The distinction between personal and non-personal data is not deprived of frictions, especially as technological developments – including AI-related developments – might oblige to regard as personal data certain data sets that were not previously regarded as such.

2.2.1. Rules on the processing of personal data

Personal data are defined in EU law as **'any information relating to an identified or identifiable natural person** (*'data subject'*); *an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person'*.⁹

The right to personal data protection is currently recognised as an **EU fundamental right** both in the EU Charter of Fundamental Rights (Article 8) and in the Treaties (Article 16 of the Treaty on the Functioning of the EU). This right is currently recognised as an **autonomous fundamental right**, different from the right to the respect for private life, to which it has been historically closely connected in some legal frameworks. This does not mean, however, that such as right does not serve at the same time **other fundamental rights**, including the right to respect for private life. In this sense, EU secondary law giving substance to Article 8 of the EU Charter simultaneously aims at guaranteeing the respect of **all fundamental rights**.¹⁰

The EU legal framework on data protection is composed of a number of different legal instruments, of which must be mentioned:

- Regulation (EU) 2016/679,¹¹ or **General Data Protection Regulation (GDPR)**, named 'general' because it applies 'generally' to the processing of personal data, within the limits of its material

⁸ Declaration 21 of the Declarations Annexed to the Final Act of the Intergovernmental Conference which adopted the Treaty of Lisbon, signed on 13 December 2007.

⁹ Article 4(1) of the General Data Protection Regulation (GDPR).

¹⁰ This has practical implications for the interpretation of the GDPR. Its Article 35, for instance, refers to the need to take into account a likely high risk 'to the rights and freedoms of individuals'. This has been interpreted as primarily concerning the rights to data protection and privacy, but as also potentially involving other fundamental rights such as freedom of speech, freedom of thought, freedom of movement, prohibition of discrimination, right to liberty, conscience and religion (Art. 29 WP, WP 248 rev.01, 6).

¹¹ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), OJ L 119, 4.5.2016, p. 1–88.

and territorial scope (but does not apply to personal data processing by EU institutions, bodies, offices and agencies);

- Directive (EU) 2016/680,¹² known as the **Law Enforcement Directive (LED)**, which applies generally to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, but only when these authorities process personal data for such purposes (but does not apply to personal data processing by EU institutions, bodies, offices and agencies);
- Regulation (EU) 2018/1725,¹³ or **EU-DPR**, which is generally applicable to the processing of personal data by EU institutions, bodies, offices and agencies, but includes special rules for the processing of 'operational personal data',¹⁴ for instance by Eurojust, and leaves out of its scope Europol and the European Public Prosecutor's Office, pending a possible change in 2022;
- Directive 2002/58/EC,¹⁵ or **e-Privacy Directive**, applying to the processing of personal data in connection with the provision of publicly available electronic communications services in public communications networks.

In addition to such instruments, other legal instruments include particularly important data protection rules applying to specific data processing activities in this field, such as, for instance:

- Regulation (EU) 2016/794,¹⁶ or the **Europol Regulation**, on the EU Agency for Law Enforcement Cooperation (Europol) – regarded as '*a hub for information exchange in the Union*',¹⁷ which has its own data protection provisions;
- Council Regulation (EU) 2017/1939,¹⁸ of **EPPO Regulation**, on the European Public Prosecutor's Office, which equally has its own data protection provisions;
- Regulation (EU) 2018/1727,¹⁹ or the **Eurojust Regulation**, on the European Union Agency for Criminal Justice Cooperation (Eurojust), which has data protection rules to be regarded as *lex specialis* to the relevant provisions of the EU-DP Regulation;

¹² Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA [2016] OJ L 119/89, p. 89–131.

¹³ Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC, OJ L 295, 21.11.2018, p. 39–98.

¹⁴ 'Operational personal data' are defined all personal data processed by EU bodies, offices or agencies when carrying out activities which fall within the scope of Chapter 4 (Judicial cooperation in criminal matters) or Chapter 5 (Police cooperation) of Title V of Part Three TFEU to meet the objectives and tasks laid down in the legal acts establishing those bodies, offices or agencies (Art. 3(2) of Regulation (EU) 2018/1725).

¹⁵ Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications services (Directive on privacy and electronic communications), OJ L 201, 31.7.2002, p. 37–47, as subsequently amended.

¹⁶ Regulation (EU) 2016/794 of the European Parliament and of the Council of 11 May 2016 on the European Union Agency for Law Enforcement Cooperation (Europol) and replacing and repealing Council Decisions 2009/371/JHA, 2009/934/JHA, 2009/935/JHA, 2009/936/JHA and 2009/968/JHA, OJ L 135, 24.5.2016, p. 53–114.

¹⁷ Regulation (EU) 2016/794 of the European Parliament and of the Council of 11 May 2016 on the European Union Agency for Law Enforcement Cooperation (Europol) and replacing and repealing Council Decisions 2009/371/JHA, 2009/934/JHA, 2009/935/JHA, 2009/936/JHA and 2009/968/JHA, OJ L 135, 24.5.2016, p. 53–114.

¹⁸ Council Regulation (EU) 2017/1939 of 12 October 2017 implementing enhanced cooperation on the establishment of the European Public Prosecutor's Office ('the EPPO'), OJ L 283, 31.10.2017, p. 1–71.

¹⁹ Recital (12) of (EU) 2016/794.

The panorama of EU data protection law is actually **more complex** than what this short presentation might suggest. Even if the establishment of '*a harmonised framework for all data processing activities by law enforcement authorities for law enforcement purposes*' has been a major, recurrent concern of the European Parliament for many years (EP, 2014, 15), EU law does not currently offer a fully homogenous outlook. The approach notably enshrined by the LED in conjunction with the EU-DPR, whereby certain EU agencies processing data for law enforcement purposes are excluded from the reach of general provisions in the field, has been described as leading '*to a multi-level data protection regime where different legal instruments, and therefore different standards affecting individuals in exercising their data protection rights, apply*' (Belfiore, 2013, 367). There exist also a number of specific provisions on the protection of personal data in certain EU instruments that remained unaffected by the entry into force of the LED and have not been revised since – specific provisions for the protection of personal data that had entered into force before May 2016 in the field of judicial cooperation in criminal matters and police cooperation. In June 2020, the European Commission announced its action plan to progressively align with the LED the provisions still requiring alignment, a total of 10 according to its assessment (COM(2020) 262 final).

Often, discussions on the possible need for future EU-level regulation of AI take as a starting point an **explicit or implicit assumption according to which the general standards of the GDPR are the central element** of the current legal framework, if not the only element worth being reviewed (Penner, 2019). When focusing on personal data processing for law enforcement and criminal justice, taking such starting point is, however, unsuited, for two main reasons:

- first, the GDPR might in many relevant cases not apply, as other instruments will be applicable;
- second, even when the GDPR does apply, restrictions grounded on the fact that the processing relates to law enforcement and criminal justice might apply, *de facto* modulating the general safeguards it foresees.

It is thus crucial to appropriately situate discussions on AI regulation in this field by grounding them on a more refined understanding of applicable rules. The Annex to this study provides an overview of the some of the most relevant provisions of the GDPR and the LED in relation to algorithmic data processing, illustrating how they are not fully coincidental.

As explained by the European Data Protection Supervisor (EDPS), in the field of law enforcement '*data processing activities used are often opaque to individuals, which makes it difficult for them to know who is processing their data and for what purposes*', and '*(i)ndividuals' rights are often restricted*' in spite of the fact that '*the impact of data processing activities on their rights and freedoms is significant*' (EDPS, 2019, 38).

An additional challenge from a data subject's perspective is the fact that to some data processing activities might apply intricate combinations of different provisions, depending on the exact purpose of each processing operation and the identity of the controller.²⁰ In the AFSJ, the fact that data processing activities related to some information systems are governed by '*complex data protection regimes due to their multiple possible uses*' can result in a problematic lack of clarity (FGB, 2019, 4).

The question of whether the **GDPR** as such offers appropriate protection for individuals in the face of the proliferation of automated decision making is in itself not settled (Penner, 2019), and it is generally apprehended as an extremely delicate question, to the extent that it could lead to a push for a review of the GDPR eventually worsening its standards (De Brouwer, 2020). In relation to the safeguards useful

²⁰ See, for instance, the case of the *European Travel Information and Authorization System* (ETIAS) described below.

in an AI context, an often-quoted provision of the GDPR is its **Article 22**, devoted to '**automated individual decision-making, including profiling**'.²¹ Many, including the European Commission, refer specifically to Article 22 of the GDPR when considering the future of AI in Europe (COM(2019) 168 final, 4).

There is however much disagreement about the exact scope of Article 22 of the GDPR, and on its very nature - whether it imposes a general ban on certain types of automated decisions, or whether it provides data subjects with some rights whether certain types of automated decisions are taken (Brkan, 2019, 98; A29WP, 2017/2018, 19). There is equally controversy on the legal implications of Article 22 of the GDPR. Finally, there is also controversy on whether the provision is actually of any use in practice – whether it is effectively applied. The existence of a '**right to explanation**' under the GDPR has also been widely debated, leading for instance to the assertion that a legal basis for such a right is 'rather shaky and inconclusive' (Temme, 2017, 482). These issues have generated what some have described as '*vast debate in the research community*', intertwined with a similarly vast debate surrounding the very explainability of AI systems (Sartor and Lagioia, 2020, 54).

Whenever some of the mentioned instruments apply, the fact that certain data processing operations relate to **law enforcement and criminal justice** can have important consequences on the scope of applicable rights and obligations.²² In general terms, **exemptions and derogations based on law enforcement grounds are permitted** in EU data protection law, to the extent that they would be necessary and proportionate and if accompanied by safeguards. In this sense, Article 23 of the GDPR foresees that both EU and national laws may restrict a the scope of certain rights and obligations, including the mentioned **Article 22 of the GDPR**, if necessary for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security.²³

Article 23 of the GDPR, described as '*highly controversial from the beginning*', authorises deviations '*from core contents of the GDPR*', even if '*with equally high requirements*' (Wagner & Benecke, 2016, 357). There is at the moment no official overview of the use at EU and national level of the exemptions and derogations grounded on this provision.²⁴ The rights granted to data subjects under the **LED** can also be restricted, by means of legislative measures, in order '*to avoid obstructing official or legal inquiries*,

²¹ Profiling is defined in Article 4(4) of the GDPR as '*any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements*'.

²² The e-Privacy Directive, in this sense, foresees in its Article 15(1) that 'Member States may adopt legislative measures to restrict the scope' of certain rights and obligations when such restrictions '*constitutes a necessary, appropriate and proportionate measure within a democratic society to safeguard national security (i.e. State security), defence, public security, and the prevention, investigation, detection and prosecution of criminal offences*'.

²³ Article 23(1)(d) of the GDPR.

²⁴ It is known that in any case the majority of Member States have made use of it (cf., for instance, Section 41 of the Dutch General Data Protection Regulation Implementation Act). Typically law enforcement considerations emerge: for example, in the Czech Republic, when what are at stake are 'protected interests' such as the prevention, investigation or detection of criminal offences, prosecution of criminal offences, execution of criminal penalties and protective measures, compliance with obligations related to the exercise of any of the data subject rights from Art. 12 to 22 of the GDPR might be postponed (*Zákon ze dne 12. března 2019 o zpracování osobních údajů* (original); Czech Act No. 110/2019 Coll., act of 12 March 2019 on personal data processing, Section 6(2)(b) and Section 11(1)). To allow for this it is simply required to notify the DPA, a notification which would nonetheless not be necessary if the processing to which this applies is carried out by courts (*ibid.*, Section 11(2)).

investigations or procedures, to avoid prejudicing the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties'.²⁵

There is significant case law by the **Court of Justice of the EU (CJEU)** on EU data protection law. The Court has notably emphasised the need to interpret its provisions in a manner that guarantees the 'effective and complete protection' of data subjects.²⁶ Particularly relevant here is the case law about the obligations imposed on the providers of public telecommunications services and networks with respect to **data retention** in order to ensure that data related to communication are available for the purpose of the investigation, detection and prosecution of serious crime, as often AI solutions in the field of law enforcement and criminal justice rely on the use for such purposes of data originally processed for other, sometimes unrelated, purposes, and involve both private companies and public authorities.

Notable in this context was the judgment of the CJEU in *Digital Rights Ireland*,²⁷ which annulled Directive 2006/24/EC²⁸ (known as the 'Data Retention Directive'), which, building on Article 15(1) of the e-Privacy Directive, imposed EU-wide restrictions of the rights and obligations set by such instrument in a way that the Court described as exceeding '*the limits imposed by compliance with the principle of proportionality in the light of Articles 7, 8 and 52(1) of the Charter*' (paragraph 69). Concretely, the CJEU decreed that Directive 2006/24/EC covered, '*in a generalised manner, all persons and all means of electronic communication as well as all traffic data without any differentiation, limitation or exception being made in the light of the objective of fighting against serious crime*' (paragraph 57), and warned that '*the fact that data are retained and subsequently used **without the subscriber or registered user being informed** is likely to generate in the minds of the persons concerned **the feeling that their private lives are the subject of constant surveillance***' (paragraph 37).²⁹

In spite of its importance for individuals, the protection of personal data processed for law enforcement and criminal justice has received relatively **limited attention** from the European Commission, at least

²⁵ Recital (44) of the LED. The Czech instrument transposing the LED, to continue with a similar example, foresees that competent authorities might not comply with requests to exercise the right of access if doing so would endanger the performance of a task in the area of prevention, investigation and detection of criminal offences, prosecution of criminal offences, execution of criminal penalties and protective measures (ibid., Section 28(2)). As a safeguard is imposed the obligation to keep a record for the reasons justifying such decision, for a period of three years (Section 28(4)).

²⁶ Cf. *Google Spain*, Judgment of the Court (Grand Chamber) of 13 May 2014, Case C-131/12, ECLI:EU:C:2014:317, paras. 34, 38, 53, 58.

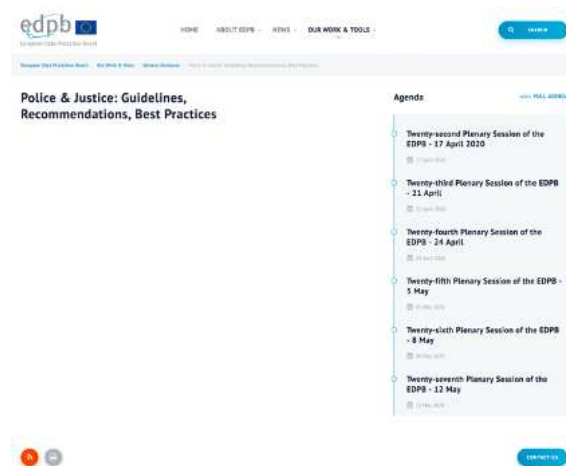
²⁷ *Digital Rights Ireland*, Judgment of the Court (Grand Chamber) of 8 April 2014, Joined Cases C-293/12 and C-594/12, ECLI:EU:C:2014:238.

²⁸ Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, OJ L 105, 13.4.2006, p. 54–63.

²⁹ In the subsequent *Tele2 Sverige and Watson* judgment, the CJEU held that the Member States cannot impose on the providers of electronic communication services an obligation of general and indiscriminate retention of data, noting that EU law, read '*in the light of Articles 7, 8 and 11 and Article 52(1) of the Charter, must be interpreted as precluding national legislation which, for the purpose of fighting crime, provides for the general and indiscriminate retention of all traffic and location data of all subscribers and registered users relating to all means of electronic communication*' (Judgment of the Court (Grand Chamber) of 21 December 2016, Joined Cases C-203/15 and C-698/15, ECLI:EU:C:2016:970, para. 112). Since then, the CJEU has been regularly confronted with further cases around data retention, and has notably still to adjudicate on relevant references for a preliminary ruling, sent by the French *Conseil d'État* (Joined Cases C-511/18 and C-512/18), the *Cour constitutionnelle de Belgique* (Case C-520/18) and the Investigatory Powers Tribunal (UK) (Case C-623/17) in which the primary issue is the application of the e-Privacy Directive to activities relating to national security and combatting terrorism – notably in light of Article 4 of the Treaty of the EU, under which national security is the exclusive responsibility of each Member State. In February 2020, the Irish Supreme Court discussed the submission of a request for a preliminary ruling in the case *Dwyer v Commissioner of An Garda Síochán* (Judgment of Mr. Justice Clarke, Chief Justice, delivered the 24th of February, 2020, Supreme Court, *Graham Dwyer vs. the Commissioner of An Garda Síochána*, Record No: 2019/18), also about data retention, safeguards related to the access to data, and the validity of convictions relying on data obtained through data retention schemes (Hennessy, 2020).

compared to the GDPR as such. The work of the European Data Protection Board (EDPB) in this area has also been until now noticeably weak, as illustrated by the absence of published guidelines, recommendations or best practices on its official website (see Figure 1, below).

Figure 1: EDPB work on police and justice



Source: Screenshot of https://edpb.europa.eu/our-work-tools/general-guidance/police-justice-guidelines-recommendations-best-practices_en [accessed 7 July 2020].

2.2.2. Rules on the processing of non-personal data

AI is explicitly mentioned in Regulation (EU) 2018/1807 on the **Free Flow of Non-Personal Data in the EU**,³⁰ which aims at ensuring the free flow of non-personal data within the EU by regulating data localisation requirements, the availability of such data to competent authorities, and the porting of data for professional users. Recital (9) of that Regulation states indeed that **AI**, together with the Internet of Things and machine learning, represents a major source of non-personal data, which might take the form of *'aggregate and anonymised datasets used for big data analytics'*.

The Regulation on Free Flow of Non-Personal Data does not define 'non-personal data', but uses this term as equivalent to *'electronic data other than personal data'* (Article 2(1)). It acknowledges there might exist data sets in which personal and non-personal data are *'inextricably linked'*, in which case the its provisions shall not prejudice the application of the GDPR (Article 2(2)). The European Commission published Guidance on the interaction between this Regulation and the GDPR, which clarifies that the notion of 'non-personal data' in the Free Flow of Non-Personal Data must be defined by opposition (a *contrario*) to personal data as laid down by the GDPR, and that there can be both 'non-personal data' which were originally as such, and data which were first personal but became 'non-personal' after anonymisation (COM(2019) 250 final, 4-5). The Guidance concedes that *'(i)n most real-life situations, a dataset is very likely to be composed of both personal and non-personal data'* (ibid., 4).

The key principle of the Regulation on the Free Flow of Non-Personal Data is that such non-personal data must **flow freely in the EU**. Data localisation requirements, referring to Member States imposing the processing of data in their territory or hindering the processing of data in another Member State,

³⁰ Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, OJ L 303, 28.11.2018, p. 59–68.

are in principle not permitted, unless justified on the grounds of public security (encompassing the need to facilitate the investigation, detection and prosecution of criminal offences) in compliance with the principle of proportionality. Regardless of the localisation of data, persons subject to obligations to provide data to competent authorities shall comply with such obligations by providing and guaranteeing effective and timely electronic access to the data to competent authorities.³¹ If said persons would fail to comply, national competent authorities shall provide assistance to each other, if appropriate under instruments in the area of police cooperation and criminal justice such as the Council Framework Decision 2006/960/JHA (known as 'the Swedish initiative'),³² Directive 2014/41/EU on the European Investigation Order (EIO),³³ and the Cybercrime Convention.³⁴

All in all, the Regulation represents an important step in making available non-personal data across internal borders. Although it exceptionally allows for data localisation requirements to be established on the basis of security grounds, it nevertheless also foresees the need to comply with certain obligations to provide data across borders and reaffirms the relevance of mechanisms allowing for **cross-border access to data**, including **for law enforcement purposes**.

³¹ Recital (25) of the Regulation on Free Flow of Personal Data.

³² Council Framework Decision 2006/960/JHA of 18 December 2006 on simplifying the exchange of information and intelligence between law enforcement authorities of the Member States of the European Union, OJ L 386, 29.12.2006, p. 89-100.

³³ Directive 2014/41/EU of the European Parliament and of the Council of 3 April 2014 regarding the European Investigation Order in criminal matters, OJ L 130, 1.5.2014, p. 1-36.

³⁴ Convention on Cybercrime of the Council of Europe, CETS No 185.

3. AI IN LAW ENFORCEMENT AND CRIMINAL JUSTICE

This section provides an overview of the main current trends related to AI in the field of law enforcement and criminal justice, as relevant for an EU-level discussion of their impact on fundamental rights. They are clustered around four major themes: 'predictive policing', facial recognition, AI and criminal justice, and AI and borders.

The notion of AI espoused for these purposes is deliberately **wide**. In line with ongoing policy developments at EU level, the notion embraced here recognises as key components of AI developments their reliance on '**data**', '**algorithms**' and '**computing power**'.³⁵ Decomposing AI into these three basic elements allows to better perceive the continuities and discontinuities between AI and other notions that have previously instigated policy change and discussions in the area, such as '**smart surveillance**', '**profiling**', or '**big data**'.

The term 'smart surveillance' places the emphasis on the fact that new solutions are supposed to be 'intelligent', in the sense that they appear to 'learn' from data observed, as well as 'take decisions' potentially 'on their own' on the basis of certain algorithms. In this sense, it has been stated that what is at the core of EU-level discussions around AI are as a matter of fact '*algorithmically controlled ADM [Automated Decision Making] systems*' (Penner, 2019).

Generally speaking, the word 'profiling' is connected to the sorting out of individuals based on the '*computer analysis of large data sets*', and is as such already commonly used by law enforcement officers and border guards '*to prevent, investigate and prosecute criminal offences, as well as to prevent and detect irregular immigration*' (FRA, 2018a, 7-8).

The term 'big data' is typically used to evoke the processing of vast data sets, including data sets with data of heterogeneous origins. The EDPS noted that in its recent *White Paper on AI* the European Commission relied on a conception of AI that is broad, and that can be interpreted as encompassing what others refer to as 'big data' (EDPS, Opinion 4/2020, 7). In the past, 'big data' had been associated with changes of potentially profound repercussions - portrayed as tearing down what counts as crime-relevant knowledge, what counts as proper reasoning, and even how should crime be prevented, investigated and prosecuted (Završnik, 2017a, 3). In any case, the development of AI systems, in particular those based on machine learning, both presupposes and fosters the creation of vast data sets that can be connected to 'big data' (Sartor & Lagioia, 2020, i), and today's AI development is largely driven by data (Boucher, 2019, 5).

The term AI as emerging in these debates is, all in all, both '*poorly defined and potentially misleading*' (Babuta, Oswald & Rinik, 2018, 2). Furthermore, the field of AI being a very dynamic field, the most rigorous efforts to define AI and provide an operational definition and related taxonomies and keywords stress the necessity to do so always in the context of a dynamic process, subject to regular reviews and iterations (Samoili et al., 2020, 6).

Situating AI in the context of other, pre-existing policy discussions does not aim at negating its innovative dimension, but should help to perceive that debates around AI are not completely unique, and have not arisen in a void space. They surfaced among many other developments encouraging reliance on computers and algorithms, and supporting wide access to data, also insofar as law enforcement and criminal justice are concerned.

³⁵ Cf. the broad vision entertained by the EC according to which '*AI is a collection of technologies that combine data, algorithms and computing power*' (COM(2020) 65 final, 2).

The processing of data for law enforcement purposes of data not originally collected for such purposes is an issue that pops up regularly, also beyond AI discussions.³⁶ The constant multiplication of data in our societies has indeed been accompanied over the years with a sustained interest among the law enforcement community to **access any possibly relevant data**. Such interest includes, for instance, an interest in data held by private companies, including data that might require crossing national borders (Carrera & Stefan, 2020, 3).

Finally, AI-related developments in this field also need to be placed against the background of a progressive 'digital transformation' of the state. The United Nations (UN) Special Rapporteur on Extreme Poverty and Human Rights, in this sense, described in 2019 the emergence of the 'digital welfare state' in many countries across the globe, as systems of social protection and assistance are increasingly driven by technologies '*used to automate, predict, identify, surveil, detect, target and punish*' (UN Special Rapporteur on Extreme Poverty and Human Rights, 2019, 1).

3.1. 'Predictive policing'

'Predictive policing' has been described as a catchphrase relying broadly on the claim that through the algorithmic processing of data sets it is possible to **reveal patterns of probable future offending and victimisation**, which can thus be interdicted before they happen (Wilson, 2017, 108). Although the origins of predictive policing might be traced back to experiments of computer-assisted crime control in the 1970s, and although prediction in the area of crime and punishment has been developed and discussed more many decades (Harcourt, 2008), the term became eventually connected to the rise of 'big data' (Wilson, 2017, 109). **Prediction as an extensive trend** has noticeably marked security developments globally and in Europe over the past decade at least, typically associated to **prevention**.

'Predictive policing' can take a variety of shapes, and multitude of typologies have been put forward. Predictive policing methods can for instance be divided into four broad categories: methods aiming at predicting **crimes**, or forecasting places and times with an increased risk of crime; methods aiming at predicting **offenders**, or identifying individuals at risk of offending (or reoffending) in the future; methods aiming at predicting **perpetrators' identities**, or creating **profiles** similar to those of past offenders, and methods aiming at predicting **victims of crimes**, used to identify groups or individuals who are likely to become victims of crime (Perry et al., 2013, xiv). Typologies can also be grounded on the possible purposes of algorithmic data or intelligence analysis within the policing context: it is then possible to distinguish predictive policing **on a macro level** incorporating strategic planning, prioritisation and forecasting; and **operational intelligence** linking and evaluation which may include, for instance, crime reduction activities; and decision-making or risk-assessments relating to **individuals** (Oswald & Grace, 2016, 3).

The NGO Electronic Privacy Information Center (EPIC) in 2020 a report dating from 2014 from the US Department of Justice to former President Obama, about predictive analytics and algorithms in law enforcement.³⁷ The report states that the use of 'social network analysis' was gaining '*gaining the attention of some State and Local law enforcement agencies*' (US Department of Justice, 2014, 3). Social network analysis was by then starting to be used to **identify individuals** who may then '*receive increased scrutiny, additional support services, or both*' (idem). According to the report, most of the US Federal Bureau of Investigation (FBI) data analysis targeted particular subjects, although in some cases

³⁶ The EDPS, for instance, expressed concern in July 2020 about the possibilities granted by Microsoft to possible access to data connected to the use by EU institutions of its products and services to third parties including law enforcement or other government agencies (EDPS, *Outcome of...*, 24).

³⁷ The report was obtained following a Freedom of Information Act (FOIA) request, lawsuit, and negotiated settlement: EPIC v. DOJ, No. 18-5307 (D.C. Cir.).

it was used *'to identify sub-sets of persons who may bear further investigative scrutiny'* (ibid., 8). Information used for such data analysis was described as originating in a *'a variety of data sources'*, including commercial databases, and encompassing *'among other things, biographical information, biometric information, financial information, location information, associates and affiliations, employment and business information, visa and immigration information, travel information, and criminal and investigative history information'* (idem).

'Predictive policing' initiatives have popped up in different Member States mainly since the beginning of the 2010s these have been, for instance, initiatives somehow connected to or involving automatic license plate control systems (Alfter, 2019, 52; Van Brakel, 2019, 44). The expansion of police use of algorithms in the recent years has been connected to three factors: austerity measures that limit resources and push towards apparently more economically efficient new solutions; an increasing perception that police ought to adopt a preventative posture, placing emphasis on anticipating harm, and an increase in the volume and complexity of data available, necessitating increasingly sophisticated processing tools (Babuta & Oswald, 2020, vii).

Prediction and prevention have been particularly productive in relation to **money laundering and terrorist financing**.³⁸ Current approaches in this field are indeed built upon obligations imposed on banks and other obliged entities to take appropriate steps to identify and assess the risks of money laundering and terrorist financing, taking into account **risk factors** including those relating to their customers, countries or geographic areas, products, services, transactions or delivery channels. This risk assessment by the entities acting as 'gatekeepers' can eventually lead to the reporting of transactions regarded as **suspicious** to competent EU Financial Intelligence Units (FIUs), which have their own reporting and data sharing obligations. The eventual labelling of certain individuals as **suspects** during this process can lead to situations requiring special attention, raising issues of data protection but also potentially impacting the right to presumption of innocence (Article 48 of the EU Charter of Fundamental Rights).³⁹

Predictive approaches in law enforcement often rely at least partially on the processing of data that is not as such originally related to crime, but initially collected by **private companies** in the context of their normal business activity (e. g. banking, telecommunications, travelling). Predictive policing schemes also typically rely on predictive software produced by private companies, be it vendors specialising in this field (the most often quoted being probably the company PredPol) or large technological corporations (Wilson, 2017, 114).

In the US, the media have reported practices by private companies targeting the **AI market** that would actively involve collecting data in a not particularly transparent manner for, or in view of eventually making it available to, law enforcement authorities (Koebler, Maiberg & Cox, 2020). For instance, in March 2020 surfaced accusations against an AI company, Banjo, of designing apps specifically conceived to covertly collect social media data (Cox & Koebler, 2020).

Many relevant developments have materialised in the United Kingdom (UK). Especially interesting are discussions around the **Gangs Violence Matrix (GVM)**, a tool used since 2012 by the Metropolitan

³⁸ See, notably: Directive (EU) 2015/849 of the European Parliament and of the Council of 20 May 2015 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, amending Regulation (EU) No 648/2012 of the European Parliament and of the Council, and repealing Directive 2005/60/EC of the European Parliament and of the Council and Commission Directive 2006/70/EC, OJ L 141, 5.6.2015, p. 73–117. Also referring to money laundering and terrorist financing: LIBE, 2020/2016(INI), 5.

³⁹ Data protection issues were at the forefront of the ban imposed by the EDPS on 19 December 2019 on processing operations carried out by Europol in the technical operation of FIU.net, a decentralised information network designed to support national FIUs (EDPS, 2019, 41).

Police in order to identify and risk-assess gang members⁴⁰ across London involved in gang violence, and to identify those at risk of victimisation. The GVM measures the harm that individuals pose by scoring them based on evidence of them committing violence and weapons offences, as well as police intelligence relating to them having access to weapons, or them being involved in, or at risk from, gang violence: individuals are thus graded as red, amber or green, denoting the level of risk (for victims) or harm (for offenders) they present.⁴¹ The database encompasses both adults and minors (MOPAC, 2018, 4).

According to the NGO Amnesty International, the Metropolitan Police has refused to divulge information about the precise criteria used to assign automated 'harm scores' to individuals on the matrix, and it would also be unclear how such scores might relate to any enforcement action (Amnesty International, 2018, 13-14). More generally, the NGO has denounced that the data processing practices around the database and the lack of proper safeguards generate a risk that the tool would '*discriminate against already marginalised young people, with disproportionate impact on black boys and young men*' (ibid., 3). An enforcement notice was served on the Metropolitan Police Service by the Information Commissioner's Office (ICO) in 2018, relating to lack of compliance with data protection law. Since then, a significant number of individuals are believed to have been removed from the database (Dodd, 2020).

'Predictive policing' initiatives can sometimes rely on the use of facial recognition technologies, which in their turn rarely develop separately from other developments more broadly connected to the monitoring of public spaces, for instance to detect 'abnormal behaviour' (e. g., Kayser-Bril, 2020).

3.2. Facial recognition

Facial recognition technology allows for the automatic identification – or authentication - of an individual by matching facial features from two or more digital images (EU FRA, 2019, 2). Under EU data protection law, even if the processing of photographs as such is not systematically considered as processing deserving **special protection** beyond general data protection rules, when data are processed in order to allow for such identification or authentication they are to be regarded as biometric data deserving special protection.⁴²

The use of facial recognition technologies is not limited to the law enforcement and criminal justice field. Facial recognition has, a matter of fact, generated much interest and disputes globally also in other fields, as illustrated by a class-action lawsuit in the US against Facebook's automated photo-labelling and its use of face-matching software to suggest names of people in users' photos.⁴³

From a global perspective, the uses of facial recognition by some specific states have been particularly controversial; in this sense, must be mentioned international concerns about the use of facial recognition in China to identify members of the Uighur minority, a largely Muslim group (Mozur, 2019). Facial recognition in public spaces is said to be growing particularly rapidly in Russia (Light, 2019).

In the field of law enforcement, the use of facial recognition that has gathered more attention concerns **live facial recognition**.

In 2016, a **US** report warned of the widespread use of live facial recognition in the country, alerting that already at the time several major police departments were actively exploring real-time face recognition

⁴⁰ A gang is defined in this context as a relatively durable, predominantly street-based group of young people who see themselves (and are seen by others) as a discernible group, and engage in a range of criminal activity and violence.

⁴¹ As explained by the Metropolitan Police's [website](#).

⁴² See notably Recital (51) and Arts. 4(14 and 9(1) of the GDPR; Art. 3(13) and 10 of the LED.

⁴³ Lawsuit settled in January 2020: Singer & Isaac, 2020.

on live surveillance camera video (Garvie, Bedoya & Frankle, 2016).⁴⁴ Documenting a variety of issues related to these developments, including privacy and bias issues but also threats to free speech, the authors emphatically called for **legislation** to regulate law enforcement use of facial recognition searches, and notably for more transparency, under the admonition that *'Face recognition is too powerful to be secret'* (ibid.).

Some US cities such as San Francisco and Boston have banned certain uses of facial recognition in their territory (Vaas, 2020). In Detroit, there is **ongoing litigation** on the matter between the American Civil Liberties Union (ACLU) and the Detroit Police Department, related to the case of a man who was wrongfully arrested on the basis of a misidentification by facial recognition software (Lee, 2020). Still in the US, campaigners have been countering the spread of facial recognition at university campuses, notably on the grounds of what they described as **systemic racial and gender bias** in the deployed systems (Fight for the Future, 2020).

In June 2020, the US Technology Policy Committee (USTPC) of the Association for Computing Machinery (ACM) urged an immediate suspension of private and governmental use of facial recognition technologies *'in all circumstances known or reasonably foreseeable to be prejudicial to established human and legal rights'*, pending the adoption of appropriately comprehensive law and regulation to govern its use, oversee its application, and mitigate potential harm (ACM USTPC, 2020, 1). The call was based on the assessment that the technology too often produces results demonstrating clear **bias based on ethnic, racial, gender, and other human characteristics**, characteristics that the ACM USTPC nevertheless described as being *'recognizable by computer systems'* (idem).

Policy discussions around facial recognition are prominent in the **UK**, where a series of police forces rely on these systems. In London, for instance, the Metropolitan Police uses Live Facial Recognition to *'to help tackle serious violence, gun and knife crime, child sexual exploitation and help protect the vulnerable'*.⁴⁵ The UK is notably funding a research project on fostering unconstrained face biometrics capability, presented as potentially significantly contributing to the UK government's security agenda in the framework of smart cities and national security.⁴⁶ The Metropolitan Police Service is one of the project's collaborators, together with inter alia a Chinese academic institution, and the UK Home Office is one of the project's partners.⁴⁷

In 2019, the High Court of England and Wales declared that the use of automated facial recognition technology to match the faces of members of the public against police watchlists was **lawful**, in a case concerning the programme piloted for already a number of years by **South Wales Police**, called AFR Locate.⁴⁸ The programme is described by South Wales Police as a 'live-time' deployment of automated facial recognition technology which compares live camera feeds of faces against a predetermined **watchlist**, in order to 'locate' persons of interest, generating possible matches that are reviewed by the operator(s).⁴⁹ A 2018 evaluation of the use of facial recognition technology by South Wales Police noted that legal instruments surrounding it *'and the ethical principles they expound pre-date the kinds of*

⁴⁴ The original deployment of a 'smart CCTV system' in public streets in the US has been traced back to 2001; the first uses of these technologies are believed to have been in prison settings (Gates, 2011, 54 and 64).

⁴⁵ See the MET [online information about Live Facial Recognition](#).

⁴⁶ Project [Face Matching for Automatic Identity Retrieval, Recognition, Verification and Management](#), FACER2VM, EP/N007743/1, January 2016- September 2021.

⁴⁷ Idem. See also: Privacy International, 2020d.

⁴⁸ R (on the application of Edward Bridges) v The Chief Constable of South Wales [2019] EWHC 2341.

⁴⁹ See *ad hoc* website by South Wales Police: <http://afr.south-wales.police.uk/>. 'Persons of interests' are defined as people wanted on warrants and suspects for particular crimes, but also potentially missing persons and vulnerable people, and people sought for intelligence purposes.

affordances AFR technologies provide', adding that '[a]s such, new regulatory frameworks will probably be necessary as this technology becomes more widely adopted' (Davies, Innes & Dawson, 2018, 41).

Over the years, a number of calls for a **suspension of the deployment of certain uses** of facial recognition technology have been voiced out internationally by civil society organisations. The Public Voice's 2019 Declaration *A Moratorium on Facial Recognition Technology for Mass Surveillance* not only called for such a suspension, but also urged countries *'to establish the legal rules, technical standards, and ethical guidelines necessary to safeguard fundamental rights and comply with legal obligations before further deployment of this technology occurs'*.⁵⁰

A particularly controversial issue concerns the role of **private companies** in the deployment of these technologies. In February 2020, media reported that according to leaked documents the company Clearview AI had contracts with thousands of law enforcement agencies, companies, and individuals around the world, and also in Europe (Mac, Haskin & McDonald, 2020). Clearview AI services, which it formally targets to *'active law enforcement personnel'*,⁵¹ can be described as consisting in matching persons to online images taken from a variety of platforms (Hill, 2020). The service allows for instance an investigating officer to upload a photo of an individual 'of interest' and search a database, compiled by Clearview AI, of what are presented as 'publicly available images', posted online by somebody (Blatt, 2020) – not necessarily the person at stake. Some of the platforms from which Clearview AI appeared to capture images have formally objected to such scraping of pictures by Clearview AI (Porter 2020).

In March 2020, the Swedish Data Protection Authority (Datainspektionen) disclosed it was investigating the possible use of Clearview AI services by Swedish public authorities (Datainspektionen, 2020). In July 2020, the ICO and the Australian Information Commissioner announced they had opened a joint investigation into Clearview AI, focusing on its use of 'scraped' data and biometrics of individuals (ICO, 2020).

In a letter in response to concerns raised by a number of MEPs regarding **the possible use of Clearview AI products by EU law enforcement authorities**, dating from 10 June 2020, the EDPB noted that such authorities may under certain circumstances use facial recognition technologies to check images against templates *'in databases that are under the control of the official authorities and that have been established under Union or Member State law'*, but that the *'possible use of a service such as offered by Clearview AI by law enforcement authorities would, however, be fundamentally different, in that it would imply, as part of a police or criminal investigation, the sharing of personal data with a private party outside the Union and the biometric matching of such data against the latter's mass and arbitrarily populated database of photographs and facial pictures accessible online'* (Jelinek, 2020). The EDPB added it had *'doubts as to whether any Union or Member State law provides a legal basis for using a service such as offered by Clearview AI'*, noted that *'several EDPB members have already started to further inquire about the use of such facial recognition technologies in their respective jurisdictions'*, and expressed its willingness to contribute to further debate on these matters (idem).

Clearview AI is not the only company about which reports of agreements with law enforcement authorities have been published. Amazon's owned company Ring, a 'smart security' device company mostly known for selling interactive video doorbells, has been reported to have a variety of agreements with law enforcement authorities, including in Europe (Privacy International, 2020b), which may or not relate specifically to facial recognition or other AI-related technologies.

⁵⁰ Full text and list of signatories available [here](#).

⁵¹ See, in this sense: <https://clearviewai.typeform.com/to/SFnULY>.

The deployment of **facial recognition technologies in the public space** has provoked controversy in many countries worldwide. It is the case for instance in Brazil, where the installation in the public metro of a system to deliver ads based on automated ‘detection’ of emotion, gender, and age of passers-by led to litigation by the Brazilian Institute of Consumer Protection (IDEC) (Arroyo & Leufer, 2020).

In the EU, the deployment of facial recognition in public spaces, including in test mode, has generally provoked a variety of reactions. In 2017, the public was said to be divided when a field test of a facial recognition programme was launched at Berlin’s railway stations (Fürstenau, 2017).

There have also been reactions from the judiciary and DPAs. In Belgium, the DPA halted a deployment of facial recognition cameras at Zaventem Airport that had no legal basis (Peeters, 2020). In France, **an administrative court declared invalid** in February 2020 a decision taken by a regional council allowing to test **a surveillance system relying on facial recognition** to monitor access to two high schools.⁵² The judgment criticised *inter alia* the lack of prior impact assessment and the attempt to ground the processing of personal data on the consent of the individual concerned, or their parents, despite the fact that their relation towards the institutions rendered impossible the provision of genuinely free consent. The judgment followed a complaint introduced together with other associations by the NGO La Quadrature du Net, also active in contesting the use of facial recognition to monitor participants to demonstrations (La Quadrature du Net, 2019).

In August 2019, the Swedish DPA fined a municipality 200 000 for using facial recognition technology to monitor the attendance of students in school (EDPB, 2019). Facial recognition had been deployed in the context of a pilot based on the consent of individuals, but the DPA considered that consent was not a valid legal basis in such context given the clear imbalance between the data subject and the controller (*idem*). The same DPA has, however, authorised certain uses of facial recognition by law enforcement authorities (Barik, 2019).

Facial recognition technologies have been until now deployed in public, or semi-public spaces, both by public authorities and private companies. In Spain, for instance, facial recognition systems have been recently deployed in some supermarkets presumably in order to automatically detect individuals who might be subject to a restraining order not allowing them to enter the premises or approach the supermarket’s workers (Rubio, 2020). EU DPAs have sometimes authorised certain uses of live facial recognition by private actors – for example, the Danish DPA authorised its use by a football club at the entrance of its stadium, to identify persons banned from attending football matches (IT-Pol, 2019).

In 2020, some large companies such as IBM announced they would refrain in the future from certain commercial activities related to facial recognition (Peters, 2020).

3.3. AI and criminal justice

AI-related developments, relying on an increasingly extensive use of data, computation and algorithms, have also permeated the administration of justice,⁵³ including criminal justice. The origins of some current developments may be traced back to the emergence of ‘computerised justice’ as a trend aiming at improving the accountability and predictability of judicial decisions, by reducing discretionary powers in the hands of judges regarded by some, in some circumstances, as excessive (Franko Aas, 2005, 65).

⁵² Tribunal administratif de Marseille (9ème chambre), N° 1901249, 27 February 2020.

⁵³ Justice has embraced the digital in a number of ways, such as online dispute resolution, where algorithms play an increasing role (Katsh & Rabinovich-Einy, 2017).

AI is used or being explored in European legal systems for a variety of purposes, such as facilitating access to law (e.g. through chatbots), supporting alternative dispute settlement measures in civil matters, online dispute resolution, or 'judge profiling'.⁵⁴ Some of the ways in which AI-related practices have entered the courts relate to the calculation of the risks of misconduct, for instance for algorithmic probation (Wilson, 2017b, 144). The use of predictive tools by judges in criminal trials in Europe, sometimes labelled 'algorithmic justice' or 'automated justice', has nonetheless at least until recently been described as very rare (Ronsin & Lampos, 2018, 37).

Recidivism algorithms have gathered much attention especially in the **US**. Many of the concerns raised over the use of algorithms for criminal justice purposes refer to indirect racial **bias** in models that predict offending and re-offending, for instance by the use of proxy variables that are not neutral. A ProPublica investigation conducted in 2016 into the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system found that black defendants were almost twice as likely to be deemed at risk of offending than white defendants (Babuta, Oswald & Rinik, 2018, 7, noting however the system's developer contested the analysis; Završnik, 2017b, 145).

Issues around **due process** safeguards have also come to the fore, most famously in *State v. Loomis*,⁵⁵ a judgment by the Wisconsin Supreme Court held that a trial court's use of an algorithmic risk assessment in sentencing did not violate the defendant's due process rights (even though the methodology used to produce the assessment at stake was not disclosed), nevertheless introduced a procedural safeguard to alert judges about the limitations of such assessments. Another interesting US case regarded litigation in Washington, D.C. challenging the use of the Structured Assessment of Violence and Risk in Youth (SAVRY) risk assessment tool in a criminal juvenile sentencing proceeding: the tool had determined that a young person who had been promised to be given probation in exchange of pleading guilty was nevertheless 'high risk', and had to be incarcerated (Richardson, Schultz, Southerland, 2019, 9). The defence lawyers challenged the SAVRY assessment contesting its scientific robustness, and won the argument, convincing the judge to disallow the use of the tool in the specific case at stake (ibid., 10).⁵⁶

The **UK** provides a good example of developments in this area with the algorithmic 'Harm Assessment Risk Tool' (HART) developed in 2017 by Durham Constabulary to assess the risk of individuals reoffending, in order to support custody decision-making (Babuta, Oswald & Rinik, 2018, 6). The tool generates a risk score that becomes one among a number of factors for human officers to take into account when making their overall risk assessment (idem). The tool generated controversy *inter alia* for originally relying on the use of consumer segmentation data from a private company, later put aside (Grace, 2019, 6).

Generally speaking, the use of AI in judicial systems is critically dependent on the **availability of data**, which has led to recognising an open data approach to judicial decisions as a prerequisite for the work of legal tech companies specialising in search engines or trend analysis (so-called 'predictive justice') (Ronsin & Lampos, 2018, 16). The processing of such data raises however a number of issues, some of

⁵⁴ Cf. Appendix II accompanying the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, adopted by the CEPEJ during its 31st Plenary meeting (Strasbourg, 3-4 December 2018), CEPEJ(2018)14.

⁵⁵ 881 N.W.2d 749 (Wis. 2016).

⁵⁶ Other relevant cases include *Louisiana v. Hickerson*, concerning the use by law enforcement authorities in the city of New Orleans of a risk-assessment database called Gotham, created by the company Palantir (Richardson, Schultz, Southerland, 2019, 13); The database had been in principle deployed to identify individuals likely to become a perpetrator or victim of gun violence, to whom awareness messages about their lifestyle were to be sent, but appeared to have played a role in investigations leading to a prior trial against Mr. Hickerson (ibid., 14).

which concern data protection law, but also others which concern the possibility for change in the formation of case-law (*idem*).

Researchers have noted that although this type of tools are currently being used for limited purposes, there is potential for the technology to do much more, and the lack of a regulatory and governance framework for its use is concerning. (Babuta, Oswald & Rinik, 2018, 7).

3.4. AI and borders

There are a variety of instances in which algorithms are already relevant in a border context. Automated decision systems are becoming more and more prevalent in relation to borders globally, and can include systems that **classify** individuals (e. g. as high risk, or high priority); that **generate scores**, probability assessments, and other indicators supporting human decision-making; that **'flag'** cases for review or investigation; that **provide recommendations** on applications, or even that render full decisions (Molnar & Gill, 2018, 3).

3.4.1. General trends

In the EU, border-related largescale information systems incorporating algorithmic processing are typically **connected, through a variety of data flows, to law enforcement authorities** – national authorities, Europol, or often both. Data emanating from law enforcement authorities might be fed into border-related systems, and data from these systems might eventually be supplied or rendered accessible to law enforcement authorities. **Security and border management** have been major drivers for the growth of both centralised and decentralised information systems in the AFSJ for many years already.⁵⁷

The interest among public authorities in **data** to be processed for migration and asylum purposes appears to be increasing in many Member States. Some of them have for instance introduced over the past years legal changes to their national asylum procedures, *inter alia* to make it possible for authorities to seize and analyse data in personal devices (e.g. smartphones) from asylum seekers,⁵⁸ also increasing the applicants' duty to cooperate (EMN, 2019, 6). The data might be analysed to help establishing their identity and travel routes (*ibid.*, 21). Member States are also enhancing border control by investing in technical equipment for border checks (e.g. 'e-gates', biometric ID checks, kiosks, etc.), including facial recognition tools (*ibid.*, 54). Technologies used at the borders are manifold and can include automated surveillance systems with different detection capabilities (e.g., heartbeat detection and thermal cameras) (*idem*).

Globally, there has been a trend over the years to increase data disclosure obligations for individuals crossing borders. In the US, ongoing litigation is challenging an obligation imposed since 2019 on visa applicants to disclose social media accounts (EFF, 2020).

At EU level, a notable development was the issuing in 2019 by the EDPS of a temporary ban on the production of **social media monitoring** reports by the European Asylum Support Office (EASO). EASO was using the such reports to provide management and relevant stakeholders information '*on shifts in asylum and migration routes and smuggling offers, as well as an overview of conversations in the social media community relating to key issues, such as flight, human trafficking and other asylum systems and processes*' (EDPS, 2019, 30). The EDPS decried this was occurring without the necessary legal basis and

⁵⁷ For an overview, see, for instance: COM(2016) 205 final.

⁵⁸ In Germany there is ongoing litigation questioning the legality of the measures, in place since 2017, allowing public authorities to access data stored in the mobile devices of asylum seekers in view of preventing asylum fraud (Reuters, 2020).

appropriate safeguards, and has highlighted that *'(s)ocial media monitoring tools in general raise a number of serious data protection concerns', which in that case 'included a **chilling effect** - the tendency for users to self-censor their online content if they think it might be monitored -, the high risks posed to the fundamental rights of the individuals and groups monitored, **the lack of fairness and transparency** involved in processing this data and the vast number of social media users implicated'*.⁵⁹

In EU's AFSJ, large scale information systems increasingly incorporate **algorithmic decision-making**. This is generally presented under a variety of terms in the relevant legislative documents, different from AI, including for instance references to the *'the data-analytics method'* (COM(2018) 302 final, 9). Regardless of terminology, what is generally at stake is that beyond the detection of 'known' specific suspects, some instruments contain an algorithmic profiling functionality which identifies 'unknown' individuals who may be of interest to law enforcement and border management authorities (EU FRA, 2018a, 114).

In 2016, the European Parliament requested the European Commission to launch a Pilot Project for a fundamental rights review of EU data collection instruments and programmes. The project eventually voiced out **concerns on the impact on fundamental rights** of the **'profiling functionalities'** included in both existing and upcoming EU information systems (FGB, 2019, 4).

An important number of developments in this field converge in **eu-LISA**, the EU Agency ensuring the operational management of the EU large-scale Information Technology (IT) systems, and their respective communication infrastructure, in the AFSJ: Eurodac (the European database of digitalised fingerprints of asylum seekers and irregular migrants), the Schengen Information System (SIS) (facilitating the exchange of information on persons and objects between national police, border control, customs, visa and judicial authorities), and the Visa Information System (VIS), which supports the implementation of the EU's common visa policy by enabling dedicated national authorities to enter and consult data, including biometrics, for short-stay visas to the Schengen Area. Eu-LISA is currently developing other upcoming EU-wide information systems: the **Entry/Exit System (EES)**, the **European Travel Information and Authorization System (ETIAS)** and the European Criminal Records Information System for Third-Country Nationals (ECRIS-TCN), as well as their interoperability components, and plays a crucial role in the technical implementation and development of the interoperability of EU information systems as foreseen by the 2019 **interoperability package**.⁶⁰

A more detailed examination of ETIAS can help understanding some of the fundamental rights challenges related to these developments, which are relatively paradigmatic, and thus useful to situate the EES and an upcoming 'upgrade' of VIS.⁶¹ Before that, issues surrounding the algorithmic processing of data related to travelling shall be considered.

⁵⁹ EDPS, *Annual Report 2019, Luxembourg: Publications Office of the European Union, 2019*, p. 30.

⁶⁰ Cf. Regulation (EU) 2019/817 establishing a framework for interoperability between EU information systems (borders and visa) and Regulation (EU) 2019/818 establishing a framework for interoperability between EU information systems (police and judicial cooperation, asylum and migration), OJ L 135, 22.5.2019, p. 27–84; and Regulation (EU) 2019/818 of the European Parliament and of the Council of 20 May 2019 on establishing a framework for interoperability between EU information systems in the field of police and judicial cooperation, asylum and migration and amending Regulations (EU) 2018/1726, (EU) 2018/1862 and (EU) 2019/816, OJ L 135, 22.5.2019, p. 85–135..

⁶¹ In 2018, the European Commission published a proposal for a Regulation of the European Parliament and of the Council on upgrading the Visa Information System (VIS) (COM(2018) 302 final). The upgrade presented in the proposal relies notably on introducing the use of *'risk indicators'* containing *'data analytics rules, as well as specific values provided by Member States and statistics generated from other relevant border management and security databases'* (ibid., 9).

3.4.2. PNR

Profiling travellers is not a novelty in the EU. Significant legal and policy discussions have taken place in relation to the processing of **Passenger Name Record (PNR) data**, be it the transfer of such data to **third countries**, or in the context of the **EU PNR system**. Originally, what was especially challenging was the question of how to frame, from a EU law perspective, the use for law enforcement purposes of data originally collected by private companies for commercial purposes.⁶² Eventually, questions focused on the substantive requirements in terms of fundamental rights for the processing of data under PNR schemes.

The European Parliament requested to the CJEU an Opinion of an agreement for the transfer of **PNR data to Canada** that had been negotiated with Canadian authorities, in particular regarding the envisaged agreement's compatibility with obligations stemming from the EU Charter. The Opinion, delivered in July 2017, stressed a number of significant failures.⁶³ Noting the envisaged agreement permitted *'the systematic and continuous transfer of PNR data of all air passengers flying between the European Union and Canada'*,⁶⁴ to be used as *'an intelligence tool'*,⁶⁵ the Court observed that the PNR data were *'intended to be analysed systematically before the arrival of the aircraft in Canada by automated means, based on pre-established models and criteria'*, as well as automatically verified by cross-checking with other databases,⁶⁶ and that those analysis could *'give rise to additional checks at borders in respect of air passengers identified as being liable to present a risk to public security and, if appropriate, on the basis of those checks, to the adoption of individual decisions having binding effects on them'*.⁶⁷ It noted that these analyses were to be carried out *'without there being reasons based on individual circumstances that would permit the inference that the persons concerned may present a risk to public security'*.⁶⁸

The CJEU stated that *'the extent of the interference which automated analyses of PNR data entail in respect of the rights enshrined in Articles 7 and 8 of the Charter essentially depends on the pre-established models and criteria and on the databases on which that type of data processing is based'*,⁶⁹ and, as a consequence, *'the pre-established models and criteria should be specific and reliable'*,⁷⁰ making it possible (...) to arrive at results targeting individuals who might be under a **'reasonable suspicion'** of participation in terrorist offences or serious transnational crime and **should be non-discriminatory'**.⁷¹

⁶² Cf. Judgment of the Court (Grand Chamber) of the CJEU of 30 May 2006, *European Parliament v Council of the European Union* (C-317/04) and *Commission of the European Communities* (C-318/04), ECLI:EU:C:2006:346.

⁶³ CJEU, *Opinion 1/15 of 26 July 2017* ('EU Canada PNR'), ECLI:EU:C:2017:592.

⁶⁴ *Ibid.*, para. 127.

⁶⁵ *Ibid.*, para. 131.

⁶⁶ *Idem.*

⁶⁷ *Ibid.*, para. 132.

⁶⁸ *Idem.*

⁶⁹ *Ibid.*, para. 172.

⁷⁰ In this regard, and *'since the automated analyses of PNR data necessarily involve some margin of error'*, *'any positive result obtained following the automated processing of that data must (...) be subject to an individual re-examination by non-automated means before an individual measure adversely affecting the air passengers concerned is adopted'* (para. 173).

⁷¹ *Idem.* In relation to this last point, the CJEU proclaimed that *'in order to ensure that, in practice, the pre-established models and criteria, the use that is made of them and the databases used are not discriminatory and are limited to that which is strictly necessary, the reliability and topicality of those pre-established models and criteria and databases used should, taking account of statistical data and results of international research, be covered by the joint review of the implementation of the envisaged agreement'* (para. 174).

The **EU PNR** system was set up by the Directive (EU) 2016/681 (the 'EU PNR Directive') in 2016.⁷² It applies as a general rule to data of 'extra-EU flights', although Member States can also extend it to 'intra-EU flights'. The system it puts in place is in a way reminiscent of the general approach described above for the prevention of money laundering, to the extent that data processing obligations are imposed on air carriers, which must be in contact with *ad hoc* national **Passenger Information Units (PIUs)**. In this scenario, however, the risk assessment procedures are not to be undertaken by private companies, which are only responsible for the provision of collected data. PIUs are responsible for collecting, storing and **further processing** such PNR data, for transferring them to the competent authorities, and for exchanging them with the PIUs of other Member States, as well as with Europol.

One of the most discussed issues over the implementation of the EU PNR Directive has been the involvement of **Europol** in the development and dissemination of a '*common sets of indicators, targeting rules and risk profiles*' (Presidency of the Council, 6300/19, 2019). In 2019, a majority (88%) of Member States declared that it would be a '*great idea*' if Europol could contribute to the developing targeting rules and indicators, '*due to its central position and its important role in gathering and sharing information on targeting rules and/or risk profiles*', and act as '*depository*' to the common rules and indicators (ibid., 14). Some Member States, however, noted it should be up to each PIU and Member State to decide if they would invite Europol to develop such sets and to what extent these sets would be considered as obligatory. There were also discussions on whether the idea that Europol could offer a central storage of all indicators and targeting rules used nationally, presumably acceptable by some, was actually a good idea, with some Member States regarding that as a potential information security breach that would jeopardise their (national PIUs) effectiveness (ibid. 9).

The CJEU is expected to rule on the **compatibility of the EU PNR Directive and EU fundamental rights**. A pending case (Case C-817/19) concerns indeed a request for a preliminary ruling from the Belgian Constitutional Court, lodged in October 2019, following action by the association *Ligue des droits humains*, on transposition of the EU PNR Directive in Belgium. The questions submitted to the Court refer notably to the systematic prior assessment of the data of all passengers (Belgian Constitutional Court, 2019), that is, the 'advance assessment' made by comparing passenger data against databases and predetermined criteria, which applies '*in a systematic and generalised manner, regardless of whether there is any objective ground for considering that the passengers concerned may present a risk to public security*'.⁷³ In January 2020, the District Court of Cologne also submitted to the CJEU a request for a preliminary ruling, in the context of proceedings initiated by the Gesellschaft Für Freiheitsrechte (GFF) and a number of plaintiffs (Turß, 2020). One of the plaintiffs litigating in Germany is the former Head of the Secretariat of the Committee on Civil Liberties, Justice and Home Affairs in the European Parliament (LIBE Committee) (De Capitani, 2019).

3.4.3. ETIAS

The **European Travel Information and Authorisation System (ETIAS)** concerns travel authorisations for third-country nationals exempt from visa requirements to enter the Schengen area, on the grounds of a '*consideration of whether the presence of those third-country nationals in the territory of the Member*

⁷² Directive (EU) 2016/681 of the European Parliament and of the Council of 27 April 2016 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime, OJ L 119, 4.5.2016, p. 132–149.

⁷³ See [application to the CJEU in Case C-817/19](#).

States would pose a **security, illegal immigration or high epidemic risk**.⁷⁴ Member States' designated authorities and Europol can consult data stored in the ETIAS Central System for the purposes of the prevention, detection and investigation of terrorist offences or of other serious criminal offences.⁷⁵ With ETIAS, the relevant considerations to decide on travel authorisation occur prior to the travellers' arrival at external border crossing points. Its use shall be optional only during the first six months from the date on which it starts operations,⁷⁶ but compulsory afterwards. A travel authorisation, it must be noted, does not confer an automatic right of entry or stay.⁷⁷

In practice, applicants shall submit an application online or via an app,⁷⁸ providing the requested personal data.⁷⁹ The submitted data are then compared to the data present already in the ETIAS Central System, but also to data in SIS, EES, VIS, Eurodac, and Europol and Interpol data,⁸⁰ and to the ETIAS watchlist.⁸¹ Some personal data are also compared to a series of **risk indicators**: that applies notably to date and place of birth, gender, nationality, first name(s) of the parents of the applicant, home address, education and occupation.⁸²

If there is a hit – that is, a coincidence, this information is recorded in the application file.⁸³ If the hit is due to coincidence for data emanating from SIS, for instance because the applicant has been registered as deserving 'discreet checks or specific checks', a reference to this will be added to the application file.⁸⁴ Eventually, in case of hit, it is up to the ETIAS National Unit of the Member State identified as responsible shall to issue or refuse a travel authorisation.⁸⁵ When the hit is due to a coincidence with a risk indicator, *'the ETIAS National Unit of the Member State responsible shall assess the security, illegal immigration or high epidemic risk and decide whether to issue or refuse a travel authorisation'*, and *'(i)n no circumstances may the ETIAS National Unit of the Member State responsible take a decision automatically on the basis of a hit based on specific risk indicators'*, as that risk shall be assessed *'individually'*.⁸⁶

Article 33 of the ETIAS Regulation explains that what the system foresees constitutes a type of **'profiling'** as defined in the GDPR: data provided by applicants shall be compared with *'specific risk*

⁷⁴ See Art. 1(1) of Regulation (EU) 2018/1240 of the European Parliament and of the Council of 12 September 2018 establishing a European Travel Information and Authorisation System (ETIAS) and amending Regulations (EU) No 1077/2011, (EU) No 515/2014, (EU) 2016/399, (EU) 2016/1624 and (EU) 2017/226, 2018 L 236, 19.9.2018, p. 1–71 (hereafter, 'ETIAS Regulation').

⁷⁵ Art. 1(2) of the ETIAS Regulation.

⁷⁶ Art. 83(1) of the ETIAS Regulation.

⁷⁷ Art. 36(6) of the ETIAS Regulation.

⁷⁸ Art. 15(1) of the ETIAS Regulation.

⁷⁹ Art. 17(2) of the ETIAS Regulation.

⁸⁰ Interpol data are data from the Interpol Travel Documents Associated with Notices database (TDAWN) and the Interpol Stolen and Lost Travel Document database (SLTD). In response to a question about concerns regarding the abuse for political purposes of the 'red notices' in those databases, the European Commission clarified in 2019 that the connection of to these databases shall be conditioned to Interpol providing necessary guarantees (Answer of Mr Avramopoulos on behalf of the European Commission, Question reference: E-000204/2019, 5 April 2019).

⁸¹ The ETIAS watchlist *'shall consist of data related to persons who are suspected of having committed or taken part in a terrorist offence or other serious criminal offence or persons regarding whom there are factual indications or reasonable grounds, based on an overall assessment of the person, to believe that they will commit a terrorist offence or other serious criminal offence'* (Art. 34(1) of the ETIAS Regulation). It is up to Europol or the Member State adding data to the ETIAS watchlist to assess *'whether the information is adequate, accurate and important enough to be included'* (Art. 35(1)(a)). They must also *'assess the potential impact of the data on the proportion of applications manually processed'*, an assessment presumably complex, and in any case to be automated, as eu-LISA is *'shall create a specific tool'* for the purpose of such assessment (Art. 35(2)), the technical specifications of which shall be determined by the European Commission by means of implementing acts (Art. 35(7)).

⁸² Art. 20(5) of the ETIAS Regulation.

⁸³ Art. 20(7) of the ETIAS Regulation. In case of a hit, the ETIAS Central System consults the ETIAS Central Unit, which accesses the file and double checks its validity (Art. 22 of the ETIAS Regulation).

⁸⁴ Article 23 of the ETIAS Regulation.

⁸⁵ Article 26(2) of the ETIAS Regulation.

⁸⁶ Article 26(6) of the ETIAS Regulation.

indicators established by the ETIAS Central Unit' and 'pointing to security, illegal immigration or high epidemic risks'. The **algorithm** enabling such profiling is called the '**ETIAS screening rules**'.

It is the European Commission that shall further define, via a delegated act, what must be regarded as '*risks related to security or illegal immigration or a high epidemic risk*'.⁸⁷ Shall be taken into account for these purposes, notably, EEE statistics on abnormal rates of overstaying and refusals of entry '**for a specific group of travellers**', ETIAS statistics on abnormal rates of refusals of travel authorisations due to a security, illegal immigration or high epidemic risk '**associated with a specific group of travellers**'; statistics on correlations between data and overstaying by travellers or refusals of entry; '*information substantiated by factual and evidence-based elements*' provided by Member States on specific security risk indicators or threats, or on abnormal rates of overstaying and refusals of entry for '**a specific group of travellers**', and information concerning specific high epidemic risks provided by Member States, as well as epidemiological surveillance information and risk assessments provided by the European Centre for Disease Prevention and Control (ECDC) and disease outbreaks reported by the WHO.

On the basis of the risks described will be determined '**risks indicators**', which will then allow the ETIAS Central Unit to establish 'specific risk indicators consisting of a combination of data including one or several of' these categories of data: (a) age range, sex, nationality; (b) country and city of residence; (c) level of education (primary, secondary, higher or none); (d) current occupation (job group). The Regulation notes that these '*specific risk indicators shall be targeted and proportionate*'. It also forbids using risk indicators 'based solely on a person's sex or age', which leaves certain ambiguity as to whether would be permissible risk factors based on a combination of gender and age (such as 'women older than...', or 'men younger than...'). Risk indicators '*shall in no circumstances be based on information revealing a person's colour, race, ethnic or social origin, genetic features, language, political or any other opinion, religion or philosophical belief, trade union membership, membership of a national minority, property, birth, disability or sexual orientation*'.⁸⁸ They can nevertheless, as noted, be based for instance on level of education, or job group. The doors leading to possible undue discrimination are thus not fully closed.⁸⁹

The ETIAS Information System, which includes the ETIAS Central System with the ETIAS watchlist, will be managed by **eu-LISA**.⁹⁰ The ETIAS Central Unit shall be in the hands of **Frontex**, the European Border and Coast Guard Agency.⁹¹ It is this Central Unit at Frontex that is entrusted with '*defining, establishing, assessing ex ante, implementing, evaluating ex post, revising and deleting the specific **risk indicators***' used to profile travellers, after consultation of the ETIAS Screening Board.⁹² It is the same Central Unit that will be responsible for 'carrying out regular audits' of the implementation of the risk indicators, '*including by regularly assessing their impact on fundamental rights, in particular with regard to privacy and personal data protection*'.⁹³ The ETIAS Screening Board, which shall have an advisory function, is to work within Frontex and be composed of representatives of the different ETIAS National Units, Frontex itself, and **Europol**.⁹⁴ This advisory board may, if it deems it appropriate, '*consult the ETIAS Fundamental*

⁸⁷ Article 33(2) of the ETIAS Regulation.

⁸⁸ Article 33(5) of the ETIAS Regulation.

⁸⁹ Cf. also, in this sense: FGB, 2019, 33.

⁹⁰ Art. 6 of the ETIAS Regulation.

⁹¹ Art. 7 of the ETIAS Regulation.

⁹² Art. 7(2)(c) of the ETIAS Regulation. In the words of the EU FRA, the ETIAS system will rely on '(a)n algorithm developed by Frontex' (EU FRA, 2018a, 114).

⁹³ Art. 7(2)(e) of the ETIAS Regulation.

⁹⁴ Art. 9(1) of the ETIAS Regulation.

Rights Guidance Board⁹⁵ on specific issues related to fundamental rights, in particular with regard to privacy, personal data protection and non-discrimination.⁹⁶

Regarding the fundamental rights at stake, the ETIAS Regulation gives special prominence to the **right to non-discrimination**. Its Article 14 is titled 'Non-discrimination and fundamental rights' refers nevertheless, in addition to generally banning any personal data processing that might '*result in discrimination against third-country nationals on the grounds of sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation*', the need to fully respect '**human dignity and integrity and fundamental rights, including the right to respect for one's private life and to the protection of personal data**'. It also mentions that '*[p]articular attention shall be paid to children, the elderly and persons with a disability*', and that '*[t]he best interests of the child shall be a primary consideration*'. It must be emphasised that this provision does not prevent, as such, the determination of risks indicators based on the fact that **certain 'groups of travellers' have been associated with certain higher associated risks**, be it security risks, a presumed propensity to 'overstay', or health risks.

The origins of ETIAS and EES can be traced back to the publication of the Smart Borders package by the European Commission in 2013. That package was followed by the launch of a Smart Borders pilot project, which included a testing phase of the proof of concept with operational tests at different border crossing points. The results of the project presented by eu-LISA echoed a small-scale survey conducted by the FRA, which revealed that a majority of third-country-national travellers had expressed concerns with regard to the reliability of the system, notably in relation to possible malfunctions, and the **difficulties related to rectifying any inaccurate data** (eu-LISA, 2015a, 12). The FRA survey results highlighted that the '*most likely implications of incorrect data in the Entry Exit system concern the risks of persons mistakenly flagged as over-stayers and the use that police, immigration or other officials may make of such information*', potentially leading to '*the risk of being apprehended and detained*' (eu-LISA, 2015b, 307). The FRA also noted '*that automated systems could be programmed to identify individuals using sensitive data, such as race, ethnicity or health*' and stated that '*(m) easures to avoid **discriminatory profiling** are, therefore, required*' (idem).

The EDPS has warned that the ETIAS system '*could pose a risk to EU fundamental rights and freedoms*', and that '*applying data protection safeguards effectively under ETIAS is particularly challenging*', as the planned data processing operations proposed involve exchanges of data between controllers who are subject to different EU data protection rules (EDPS, 2019, 22). It has also cautioned that there is a risk that '*individuals will find it **very difficult to exercise their data protection rights effectively***' (idem).

Indeed, in this system the applicable data protection rules are **particularly complex**.⁹⁷ According to the ETIAS Regulation, the '*ETIAS Central Unit shall provide the general public with all relevant information*

⁹⁵ The ETIAS Fundamental Rights Guidance Board, entrusted with both an advisory and 'appraisal' function, is to consist of the Fundamental Rights Officer at Frontex and representatives of the consultative forum on fundamental rights of Frontex, the EDPS, the EDPB, and the EU FRA. The ETIAS Fundamental Rights Guidance Board shall perform regular appraisals and issue recommendations to the ETIAS Screening Board on the impact on fundamental rights of the processing of applications and profiling under ETIAS, '*in particular with regard to privacy, personal data protection and non-discrimination*', have access to the mentioned audits, and publish a yearly report (Art. 9(5) of the ETIAS Regulation).

⁹⁶ Art. 9(5) of the ETIAS Regulation.

⁹⁷ Art. 56 of the ETIAS Regulation. In the context of ETIAS, the EU-DP Regulation shall apply to the processing of personal data by Frontex and eu-LISA; the GDPR shall apply to the processing of personal data by the ETIAS National Units regarding the assessment of applications by border authorities and by immigration authorities, but not when the processing is performed at the ETIAS National Units by competent authorities assessing the applications for the purposes of the prevention, detection or investigation of terrorist offences or other serious criminal offences, in which case the LED shall apply, although when the ETIAS National Unit actually decides on the travel authorisation as such, the GDPR shall always apply. Additionally, the LED shall apply when competent national authorities access the ETIAS Central System for the

in relation to applying for a travel authorisation,⁹⁸ which shall include references to the rights of data subjects under the EU-DP Regulation, the GDPR, the Europol Regulation and the LED.⁹⁹ It is not clear whether or how such general information will effectively make clear to which processing operations apply which rights exactly, and how applicable rights would be exercised.

The ETIAS Regulation states that, when individuals provide data to the system, they shall be informed of the existing procedures applicable to exercising their rights under Articles 13 to 16 of Regulation (EU) 2001/45 (now replaced by the EU-DPR), and Articles 15 to 18 of the GDPR. It is important to note that **this individual information does therefore not encompass any information about the rights to object to profiling (Article 21) and the rights related to automated decision-making (Article 22) of the GDPR.** Are also not mentioned the equivalent rights in the EU-DPR.¹⁰⁰ As a matter of fact, the ETIAS Regulation only explicitly refers about the possible exercise of the rights of access to, of rectification, of completion, of erasure of personal data and of restriction of processing, not to additional rights.¹⁰¹

purposes of the prevention, detection and investigation of terrorist offences or of other serious criminal offences falling under their competence, and the Europol Regulation when it is Europol that accesses the data for equivalent purposes.

⁹⁸ Art. 71 of the ETIAS Regulation.

⁹⁹ Art. 71(q) of the ETIAS Regulation.

¹⁰⁰ The right to object and rights related to automated individual decisions were in Art. 18 and 19 of Regulation (EU) 2001/45.

¹⁰¹ Art. 64 of the ETIAS Regulation.

4. IMPACT ON FUNDAMENTAL RIGHTS

Many efforts have already been devoted by a variety of actors to the identification of the impact on fundamental rights of AI systems broadly speaking. The European Commission's *White Paper on AI* notes that **AI 'can do harm'**, and goes on to note that harm might be immaterial, constituting for instance *'loss of privacy, limitations to the right of freedom of expression, human dignity, discrimination for instance in access to employment'* (COM(2020) 65 final, 10). In an almost equally odd formulation, it also puts forward that the *'main risks related to the use of AI concern the application of rules designed to protect fundamental rights (including personal data and privacy protection and non-discrimination), as well as safety'* (idem, 10). The White Paper eventually concedes more straightforwardly that the *'use of AI can affect the values on which the EU is founded and lead to breaches of fundamental rights'*, be it as a result from flaws in the overall design of AI systems, or from the use of data without correcting possible bias (ibid., 11).

The previous section already provided a series of examples in which developments related to AI systems in the field of law enforcement and justice have been regarded as having an impact on fundamental rights. This section looks closer at some of the issues that are most often raised. It does not attempt to connect specific types of impact with specific developments or areas, as issues – as well as the very rights affected – are generally intertwined: it has been pointed out, for instance, that shortcomings in predictive policing systems can *'create lasting consequences that will permeate throughout the criminal justice system and society more widely'* (Richardson, Schultz & Crawford, 226). It has been stated that lack of rules on automated profiling puts in danger the very notion of **freedom** (Gräf, 2017, 451).

In 2019, the Expert Committee on Human Rights Dimensions of Automated Data Processing and Different forms of AI (MSI-AUT), at the Council of Europe, argued that algorithmic decision-making systems that rely on data-driven profiling techniques may threaten **several human rights** including the rights to a fair trial and to due process, the rights to freedom of expression and information, the right to protection against discrimination in the exercise of rights and freedoms, and the rights to privacy and data protection (MSI-AUT, 2019, 8).

The Council of Europe *Recommendation on the human rights impacts of algorithmic systems* of April 2020 asserts that there are **'significant human rights challenges attached to the increasing reliance on algorithmic systems in everyday life, such as regarding the right to a fair trial; the right to privacy and data protection; the right to freedom of thought, conscience and religion; the right to freedom of expression; the right to freedom of assembly; the right to equal treatment; and economic and social rights'**.¹⁰²

There is a general consensus on the fact that AI systems might be deployed in the field of law enforcement and criminal justice renders the possible impact on fundamental rights **even more significant**.¹⁰³ Finally, it is important to note that the focus on fundamental rights of this study shall not allow ignoring that the use of AI in this field can also raise other types of questions on their impact.¹⁰⁴

¹⁰² Recommendation CM/Rec(2020)1, Appendix, paragraph A.4.

¹⁰³ In this sense, Consideration of the impact of AI when used in criminal proceedings has for instance been described as *'even more important'* because of their direct impact on the individuals' personal freedoms (Ronsin & Lamos, 2018, 39).

¹⁰⁴ For instance, referring to the impact of predictive policing on police accountability: Hardyns & Rummens, 2017, 214.

4.1. Privacy and data protection

Although the rights to respect for private life and to the protection of personal data are recognised separately in the EU Charter of Fundamental Rights (in Article 7 and 8 respectively), under the ECHR data protection safeguards constitute a dimension of Article 8, on the right to respect for private life.

The European Court of Human Rights (ECtHR) case law on Article 8 of the ECHR and surveillance is extensive.¹⁰⁵ The Court has notably highlighted that the possibility of governments acquiring a detailed profile of the most intimate aspects of citizens' lives by **collating different types of data** 'may result in **particularly invasive interferences with private life**', making explicit reference in this context 'to the views expressed by the Court of Justice of the European Union and the European Parliament'.¹⁰⁶ In this sense, it is crucial to keep in mind that AI-related technologies such as facial recognition make it easier to link up personal data across a variety of surveillance systems, and to merge profiles (Datenethikkommission, 2019, 102).

A key issue for the ECtHR is the requirement of **foreseeability** of any surveillance measures. The case in *Liberty*¹⁰⁷ concerned the interception of communications and the subsequent use of 'filtering' techniques, consisting of an **automated sorting system**, operated under human control, aimed at selecting communications on the basis of what was occasionally referred to as 'keyword lists', 'technical databases', or 'The Dictionary'.¹⁰⁸ The Court concluded that the applicable law at the relevant time did not indicate 'with sufficient clarity, so as to provide adequate protection against abuse of power, the scope or manner of exercise of the very wide discretion conferred on the State to intercept and examine external communications'.¹⁰⁹ In particular, the law failed to 'set out in a form accessible to the public any indication of the procedure to be followed for selecting for examination, sharing, storing and destroying intercepted material', and thus the interference with the rights of the applicants under Article 8 of the ECHR could not be regarded as being 'in accordance with the law'.¹¹⁰

Among the necessary safeguards regularly brought to the fore by the ECtHR stands out the **notification** of individuals subject to surveillance. Notification does not need to take place during a surveillance measure, but might take place afterwards, that is, as soon as it would not jeopardise the purpose of the surveillance measure. Such 'subsequent notification of surveillance measures is inextricably linked to the **effectiveness of remedies** before the courts and hence to the existence of effective safeguards against the abuse of monitoring powers', has stated the Court.¹¹¹

The recent *Gaughran* judgment, of 13 February 2020,¹¹² concerned the retention of the DNA profile, fingerprints and photograph of a convicted person.¹¹³ During the proceedings was discussed the

¹⁰⁵ The ECtHR has notably stressed, in relation to the fight against terrorism, that 'it would defy the purpose of government efforts to keep terrorism at bay, thus restoring citizens' trust in their abilities to maintain public security, if the terrorist threat were paradoxically substituted for by a perceived threat of unfettered executive power intruding into citizens' private spheres by virtue of uncontrolled yet far-reaching surveillance techniques and prerogatives' (*Szabó and Vissy v. Hungary*, judgment of 12 January 2016, paragraph 68).

¹⁰⁶ *Szabó and Vissy*, paragraph 70, in relation to the *Digital Rights Ireland* judgment of the CJEU, and the EP Resolution on the US NSA surveillance programme, concerned inter alia with 'the increasingly blurred boundaries between law enforcement and intelligence activities, leading to every citizen being treated as a suspect and being subject to surveillance' as perceived after the 2013 'Snowden revelations' (EP, 2014, 7).

¹⁰⁷ *Liberty and Others v. the United Kingdom*, judgment of 1 July 2008.

¹⁰⁸ *Ibid.*, paragraph 43.

¹⁰⁹ *Ibid.*, paragraph 69.

¹¹⁰ *Idem.*

¹¹¹ *Weber and Saravia v. Germany*, 29 June 2006, paragraph 135.

¹¹² *Gaughran v. the United Kingdom*, Application no. 45245/15, ECLI:CE:ECHR:2020:0213JUD004524515, [2020] ECHR 144.

¹¹³ Concretely, the indiscriminate nature of the powers to retain such data, without reference to the seriousness of the offence, or the need for indefinite retention, and in the absence of any real possibility of review. The ECtHR concluded that

question of whether the database storing the photographs could be accessed for the purposes of **facial recognition**, which apparently was not possible initially (paragraph 13), but was possible eventually (paragraphs 37, 68), as the data could be transferred to another database which did have facial recognition functionalities (paragraph 69). In the judgment, the ECtHR explicitly dismissed the argument submitted by the UK government according to which *'the more data is retained, the more crime is prevented'*, noting that *'accepting such an argument in the context of a scheme of indefinite retention would in practice be tantamount to justifying the storage of information on the whole population and their deceased relatives, which would most definitely be excessive and irrelevant'* (paragraph 89).

Another very interesting judgment was issued earlier this year in the Netherlands, concerning **risk scoring by public authorities**. In February 2020, the Hague District Court delivered indeed a landmark ruling¹¹⁴ about the Systemic Risk Indication (SyRI), a tool that the Dutch government had put in place to combat fraud. The proceedings had been initiated by several civil society groups, including the Dutch Section of the International Commission of Jurists and two private individuals. The Dutch legislator defined SyRI as a technical infrastructure with associated procedures through which data could be linked and analysed in order to generate risk reports. The existence of a risk report would mean that a person was deemed worthy of investigation with regard to possible fraud, perpetration of abuse, or non-compliance with legislation.¹¹⁵

In its judgement, the District Court noted that *'(n)ew technologies – including digital options to link files and analyse data with the help of algorithms – offer (more) possibilities for the government to exchange data among its authorities in the context of their statutory duty to prevent and combat fraud'*, and stated that *'those new technological possibilities to prevent and combat fraud should be used'* (paragraph 6.4). The Court, however, also observed that *'the development of new technologies also means that **the right to the protection of personal data increasingly gains in significance**'* (paragraph 6.5). As these technologies can interfere extensively with the lives of individuals, but it is difficult for individuals themselves to gauge such effect, the legislator bears a *'special responsibility'* when applying instruments such as SyRI (paragraph 6.85).

The Court's analysis led to the conclusion that SyRI legislation did not meet the requirement laid down in Article 8(2) of the ECHR that only permits interferences with the right to respect for private life that may be regarded as necessary in a democratic society, meaning that they should be necessary, proportionate and subsidiary in relation to the intended purpose. In this context, the Court noted that having examined the system in place in light of basic data processing principles established by the EU Charter and the GDPR, concretely the principles of transparency, purpose limitation and data minimisation, the system was *'insufficiently transparent and verifiable'* (paragraph 6.7).

In relation to these gaps, it noted that the legislation in place did not provide clear information on the factual data that could justifiably lead to the conclusion that there was an increased risk (paragraph 6.87), on the functioning of the risk model, on the risk analysis methods, the validation of the risk model

such retention could not be deemed proportionate in the circumstances at stake, which included the lack of any relevant safeguards and the absence of any real review, and thus amounted to a violation of Article 8 of the ECHR.

¹¹⁴ Judgment of The Hague District Court of 5 February 2020, case number C/09/550982, ECLI:NL:RBDHA:2020:865. The judgment was notably celebrated to the extent that it stopped an initiative based on AI to *'detect fraud before it happens'*, and putting under surveillance for such purpose welfare claimants, including people in most vulnerable situations (Privacy International, 2020c).

¹¹⁵ The instrument could be used at the request of certain government bodies or other bodies with a public function, such as municipal authorities, the Netherlands Tax and Customs Administration, the Social Affairs and Employment Inspectorate, or the Immigration and Naturalisation Service. Personal data processed by SyRI included data about work, education, taxes, property, housing, administrative measures and sanctions, civic integration, grounds for exclusion from benefits, pension, debt burden, reintegration, social benefits, and some health care insurance data.

and the verification of the risk indicators (paragraph 6.89). It was **unclear how data subjects 'could be able to defend themselves against the fact that a risk report has been submitted'** about them, and in general know if the data about them were processed on correct grounds, going against the principle that *'a data subject must reasonably be able to track their personal data'* (paragraph 6.90).

In relation to challenges triggered by AI for privacy and data protection law, the European Commission's *White Paper on AI* argues that the GDPR and the e-Privacy Directive *'address these risks'*, but concedes *'there might be a need to examine whether AI systems pose additional risks'* (COM(2020) 65 final, 11).¹¹⁶ It hints to the fact that persons will *'increasingly be subject to actions and decisions taken by or with the assistance of AI systems, which may sometimes be difficult to understand and to effectively challenge where necessary'*, but also that *'AI increases the possibilities to track and analyse the daily habits of people'*, including *'by state authorities or other entities for mass surveillance'* (idem). Moreover, *'AI may also be used to retrace and de-anonymise data about persons, creating new personal data protection risks even in respect to datasets that per se do not include personal data'*, states the EC.

4.2. Freedom of expression and information, and freedom of assembly and association

Some interferences with the rights to privacy and data protection might also simultaneously have a direct negative impact on the rights to freedom of expression and information (Article 11 of the EU Charter), and/or the right to freedom of assembly and association (Article 12 of the EU Charter), which might actually also been affected even when Article 7 or 8 of the EU Charter are not at stake. Notable challenges emerge when AI technologies are deployed to allow for the monitoring of public spaces – for instance, with **live facial recognition**, and the monitoring on communications, for instance in **social media**. Group anonymity can be under special pressure (EU FRA, 2019, 29).

The European Commission's *White Paper on AI* observes that AI-related processing of personal data can trigger new fundamental rights risks that affect rights different than the protection of personal data and privacy, such as the right to freedom of expression, and political freedoms – in particular when AI is used by online intermediaries to prioritise information and for content moderation (COM(2020) 65 final, 11).

This connects to concerns growing over the recent years on the use of automated processes of online content moderation, which can have a serious impact on the right to freedom of expression, but also on the right to non-discrimination.¹¹⁷ These concerns are further exacerbated by a general problem of lack of transparency with regard to communication processes and the enforcement by Internet services of initiatives against disinformation (Van Hoboken et al., 2019, 124). It is important to stress that in relation to online misinformation often coexist calls to support AI tools for monitoring the veracity of information, and calls to reinforce the transparency of algorithms used, notably towards users (High-Level Group on Fake News and Online Disinformation, 2018).

4.3. Non-discrimination

The risks of bias and discrimination triggered by algorithmic decision-making are very frequently highlighted. The EU Fundamental Rights Agency noted that *'[d]irect or indirect discrimination through the use of algorithms using big data is increasingly considered as one of the most pressing challenges of the use of new technologies'* (EU FRA, 2018b, 3). The Council of Europe's *Recommendation on the Human*

¹¹⁶ Specifically, in relation to the GDPR, the EC proclaims that it will *'be monitoring and assessing'* its application *'on a continuous basis'* (COM(2020) 65 final, 11).

¹¹⁷ See, for instance: Committee of Experts on Quality Journalism in the Digital Age (MSI-JOQ), 2019, 15.

Rights Impacts of Algorithmic Systems of April 2020 observes that '(m)any algorithmic systems use optimisation techniques' which rely on prioritising certain values over others, which 'may generate adverse effects, particularly for **minorities and marginalised or disadvantaged groups**'.¹¹⁸

In relation to 'predictive policing', an important issue concerns the very pertinence of the possible inclusion of algorithmic variables such as criminal history and family background that can make the past behaviour of a certain group decide the fate of an individual, who is, nevertheless, 'a unique human being with a specific social background, education, skills, degree of guilt and distinctive motivations for committing a crime' (Ronsin & Lampos, 2018, 39). This relates to the danger of creating 'echo chambers within which **pre-existing prejudice** may be further cemented' (MSI-NET, 2018, 11), and in which individuals might end up being trapped without their knowledge.¹¹⁹

Additionally, researchers have shown how some systems used, notably in the US, are built on data produced during periods of flawed, racially biased, and even unlawful practices and policies ('dirty policing') which have inescapably shaped the very production of data, and raise the risk of creating inaccurate, skewed, or systemically biased data ('**dirty data**') (Richardson, Schultz & Crawford, 2019, 192).¹²⁰ Authors have pointed out that the 'predictive policing' relying on historical crime data risk fuelling a cycle of distorted enforcement: as underlined by criminologists, crime reports and statistics gathered by the police are partly a record of law enforcement's responses to what happens in a community, and not necessarily an accurate record of all the crime that occurs in said community (Robinson & Koepke, 2016). Even 'predictive policing' systems which do not appear to use any personal data can have a negative impact. For instance, 'location-related risk prognoses' can lead to excessive police checks in certain neighbourhoods identified as hotspots and therefore to the ethnic or social profiling of population groups living there (Datenethikkommission, 2019, 215).

The General Findings section of the Stop Secret Surveillance Ordinance banning certain uses of facial recognition in San Francisco notes that '[w]hile surveillance technology may threaten the privacy of all of us, surveillance efforts have **historically been used to intimidate and oppress certain communities and groups more than others**, including those that are defined by a common race, ethnicity, religion, national origin, income level, sexual orientation, or political perspective'.¹²¹

In relation to non-discrimination and equality issues, the Council of Europe Commissioner for Human Rights' has argued that 'Member states should **apply the highest level of scrutiny when using AI**

¹¹⁸ Recommendation CM/Rec(2020)1, Appendix, paragraph A.7. In the already introduced judgement about the fraud detection SyRI system, the District Court of The Hague observed that 'given the large amounts of data that qualify for processing in SyRI, including special personal data, and the circumstance that risk profiles are used', there was 'a risk that SyRI inadvertently' generated connections based on bias, for instance related to immigration background (paragraph 6.93). The Court also noted that the analysis of large data sets may lead to 'undesirable results, including unjustified exclusion or discrimination' (paragraph 6.91), and echoed views supported by the UN Special Rapporteur on extreme poverty and human rights, Philip Alston, according to which SyRI had a discriminatory and stigmatising effect, by spreading negative stereotyping to all occupants of neighbourhoods investigated as problem areas (paragraph 6.92). It was a major concern indeed of Alston that the system actually targeted poor neighbourhoods in the Netherlands, focusing attention on its inhabitants without any concrete suspicion of individual wrongdoing (UN Independent Human Rights Experts Press Service, 2020).

¹¹⁹ See also: US Department of Justice, 2014, 22 ('Likewise, the length of a defendant's prison term should not be adjusted simply because a statistical analysis has suggested that other offenders with similar demographic profiles will likely commit a future crime. Instead, equal justice demands that sentencing determinations be based primarily on the defendant's own conduct and criminal history').

¹²⁰ As a consequence, the systems informed by such data will be problematic unless they offer technical safeguards to adequately mitigate or segregate the 'dirty data', technical safeguards which have not, according to their analysis, be proven to exist (Richardson, Schultz & Crawford, 2019, 226). The researchers emphasised thus the need to develop reliable mechanisms for assessing the harms inherent in the use of historical police data, 'backed by strong public transparency and accountability measures' (idem).

¹²¹ Stop Secret Surveillance Ordinance, Section 1, § c).

systems in the context of law enforcement, especially when engaging in methods such as predictive or preventive policing' (Council of Europe Commissioner for Human Rights, 11). The Commissioner advocates that 'such systems need to be independently audited prior to deployment for any discriminatory effect that could indicate de facto profiling of specific groups', and that '(i)f any such effects are detected, the system cannot be used'.

Bias and discrimination can surface at many levels in AI systems.¹²² Research exposed, for instance, that commercial products classifying individuals into male and female using automated facial analysis algorithms suffered from substantial disparities in accuracy depending on the colour of the person's skin and their actual gender, with darker-skinned females being the most misclassified group (Buolamwini and Gebru, 2018).

The European Commission's *White Paper on AI* mentions **gender-based discrimination** as one of the main potential risks entailed by AI (COM(2020) 65 final, 1), and puts forward as a possible requirement to mitigate this risk the establishment of '*particular obligations to use data sets that are sufficiently representative, especially to ensure that all relevant dimensions of gender, ethnicity and other possible grounds of prohibited discrimination are appropriately reflected in those data sets*' (ibid. 19). Algorithmic discrimination, however, cannot be prevented or mitigated solely by regulating data sets, or by encouraging diversity in AI-related fields,¹²³ as it might also be triggered by the design of the algorithms, or the design of the system implementing it.

In relation to gender-based discrimination,¹²⁴ it is crucial to note that although EU data protection law has special rules to limit automated individual decision-making based on 'sensitive data',¹²⁵ a notion covering a variety of categories of data often used for discriminatory practices, these **data protection rules on 'special categories of data' do not cover data about gender**. Indeed, gender is not a category of data regarded by EU data protection law as deserving special protection as a 'special category of data',¹²⁶ and this despite the fact that there is a broad consensus on the reality of AI-related gender-based discrimination.¹²⁷

¹²² It is generally acknowledged that in many instances '*discrimination-by-computer does not rest on the use of overtly discriminatory criteria*' (Korff, 2012, 22), but rather '*creeps into the algorithms in much more insidious ways*' (ibid., 23).

¹²³ In its 2019 Communication on *Building Trust in Human-Centric Artificial Intelligence*, the European Commission noted that diversity in terms of gender '*should be ensured at every stage of AI development*' (COM(2019) 168 final, 2).

¹²⁴ In relation to gender discrimination, the *White Paper on AI* refers to preparation by the EC Advisory Committee on Equal Opportunities for Women and Men of an Opinion on Artificial Intelligence (ibid., 11), document which has in the meantime already been published (Advisory Committee on Equal Opportunities for Women and Men Opinion on Artificial Intelligence, 2020). This Opinion underscores the importance of the subject but advances as main recommendation to (further) '*reflect and analyse at the EU and Member State level whether the existing legislative and institutional framework is equipped to deal with arising legal questions related to AI and gender equality and non-discrimination or whether there is any need for reform within the general framework for promotion and protection of human rights*' (ibid., 6). In this context, it notably alludes to the work of the High-Level Expert Group on AI, and concretely to its *Policy and Investment Recommendations for Trustworthy AI*, a document primarily associating gender issues to the need to tackle the gender gap in digital skills by improving education (AI HLEG, 2019b, 32-35). For a more detailed discussion on AI challenges and EU law, see, by the author: González Fuster, 2020.

¹²⁵ Article 9(1) of the GDPR lists the special categories of data (often referred to as 'sensitive data', as in Recital (10) of the GDPR): '*Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.*'

¹²⁶ Arguably, it is possible to defend that certain data processing practices which would lead to gender-based discrimination would infringe certain principles of data processing, such as the principle of fairness under the GDPR. Recital (71) of the GDPR proclaims that '*to ensure fair and transparent processing*', the controller should '*secure personal data in a manner that takes account of the potential risks involved for the interests and rights of the data subject and that prevents, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect*'.

¹²⁷ In this sense, for instance: COMEST, 2019, 21.

4.4. Right to an effective remedy and to a fair trial

AI-related technologies can affect effective remedy rights mainly in two ways: first, they can alter the nature of decisions that individuals might have to contest, bringing in novel challenges, second, they can affect the way in which individuals can actually contest against decisions through the judicial system. The former would refer for instance to the possibility of contesting a measure such a law enforcement intervention based on – or significantly affected by – facial recognition (EU FRA, 2019, 31). The latter concerns AI developments in the justice system itself.

The need to maintain AI ‘under human control’ is often evoked in policy discussions and the literature in this field. Making visible this control has implications in terms of bringing clarity to who is responsible for fundamental rights violations. The Council of Europe Commissioner for Human Rights’ has stressed in this sense that *‘states must establish **clear lines of responsibility for human rights violations that may arise at various phases of an AI system lifecycle**’* (Council of Europe Commissioner for Human Rights, 2019, 13). The Commissioner also noted that *‘(a)nyone who claims to be a victim of a human rights violation arising from the development, deployment or use by a public or private entity of an AI system should be provided with an effective remedy before a national authority’*, adding that *‘states should provide access to an effective remedy to those who suspect that they have been subjected to a measure that has been solely or significantly informed by the output of an AI system in a non-transparent manner and without their knowledge’* (idem).

The use of AI systems in the field of law enforcement and criminal justice can generally raise concerns related to the **fair trial standards** of Article 6 of the ECHR, in particular the presumption of innocence, the right to be informed promptly of the cause and nature of an accusation, the right to a fair hearing and the right to defend oneself in person (MSI-NET, 2018, 10).¹²⁸ The possible use of algorithmic systems in the context of justice has been highlighted as a sector in which these systems can **conflict with constitutionally protected rights of overriding importance**, meaning their use might, irrespective of possible protective measures, be permitted only under very restrictive conditions, or prohibited (Datenethikkommission, 2019, 212). In this sense, even legally non-binding proposals for decisions for judgments by algorithmic systems are generally to be regarded as highly problematic from the perspective of the parties concerned (ibid., 213).

4.5. Rights of the child

The EU Charter of Fundamental Rights devotes a special provision to the rights of the child (Article 24), which does not mean that children do not enjoy, additionally, all other fundamental rights enshrined in the EU Charter – such as, for instance, the right to personal data protection. The ECtHR has emphasised that the retention by public authorities of data related to innocent individuals *‘may be especially harmful in the case of minors’*, *‘given their special situation and the importance of their development and integration in society’*.¹²⁹

There are multiple instances in which minors appear as the direct or indirect target of AI-related initiatives in the field of law enforcement.¹³⁰ An especially relevant issue concerns the collection of data

¹²⁸ Connected concerns might arise with respect to Article 5 of the ECHR, protecting against arbitrary deprivation of liberty, and Article 7 (no punishment without law) (MSI-NET, 2018, 10). Stressing that the use of AI in judicial systems can potentially interfere with the right to a fair trial, particularly with the right to a natural judge established by law, the right to an independent and impartial tribunal, and equality of arms in judicial proceedings: Ronsin & Lamos, 2018, 11.

¹²⁹ *Marper v. the United Kingdom*, judgment of 4 December 2008, applications 30562/04 and 30566/04, ECLI:CE:ECHR:2008:1204JUD003056204, paragraph 124.

¹³⁰ *The fact that data about minors might appear in law enforcement databases used for risk assessment has already been pointed out above, in relation to the Gangs Violence Matrix (GVM) case.*

about minors in the relation to **borders**. As regards to the processing of data of third-country nationals, and the automated assessment of risks associated to them, it has been pointed out that vulnerable groups, including refugees, face unique risks in case information about them would reach repressive governments in their countries, which might weaponized the data, putting individuals and their families at grave personal risk (Molnar & Gill, 2018, 43).

Developments such as the increased participation of minors in **demonstrations**, notably in relation to climate change (among other reasons, following the impulse of activist Greta Thunberg), trigger important questions as to how to protect specifically their rights in the face of the monitoring of such demonstrations, but also regarding possible online surveillance of the relevant organisational online activities.

The WEF *Framework for Responsible Limits on Facial Recognition* mentions explicitly children's rights (WEF, 2020). It states that although facial recognition should not exclude anyone and should always be accessible to and usable by all groups of people, *'there may be some instances, such as infants and children, in which an exception to this principle is appropriate and an alternative to facial identification should be offered'* (idem, 9).

According to the EU FRA, one of the opportunities for facial recognition in the field of law enforcement is to use it in order to help finding missing children (EU FRA, 2019, 18). It is very unclear, however, how could facial recognition be used for such purposes without disproportionately interfering with the fundamental rights of all children.

5. POLICY DEVELOPMENTS

This section presents relevant **policy considerations concerning the regulation of AI**, focusing on EU-level discussions but placing them in a broader context, notably by reference to Council of Europe developments. European policy initiatives are numerous, and it is beyond the scope of this study to review them all in detail.¹³¹ European initiatives inscribe themselves in an even richer context of global initiatives and debates around AI,¹³² which can be policy debates, but are also sometimes markedly more technical and academic.¹³³

International initiatives with a strong focus on AI and law enforcement and criminal justice are comparatively less common than others. The Innovation Centre of Interpol and the United Nations Interregional Crime and Justice Research Institute (UNICRI) established a Centre for AI and Robotics, and convened a global meeting on the subject in 2018, leading to the publication of a report in 2019 (Interpol and UNICRI, 2019). The report pointed out that many countries were already then exploring the application of AI and robotics in the context of law enforcement, calling for discussions on their ethical use and social impact (ibid., vi).

Initiatives emanating from the **Council of Europe** are multiple. In 2017, the Parliamentary Assembly of the Council of Europe (PACE) adopted a *Recommendation on Technological Convergence, Artificial Intelligence and Human Rights*, which notably proposed that guidelines be drawn up on ‘a common framework of standards to be complied with **when a court uses artificial intelligence**’ (point 9.2).

In 2018, the European Commission for the Efficiency of Justice (CEPEJ) adopted a **European Ethical Charter on the Use of AI in the Judicial Systems and their Environment** (CEPEJ, 2018), targeting public and private stakeholders responsible for the design and deployment of AI tools and services that involve the processing of judicial decisions and data, as well as public decision-makers in charge of the legislative or regulatory framework, of the development, audit or use of such tools and services. The Charter puts forward five principles: of respect of fundamental rights; of non-discrimination; of quality and security; of transparency, impartiality and fairness; and of ‘under user control’ (ibid., 5).

CEPEJ’s *Ethical Charter* notes that judicial decision processing by AI could in civil, commercial and administrative matters help improving the predictability of the application of the law and consistency of court decisions, but that in criminal matters its ‘*use must be considered **with the greatest reservations in order to prevent discrimination based on sensitive data, in conformity with the guarantees of a fair trial***’ (ibid., 4).¹³⁴

¹³¹ As part of the background research for the Agency’s project on ‘Artificial intelligence (AI), Big Data and Fundamental Rights’, the EU FRA has collected information on AI-related policy initiatives in EU Member States and beyond in the period 2016-2020. The collection (available [here](#)) currently includes over 290 initiatives. A ‘policy initiative’ was defined broadly to include a range of initiatives that could potentially contribute to policy making and standard setting in the area of AI. This could be, amongst others, actual (draft) legislation, soft-law, guidelines and recommendations on the use of AI, or reports that include conclusions with policy relevance. Civil society has also contributed to debates on the EU regulation of AI. Access Now Europe, for instance, stressed in 2019 that the EU must enforce and develop the highest human rights compliance standards for emerging technologies and AI systems that are designed, developed, or deployed in its territory (Access Now Europe, 2019, 5), and put forward a number of recommendations.

¹³² Illustrating global interest in the issue, the Organisation for Economic Co-operation and Development (OECD) adopted in 2019 its [Principles on AI](#). As another example, in the US, in 2016 the Obama White House’s Office of Science and Technology launched a ‘Preparing for the Future of Artificial Intelligence’ initiative on the risks and benefits of AI. More information on the initiative’s [website](#).

¹³³ Coalescing for instance around the keywords Fairness, Accountability and Transparency (FAT).

¹³⁴ In relation to civil, commercial and administrative matters, a study accompanying the Charter highlights among the main guarantees to be reaffirmed the right of access to a court, the adversarial principle, equality of arms, the impartiality and independence of judges, and the right to counsel (Ronsin & Lamos, 2018, 34-35).

The Council of the Europe's Commissioner for Human Rights published in 2019 a Recommendation titled **Unboxing Artificial Intelligence: 10 steps to Protect Human Rights** which describes a number of steps for national authorities to maximise the potential of AI while preventing or mitigating the negative impact on people's lives and rights (Commissioner for Human Rights of the Council of Europe, 2019). It focuses on 10 key areas of action: human rights impact assessment, public consultations, human rights standards in the private sector, information and transparency, independent oversight, non-discrimination and equality, data protection and privacy, freedom of expression, freedom of assembly and association, and the right to work; access to remedies; and the promotion of AI literacy. The latter is explicitly connected to the need to **promote knowledge and understanding of AI 'in government institutions, independent oversight bodies, national human rights structures, the judiciary and law enforcement'**, as well as the general public.¹³⁵

In April 2020, the Council of Europe's Committee of Ministers adopted a *Recommendation on the human rights impacts of algorithmic systems*,¹³⁶ including **Guidelines on addressing the human rights impacts of algorithmic systems**. The guidelines are designed to advise States, and public and private sector actors, in all their actions regarding the design, development and ongoing deployment of algorithmic systems, which, the document comments, are often '*neither clearly public nor clearly private*'.¹³⁷ To ensure that the human rights and fundamental freedoms of all individuals are effectively protected throughout the technological evolution, States '*shall refrain from violating human rights through the use of algorithmic systems, and shall **develop legislative and regulatory frameworks** that foster an environment where all actors respect and promote human rights and seek to prevent possible infringements*'.¹³⁸

These 2020 *Guidelines* remark that algorithmic systems are sometimes '*employed **for predictive purposes** in the context of policing and border control, for the purposes of combatting money laundering and fraud*'.¹³⁹ They note that certain applications of algorithmic systems '*can prompt a particular, higher risk to human rights*', for instance if used by States for their public service or public policy delivery. A similarly heightened risk, the Guidelines explain, '*ensues as a result of use in the context of decision-making processes, by either public authorities or private parties, in situations that carry particular weight or legal consequence*', and thus the use of algorithmic systems in the judicial field for the purpose of legal analysis, prediction or individual risk assessment must be introduced **with great care** and in conformity with the guarantees of a fair trial in Article 6 of the ECHR.¹⁴⁰ The term 'high risk' is introduced in this context to refer to '*the use of algorithmic systems in processes or decisions that can produce serious consequences for individuals or in situations where the lack of alternatives prompts a **particularly high***

¹³⁵ The Recommendation *Unboxing Artificial Intelligence* also contains a checklist to help implement the measures recommended in each key area. The 'do's' include, in relation to access to remedies, the need to assess whether laws 'provide an effective remedy to those claiming to be a victim of a human rights violation arising from the development, deployment or use of an AI system', the need to 'ensure that liability regimes clearly establish who is legally responsible for the whole spectrum of human rights violations that may occur at each phase of an AI system's lifecycle', but also to 'ensure that the judiciary and other relevant national authorities do not place inappropriate weight on the assumed/perceived accuracy or objectivity of an AI system, and that they provide an equality of arms between the victim and the defendant in cases challenging human rights violations caused by AI systems' (Commissioner for Human Rights of the Council of Europe, 2019, 23).

¹³⁶ [Recommendation CM/Rec\(2020\)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems](#), adopted by the Committee of Ministers on 8 April 2020 at the 1373rd meeting of the Ministers' Deputies.

¹³⁷ Appendix, paragraph A.12.

¹³⁸ Appendix, paragraph A.1. Concretely, as a first step, States should '*review their legislative frameworks and policies as well as their own practices with respect to the procurement, design, development and ongoing deployment of algorithmic systems to ensure that they are in line*' with the presented Guidelines (Recommendation, paragraph 1).

¹³⁹ Appendix, paragraph A.8.

¹⁴⁰ Ibid., paragraph A.11.

probability of infringement of human rights, including by introducing or amplifying distributive injustice.¹⁴¹

Still in the context of the Council of Europe, work on a Report on *Justice by Algorithm: The role of Artificial Intelligence in policing and criminal justice systems* was launched following a motion emanating from the Committee on Legal Affairs and Human Rights in 2018 of the Parliamentary Assembly (PACE).¹⁴² The **Council of Europe's Ad hoc Committee on AI (CAHAI)** was set up in order to examine the advisability, feasibility and potential elements of a legal framework for the development, design and application of AI. The EU contributes to the CAHAI in an observer capacity.

5.1. Introduction

Recent years have been particularly rich in EU-level policy developments explicitly connected to AI. Especially since 2018, EU institutions have been delimiting the contours of what increasingly took the shape of an **AI agenda**.¹⁴³ This sub-section considers in more detail input to such debate on AI from the European Commission and the European Parliament,¹⁴⁴ and grants special attention to the role of ethics and the funding of AI in this context.

A number of other EU-level developments, not formally connected to AI policy discussions, might be regarded as having a potentially significant incidence on the impact AI on EU fundamental rights in the field of law enforcement and criminal justice. Generally speaking, it must be noted that over the recent years the **AFSJ** has witnessed **a major transformation in the ways it enables data processing**. This shift has been described as a movement from a 'silo-based approach', whereby different information systems operate as separate entities, and where a strict interpretation of the principle of purpose limitation is crucial, towards a new approach where **interoperability** is the keyword. Typically, data originally processed for non-law enforcement purposes might be made available for law enforcement processing. The Europol Regulation, the interoperability package, and the ETIAS Regulation have been singled out as examples of this trend (Coudert, 2017).

Regarding **facial recognition**, a particularly relevant development is the interest in facial recognition in the context of **Prüm Decision**,¹⁴⁵ based on the principle of cross-border searches. Coinciding with Prüm's 10th anniversary, the Council encouraged Member States experts 'to evaluate the Prüm workflow for further developments with a view to possible new biometric technologies, e.g. face recognition systems' (General Secretariat of the Council, 2018). Some partially publicly available documents¹⁴⁶ echo the establishment of a Focus Group on Face Recognition connected to the Council's Working Party on

¹⁴¹ Idem. The Guidelines submit that 'States should take a precautionary approach and require the refusal of certain systems when their deployment leads to high risks of irreversible damage or when, due to their opacity, human control and oversight become impractical' (Ibid., paragraph A.15).

¹⁴² Motion for a Recommendation *Justice by algorithm – The role of Artificial Intelligence in policing and criminal justice systems*, Doc. 14628, 26 September 2018.

¹⁴³ In a way that was not dissimilar to the progressive emergence, bridging Digital Single Market and Security policies, of an EU cybersecurity agenda and policy.

¹⁴⁴ Other EU institutions are also playing an important role in the area. In February 2019, the Council adopted conclusions on AI calling for 'stronger development, deployment and uptake of Artificial Intelligence applications in all economic sectors, with the aim of making Europe a global leader in Artificial Intelligence', while stressing 'the need for establishing appropriate cyber security requirements for AI and for ensuring accountability and the protection of fundamental rights' (Council of the EU, 6177/19, 7 and 8).

¹⁴⁵ Council Decision 2008/615/JHA of 23 June 2008 on the stepping up of cross-border cooperation, particularly in combating terrorism and cross-border crime, OJ L 210, 6.8.2008, p. 1–11. Prüm is one of the EU AFSJ data processing initiatives that has since the start being openly about data processing for law enforcement purposes, based on the development of decentralised tools for information exchange.

¹⁴⁶ Available public information about these developments is limited, and some available insights are tributary to leaks (cf., e.g., Campbell and Jones, 2020).

Information Exchange and Data Protection (DAPIX) (Presidency of the Council, 2019). Other documents evoke the consideration by said Focus Group of the preliminary results of a feasibility study on amending the current Prüm legislation, a study carried out by a consultancy firm on behalf of the European Commission which presumably presented as a proposal the possibility to create a central database for facial recognition in the EU – a proposal nonetheless rejected by the experts of the Focus Group, notably on the basis of its lack of operational benefits (Austrian Delegation to the Council, 2019, 9).

Some EU-funded research efforts are already moving in this direction.¹⁴⁷ The *Towards the European Level Exchange of Facial Images* (TELEFI) Project,¹⁴⁸ according to its website, is undertaking a study on how facial recognition is currently being used for the investigation of crime across EU Member States, giving particular consideration *'to the potential for implementing the exchange of facial images within the Prüm framework'*. The project focuses on the use of facial recognition in criminal investigations, but is carrying out research on a variety of databases in order to *'provide a future opportunity to consider whether these additional databases could be legally cross-used for criminal investigations/proceedings and transnational data exchange'*.¹⁴⁹

5.2. European Commission

AI-related developments have until now primarily inscribed themselves within **Digital Single Market** discussions.¹⁵⁰ In its 2018 Work Programme, the European Commission announced under the heading *'A connected Digital Single Market'* its commitment to *'look to make the most of artificial intelligence that will increasingly play a role in our economies and societies'* (COM(2017) 650 final, 5).¹⁵¹ In the 2020 Work Programme it was under the heading **'A Europe fit for the Digital Age'** that the Commission announced it was going to put forward a White Paper on Artificial Intelligence *'to support its development and uptake and ensure full respect of European values and fundamental rights'*, in view of *'(m)aking the most of artificial intelligence'* while establishing *'an ecosystem of trust to ensure it develops within clearly defined ethical boundaries'* (COM(2020) 37 final, 4). The establishment of *'A connected Digital Single Market'* was one of the priority policy areas of Jean-Claude Juncker (Juncker, 2014, 6); the Digital Single Market has been defined as *'one in which the free movement of goods, persons, services and capital is ensured'* (COM(2015) 192 final, 2).¹⁵²

¹⁴⁷ As noted for instance in: Monroy, 2020.

¹⁴⁸ Funding via the European Union Internal Security Fund-Police (Grant Agreement: 821284), duration from January 2019 to September 2020; more information on the [TELEFI website](#).

¹⁴⁹ See *'About TELEFI Project'*.

¹⁵⁰ The European Commission's involvement in supporting AI and AI policy encompasses many initiatives. For instance, it hosts and facilitates the European AI Alliance, where representatives of businesses, consumer organisations, trade unions, and other representatives of civil society bodies, are supposed to interact with the experts of the (AI HLEG) in a forum-style setting, and to feed into the stakeholders dialogue. Additionally, a pilot project was commissioned by the European Commission on algorithmic awareness building, following calls by the European Parliament, and was active in 2018 (see the website: <https://algoaware.ipweb.eu/>). In 2018, the European Commission launched [AI Watch](#), a knowledge service to monitor the development, uptake and impact of AI for Europe. AI Watch is supported by the Joint Research Centre (JRC) of the European Commission in collaboration with the Directorate-General for Communications Networks, Content and Technology (DG CONNECT). AI Watch has announced it monitors also the usage and impact of AI technologies used in public services and in public organizations.

¹⁵¹ The 2019 Work Programme noted, under the same title, that the European Commission would continue to *'work on the emerging challenge of Artificial Intelligence by enabling coordinated action across the European Union'* (COM(2018) 800 final, 3).

¹⁵² It expanded on the fact that already in 1995 the EU legislator had established the principle of the free movement of personal data inside the EU (Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ L 281, 23.11.1995, p. 31–50), notably to, in 2018, establish the free movement of non-personal data.

These constitute only a few of the highlights in a rather long history of support by the European Commission of **data processing as key enabler of European economy's growth**, the origins of which can be traced back to the 1970s.¹⁵³ The year 2018 was particularly prolific in relation to AI policy. The European Commission published a Communication embracing the notion of 'human-centric AI' and a three-pronged approach to boost the EU's technological and industrial capacity and AI uptake across the economy; prepare for socio-economic changes, and ensure an appropriate ethical and legal framework (COM(2018) 237 final). The Commission also developed together with Member States a coordinated plan on AI to create synergies, pool data, and increase joint investments (COM(2018) 795 final).

Even if previously the term AI as such had not been particularly prominent, there is a clear continuity with prior discussion around 'big data'. In its 2014 Communication ***Towards a Thriving Data-driven Economy***, the European Commission described '**big data**' as referring '*to large amounts of different types of data produced with high velocity from a high number of various types of sources*', and observed that handling such data required '*new tools and methods, such as powerful processors, software and algorithms*' (COM(2014) 442 final, 4), noting also the advent of '*a new industrial revolution driven by digital data, computation and automation*' (ibid., 2).

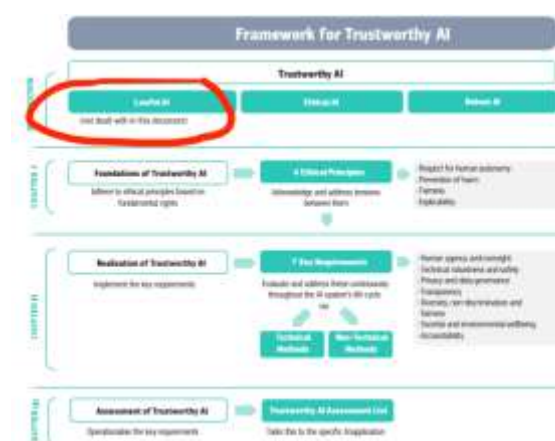
In the 2019 Communication ***Building Trust in Human-Centric Artificial Intelligence***, the 'trust' mentioned in the title was overtly connected by the European Commission primarily to testing '*the practical implementation of ethical guidance for AI development and use*' (COM(2019) 168 final, 1). AI applications, stated the EC, '*should not only be consistent with the law, but also adhere to **ethical principles***', and respect citizens' fundamental rights. The exact relations between these ideas were not clearly defined, but the inferred point of action was that '*there is a need for **ethics guidelines** that build on the existing regulatory framework*', which is why the EC had set up a High-Level Expert Group tasked with drafting such guidelines (COM(2019) 168 final, 2).

The **High-Level Expert Group on Artificial Intelligence** (AI HLEG) had delivered in 2018 the first draft of its *Ethics Guidelines on Artificial Intelligence*, and published a revised document in 2019 following an open consultation which generated extensive feedback (AI HLEG, 2019a). The AI HLEG *Ethics Guidelines on AI* list requirements of 'trustworthy' AI systems, requirements which were put through a piloting process expected to conclude with the publication of a further revised document in 2020.

The role granted to law and fundamental rights in the AI HLEG's *Ethics Guidelines* is relatively elusive, and potentially misleading. The AI HLEG indeed declares that '**trustworthy AI**' should be **lawful, ethical, robust**, but notes that its *Ethics Guidelines* do '*not explicitly deal*' with the lawfulness component of trustworthy AI, opting instead to offer guidance on the ethical and robustness components (AI HLEG, 2019a, 2). A footnote further clarifies that the Guidelines' statements are '*not meant to provide legal advice or to offer guidance on compliance with applicable laws*', while however, at the same time, acknowledging that '*many of these statements are to some extent already reflected in existing laws*' (idem). Later in the document, another footnote insists on the fact that it '*does not provide any advice on ensuring legal compliance (lawful AI)*' (ibid., 3). The Figure below illustrates how the AI HLEG openly affirms to put aside considerations of lawfulness in relation to AI.

¹⁵³ Already then, as the European Commission delineated the contours of its seminal Community data policy, it observed however that the 'data banks' under development could hold police data and that, therefore, rules on access to such information were 'vital' (SEC(73) 4300 final, 13).

Figure 2: AI HLEG Framework for Trustworthy AI



Source: Figure from: AI HLEG, 2019a, 8. Circle in red by the author of this study.

There is no real solid substantive justification in the AI HLEG Guidelines themselves as to why no consideration would be given by this group to the requirements of 'lawful AI'. Somehow puzzlingly, the same Expert Group insists on the fact that **'trustworthy AI'** would be crucial to reap the benefits of AI *'in a way that is aligned with our foundational values of respect for human rights, democracy and the rule of law'* (AI HLEG, 2019a, 4). More confusingly, the same document *Ethics Guidelines* eventually announces it shall *'briefly touch upon'* the requirements of 'lawful AI', only to note that that it proceeds on the assumption that all relevant legal rights and obligations, which are not detailed, *'remain mandatory and must be duly observed'* and that the full realisation of ethical and robust AI may (or may not) *'go beyond existing legal obligations'* (ibid., 6), whichever they might be.

The most ambivalent choice of the *Ethics Guidelines*, nevertheless, is probably the identification of what are called the 'foundations of trustworthy AI' with a 'fundamental-rights based approach' (AI HLEG, 2019a, 7). Such as a proposal is accompanied by a footnote arguing that fundamental rights *'underpin the legally enforceable rights guaranteed by the EU Treaties and the EU Charter'*,¹⁵⁴ and are as such an element of 'lawful AI' to the extent that they are legally binding. However, the footnote maintains that fundamental rights can *'also be understood as reflecting special moral entitlements of all individuals arising by virtue of their humanity, regardless of their legally binding status'*, and in that sense they also form part of 'ethical AI' (idem). As an inescapable logical consequence, it must be understood that **whenever the AI HLEG alludes to fundamental rights, it might or might not be referring to rights that are legally binding**. To the extent that it openly asserts to focus on 'ethical AI' as opposed to 'lawful AI', as a matter of fact most often fundamental rights as discussed by the AI HLEG might not necessarily be related to legally binding rights.¹⁵⁵ This is, from a legal perspective, peculiarly **troublesome**.

This takes place, as a matter of fact, amidst a constant vagueness about the distinction between 'ethical and robust AI', on the one hand, and 'lawful AI', on the other. In this sense, the seven key requirements finally put forward by the AI HLEG, supposedly unconcerned with 'lawful AI', are not deprived of references to legal instruments such as the GDPR (AI HLEG, 2019a, 17). Relegating elemental considerations around the basics of regulation of AI to multiple footnotes, and sustaining an almost constant blurring between what is lawful and what is ethical, the AI HLEG *Ethics Guidelines*

¹⁵⁴ The exact ambitions of this sentence, according to which fundamental rights underpin the fundamental rights established by the EU Charter of Fundamental Rights, remains unclear.

¹⁵⁵ As confirmed notably in AI HLEG, 2019a, 10.

unfortunately did not, in any case, help establishing an intelligible policy debate about the regulation of AI in the EU and its impact on (legally binding) fundamental rights.

The European Commission officially welcomed the work of the AI HLEG in its 2019 Communication *Building Trust in Human-Centric Artificial Intelligence*, in which however it omitted the fact that the AI HLEG *Ethics Guidelines* **had not discussed applicable law as such**. On the contrary, according to the EC the Guidelines had identified key requirements of 'trustworthy AI' based on its three components: **law**, ethical principles, and robustness (COM(2019) 168 final, 3). Furthermore, the EC even celebrated that many already existing EU law provisions '*of course already reflect one or several of these key requirements, for example safety, personal data protection, privacy*' (idem). This only added an extra layer of unclarity to whether and how the discussed ethical guidelines might or not be something different from legal rules. Further entertaining the ambiguity, the EC stated that '*(e)thical AI is a win-win proposition*' visibly linking it to '*(g)uaranteeing the respect for fundamental values and rights*' (COM(2019) 168 final, 8).

In 2019, in *A Union that strives for more: My agenda for Europe*, the by then candidate to the Presidency of the European Commission announced that under in her '*first 100 days in office*' as President she would '*put forward legislation for a coordinated European approach on the human and ethical implications of Artificial Intelligence*' (Von der Leyen, 2019, 13). This was a key element of the top-level ambition of moving towards a '*A Europe fit for the digital age*', in the sense of a Europe '*grasping the opportunities from the digital age within safe and ethical boundaries*' (idem). The plan of putting forward legislation in 100 days eventually evolved into putting forward a White Paper. In the weeks preceding its publication, there were rumours that the document might have considered as a possibility working towards a general ban of facial recognition in public areas, an idea that was eventually not featured as such in the published document (Yun Chee, 2020).

The European Commission published its **White Paper on Artificial Intelligence** on 19 February 2020 (COM(2020) 65 final). The White Paper proclaims that '*(g)iven the major impact that AI can have on our society and the need to build trust, it is vital that European AI is grounded in our values and fundamental rights such as human dignity and privacy protection*' (ibid., 2). The future regulatory framework for AI in Europe, states the document, shall create an '**ecosystem of trust**' which constitutes '**a policy objective in itself**' (ibid., 3), which might, nevertheless, also contribute to other objectives such as creating '*a frictionless internal market for the further development and uptake of AI as well as strengthening Europe's industrial basis in AI*' (ibid. 10).

The announced regulatory framework shall notably ensure socially '*optimal outcomes and compliance with EU legislation, principles and values*', an objective '*particularly relevant in areas where citizens' rights may be most directly affected, for example in **the case of AI applications for law enforcement and the judiciary***' (ibid., 10). It shall be built on an examination of '*whether current legislation is able to address the risks of AI and can be effectively enforced, whether adaptations of the legislation are needed, or whether new legislation is needed*' (idem). One of the ways in which AI can bring benefits to Europe is, according to the White Paper, '*by **equipping law enforcement authorities with appropriate tools** to ensure the security of citizens*' (COM(2020) 65 final, 2).

The White Paper on AI was presented together with two Communications, one on *Shaping Europe's Digital Future* in general, and another specifically about the *European Data Strategy*.¹⁵⁶ The Communication **Shaping Europe's digital future** connected the advent of a 'true digital

¹⁵⁶ The White Paper was also accompanied by a Report of the European Commission on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics (COM(2020) 64 final).

transformation' to the need to make sure that digital applications and products are secure, an aim to be pursued *inter alia* by '**ensuring that law enforcement and judicial authorities can work effectively by developing new tools to use against cybercriminals**' (COM(2020) 67 final, 5).

The Communication on a **European Data Strategy** (COM(2020) 66 final) further develops policies around data that have been over the years connected to notions such as a 'common European data space', the 'data strategy', the 'data economy', originally '*embedded in the Digital Single Market strategy*' and later '*applied to AI*' (Penner, 2019). It addresses restrictions on the free flow of data, including legal barriers on the location of data for storage and/or processing purposes, and a series of issues relating to data such as ownership, access, reuse, portability and liability. It emphasises the value of the availability of data (COM(2020) 66, 6). Noting that data might be used '*for the public good*', it states that data '*created by society*' can, '*where necessary and proportionate, to ensure more efficient fight against crime*' (idem). As part of such Strategy, the European Commission intends to fund the establishment of EU-wide common, interoperable data spaces in strategic sectors (ibid., 16), such as '**common European data spaces for public administration**' aimed *inter alia* at **addressing law enforcement needs** and at enabling '*innovative 'gov tech', 'reg tech' and 'legal tech' applications supporting practitioners as well as other services of public interest*' (ibid., 22).

According to the European Commission's June 2020 Communication on the two years of application of the GDPR, the *White Paper on AI* '**opened up a public debate on the specific circumstances, if any, which might justify the use of artificial intelligence for remote biometric identification purposes (such as facial recognition) in public places, and on common safeguards**' (COM(2020) 264 final, 10).

Reactions to the White Paper had been varied. Some, for instance, questioned whether the confirmed commitment towards European values is compatible with keeping pace with international competitors such as the US and China (Scott, 2020). The **Secretariat of the Council of Europe** submitted responses to the European Commission's consultation on the White Paper, in which was highlighted that the 'framework of trust' to be created between all stakeholders should have as **a prerequisite** the respect for regulatory frameworks and fundamental rights, and the meaningful implementation of the relevant existing regulation (Secretariat of the Council of Europe, 2020, 5). The document comes back to the issue of fundamental rights to state that AI-related research efforts should '*focus on fundamental rights so as to promote the development of legislation, ethical standards and guidelines on AI compliant with such rights*', and promote digital humanities (ibid., 7).

The **EDPS** formally reacted to the *White Paper on AI* on 29 June 2020 (EDPS, Opinion 4/2020). In its Opinion, the EDPS notably stressed that any future regulatory framework for AI shall apply '*to both to EU Member States and to EU institutions, offices, bodies and agencies*' (ibid. 2). Regarding the substance, the EDPS called in particular for an approach to determining the level of risk in the use of AI applications that should be both '*more robust and more nuanced*' (ibid., 11), in the sense that it should be layered instead of an 'all or nothing' approach (ibid., 13).

The EDPS Opinion also supports the idea of '**a moratorium on the deployment, in the EU, of automated recognition in public spaces of human features, not only of faces but also of gait, fingerprints, DNA, voice, keystrokes and other biometric or behavioural signals**' (idem). The moratorium shall coincide with work towards '*appropriate safeguards, including a comprehensive legal framework in place to guarantee the proportionality of the respective technologies and systems for the specific use case*' (idem). In the EDPS' view, some AI applications, such as for instance live facial recognition, '*interfere with fundamental rights and freedoms to such an extent that they may call into question the essence of these rights and freedoms*' (ibid., 7), implying they might not be permissible at all. The EPDS centres part of its analysis around **Remote Biometric Identification (RBI)**, which might rely or not on AI, and includes live facial

recognition in public places but is not limited to it. These systems, the EDPS observes, *'are easily hidden, frictionless, often are presented as mere "experiment" but could easily be turned into ubiquitous and pervasive surveillance complex'* (ibid., 20).¹⁵⁷

As regards to **predictive policing**, the EDPS states that it *'may have negative effects like over-policing on collectives, as well as on individuals'*, and calls in this context to devising *'inclusive governance models which would empower organizations representing civil society (e.g. NGOs and other non-profit associations) so they also can help assessing the impact of AI applications on specific collectives and the society in general'* (ibid., 9). In relation to **non-discrimination**, the EDPS stresses that AI might have specific risks for *'vulnerable groups'*, and puts forward that in the absence of a formally adopted legal definition of vulnerable groups, a *'context-specific, pragmatic approach'* shall be endorsed, to encompass *'children, elderly, and persons with disabilities, ethnic minorities or historically marginalised groups, women, LGBTQIA+ communities, workers and others at risk of exclusion'* (ibid., 21). Finally, in its reaction to the *White Paper* the EDPS also observed that some AI-related measures put forward in the context of the fight against the pandemic represent in themselves, or enable, a kind of *'function creep'* which can have *'a chilling effect in democratic societies'* (ibid., 20).

Regarding EU data protection law, the European Commission's Communication marking the two years of application of the GDPR echoed the ongoing bilateral dialogues between the EC and Member States to ensure the appropriate implementation on the GDPR (COM(2020) 264 final, 14). The Communication notably acknowledged that private companies active in the European market might be called to **share data for law enforcement purposes** *'on the basis of a legitimate request'*, and that such requests might lead to facing conflicts of law or generating **tensions with EU fundamental rights**.¹⁵⁸

In April 2020, the European Commission made public the establishment of an Expert Group on Artificial Intelligence in the domain of Home Affairs to assist European Commission's DG for Migration and Home Affairs in the preparation of legislative proposals or **policy initiatives concerning AI in the domain of Home Affairs**; to establish cooperation and coordination between the Commission and Member States or stakeholders on questions relating to the implementation of EU legislation, programmes and policies in the field of AI in the domain of Home Affairs; and to bring about an exchange of experience and good practice in the field of AI in the domain of Home Affairs. It will be composed of representatives of Member States' and Schengen Associated Countries' authorities.¹⁵⁹

¹⁵⁷ The significance of the challenges connected to RBI do not exclude, the EDPS nuanced, that similar technologies not aimed at identifying individuals can also raise serious concerns: that would be the case, for instance, of real time so-called *'detection'* of emotions (EDPS, Opinion 4/2020, 20).

¹⁵⁸ The European Commission announced in this context that *'[t]o improve such transfers, the Commission is committed to develop appropriate legal frameworks with its international partners to avoid conflicts of law and support effective forms of cooperation, notably by providing for the necessary data protection safeguards, and thereby contribute to a more effective fight against crime'* (COM(2020) 264 final, 13). This notably translates into a commitment *'to assess how cooperation between private operators and law enforcement authorities could be facilitated, including by negotiating bilateral and multilateral frameworks for data transfers in the context of access by foreign criminal law enforcement authorities to electronic evidence'* (ibid., 18).

¹⁵⁹ Commission Expert Group on Artificial Intelligence in the domain of Home Affairs (ref. E03727 at the Register of Commission's Expert Groups and Other Similar Entities).

5.3. European Parliament

The European Parliament has been considering AI in multiple policy contexts – for instance, it adopted in February 2020 a *Resolution on Automated decision-making processes: Ensuring consumer protection, and free movement of goods and services* (P9_TA(2020)0032).¹⁶⁰

In 2017 the European Parliament adopted a **Resolution on fundamental rights implications of big data** (on privacy, data protection, non-discrimination, security and law-enforcement), which called on the European Commission, the EDPB and other DPAs 'to issue guidelines, recommendations and best practices in order to further **specify the criteria and conditions for decisions based on profiling and the use of big data for law enforcement purposes**' (EP, P8_TA(2017)0076, paragraph 25). The Resolution emphasised that 'the trust of citizens in digital services can be seriously undermined by government mass surveillance activities and the unwarranted accessing of commercial and other personal data by law enforcement authorities' (ibid., paragraph 27), and stated that 'legislation permitting public authorities to gain access to the contents of electronic communications on a generalised basis must be regarded as compromising the essence' of the right to privacy as established by Article 7 of the EU Charter (ibid., paragraph 28).¹⁶¹

The European Parliament has also witnessed a series of specific initiatives related to AI. The creation of budget for an Observatory on Artificial Intelligence was requested but rejected by the EP Committee on Budgets, leading later to concerns regarding the possible continuation of such initiative as a working group (EP, P9_TA-PROV(2020)0084, paragraph 27).

Ongoing work on a **Draft Report on Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminal matters** (LIBE, 2020/2016(INI)) was already mentioned. The document notably includes a call for a **moratorium on the deployment of facial recognition systems for law enforcement**, as well as a set of recommendations including the establishment of compulsory fundamental rights impact assessments, and the creation of a clear and fair legal regime for assigning legal responsibility for the potential adverse consequences produced by advanced digital technologies in the field.

On 18 June 2020, the European Parliament took the decision to set up a special committee on artificial intelligence in a digital age. The committee is due to begin its activities after the summer recess.

5.4. The role of ethics

The softening, or blurring, of boundaries between **law and ethics** described in relation to the *Ethics Guidelines* of the AI HLEG is not exclusive to the Group's endeavours, but rather a more pervasive trend in European discussions around digital regulation. In this context, a certain type of 'post-compliance ethics', opening up a space characterised as potentially 'unlimited', has been particularly noticeable in EU-level debates over the past years (Floridi, 2018). The intersections between data, ethics, AI and EU policy are complex, and do not systematically help delineating a consistent picture of what could be a

¹⁶⁰ Its paragraph 10 includes a reference to the legal profession: (the EP) '(u)nderlines that while automated decision-making processes can improve the efficiency and accuracy of services, humans must always be ultimately responsible for, and able to overrule, decisions that are taken in the context of professional services such as the medical, legal and accounting professions, and for the banking sector; recalls the importance of supervision or independent oversight by qualified professionals in cases of automated decision-making where legitimate public interests are at stake'.

¹⁶¹ The Resolution also echoed the fact that 'intelligence services of third countries and Member States have increasingly been relying on the processing and analytics of such [big data] datasets, which are either not covered by any legal framework or, most recently, have been the subject of legislation the compatibility of which with Union primary and secondary law raises concerns and is yet to be ascertained' (EP, P8_TA(2017)0076, paragraph D).

role of ethics in this field, in particular in relation to the (certainly different, possibly complementary) role of law.¹⁶²

Work on AI and ethics is globally **extremely voluminous**, resulting from many years of ‘intense efforts’ of a ‘diverse set of stakeholders’ including both public authorities and private companies (Jobin, Ienca & Vayena, 2019). In terms of content, although relevant documents typically mobilise coinciding terminology – referring for instance to transparency, justice and fairness, responsibility, freedom and autonomy, trust, dignity, sustainability, or solidarity – this does not necessarily mean that there is a consensus on the meaning of these notions, or their implications (idem). AlgorithmWatch launched an AI Ethics Guidelines Global Inventory to compile frameworks and guidelines that seek to set out principles of how systems for automated decision-making can be developed and implemented ‘ethically’, and it compiles more than 160 guidelines.¹⁶³

Ethics already played a significant role in global discussions around **big data** a few years ago, especially in relation to research.¹⁶⁴ Discussions on AI and ethics have surfaced in all fields.¹⁶⁵ An additional, whole strand of literature regards the possibility to embed ethics requirements into the design of AI systems.¹⁶⁶ According to some authors, the focus on ethics as the main prism through which to address AI’s risks might only constitute ‘a phase’, preceding a different, regulatory phase (Lapinski, 2018). There is any case scepticism about the added value of ‘the move to ethical principles’ in terms of measurable effects, and also because they often lack enforcement mechanisms (AI Now Institute, 2018, 9). The frequent use of the word ‘ethics’ in connection with algorithms may be, it has been stated, ‘an indicator for a tactical move by some actors who want to avoid strict regulation by pointing to non-formal normative concepts’ (MSI-NET, 2018, 42).¹⁶⁷ Some initiatives have framed themselves visibly as different from ethics-centred approaches – for instance, by openly embracing human rights considerations as prime focus.¹⁶⁸

A number of European **national AI strategies** devote particular attention to ethical issues (Van Roy, 2020; regarding Cyprus, 20; Czech Republic, 24; Finland, 35; France, 36; Germany, 42). In Denmark, the

¹⁶² According to a 2019 guide published by The Alan Turing Institute, ‘AI ethics is a set of values, principles, and techniques that employ widely accepted standards of right and wrong to guide moral conduct in the development and use of AI technologies’ (Leslie, 2019, 3). The main connection between AI ethics and the law would be that AI ethics moral vocabulary draws on human rights discourse, which in its turn ‘draws inspiration from the UN Declaration of Human Rights’ (ibid., 8). The guide calls for such development to be accompanied by an ‘ethical platform’ requiring a set of actionable principles, called FAST Track Principles, composed of four notions: Fairness, Accountability, Sustainability, and Transparency (ibid., 7).

¹⁶³ As explained by two contributors of the team working on such inventory, ‘AI ethics’ as found in that database ‘are often positioned between instrumental-economic and ethical perspectives. AI ethics in this sense is rather business ethics’ (Gießler & Haas, 2020).

¹⁶⁴ This was notably the case after the so-called ‘emotional contagion Facebook scandal’ of 2013, regarding the algorithmic manipulation of emotions of Facebook users for research purposes. At that time, US scholars were particularly interested in the fact that the research had not been subject to standard research ethics procedures because it was not publicly funded but corporate research, which in the US has strong legal implications insofar as research ethics are concerned. Thus, there were notably calls to bring closer research ethics and corporate ethics, concretely around data and big data practices (boyd, 2016).

¹⁶⁵ In February 2020, for instance, the US Department of Defense adopted five principles for the ‘ethical development’ of AI capabilities (US Department of Defense, 2020). As another example, in 2019 the French Armed Forces Ministry published a report of the AI Task Force calling for the establishment of a ministerial committee to, in particular but not exclusively, on the ethical issues that AI applications in the military sphere could raise (AI Task Force, 2019).

¹⁶⁶ Which is connected to the question of whether one can somehow operationalise in mathematical or computational terms ‘ethics requirements’; arguing for instance that designing AI systems, as well as designing fairness requirements into them, is not a purely mathematical exercise but involves a series of important trade-offs, very much like human decision-making in the criminal justice system in general: McNamara, et al, 2019, 112.

¹⁶⁷ In some cases, embracing ethics in this context is openly portrayed as a way further promoting the spread of AI, to be endorsed by those who ‘favour security over interests such as privacy’ (McCarthy, 2019).

¹⁶⁸ Such is the case for example of the [The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems](#) of May 2018.

government emphasised the need to develop and use AI *'within the relevant legislation, and with respect for the rights of citizens'*, which would inter alia require that *'businesses and the public authorities must have strong focus on **data ethics**'* (Ministry of Finance and Ministry of Industry, Business and Financial Affairs of the Danish Government, 2019, 8). In Belgium, AI 4 Belgium published a report noting it is *'important to introduce ethical guidelines to support the development, deployment and use of AI'* (AI 4 Belgium Coalition, 2019, 14), and suggested *'a new role could be created to monitor compliance with these ethical principles: **the digital ethicist**'* (ibid., 15).

In 2018, Germany's Federal Government set up the Data Ethics Commission (Datenethikkommission), and mandated it to develop ethical benchmarks and guidelines as well as specific recommendations for action, aiming at protecting individuals, preserving social cohesion, and safeguarding and promoting prosperity in the information age, to tackle concretely a number of key questions clustered around algorithm-based decision-making (ADM), AI and data (Datenethikkommission, 2019, 13). The Data Ethics Commission openly stated that, in particular for issues with heightened implications for fundamental rights, regulation is necessary and **cannot be replaced by ethical principles** (ibid., 15).

In many cases, there is currently at least some degree of **ambiguity** when **EU institutions** touch upon the role of law and ethics in relation to AI. In the *A Union that strives for more* 2019 document of Von der Leyen, there was much emphasis on the importance of *'ethical boundaries'* and the need for a coordinated *'European approach on the human and ethical implications of Artificial Intelligence'*, together with a call for legislation (Von der Leyen, 2019, 6). There was also a reference to preserving *'ethical standards'* that appear to have a connection with the GDPR, as if, perhaps, the GDPR was about ethical standards (idem).

The role of ethics in this field as inferred from the contributions of the European Parliament is variable. In the European Parliament's 2017 **Resolution on Fundamental Rights Implications of Big Data**, it called for *'Member States' law enforcement authorities that make use of data analytics to uphold the highest standards of **ethics** when analysing data'*, as a way of contributing to non-discriminatory practices (P8_TA(2017)0076, paragraph 32). The Resolution adopted by the European Parliament on 12 February 2020, on automated decision-making processes and consumer protection, describes *'ethical guidance, such as the principles adopted by the Commission's High- Level Expert Group on Artificial Intelligence'* as a *'starting-point'* (paragraph D) to move towards *'a common EU approach to the development of automated decision-making processes'* that would help secure the benefits of those processes and mitigate the risks across the EU, avoid fragmentation of the internal market, and enable the EU to better promote its approach and its values around the world' (paragraph E). The EDPS, which has as main tasks to monitor compliance and provide guidance about data protection law, devoted in the past years particular attention to *'digital ethics'*.¹⁶⁹ Somehow as a counterpoint, the Council discussed the possibility to actively encourage the *'development of AI ethics into a discipline of its own'* (Council of the EU, 12224/19, 3).

The ambivalent relation between ethics and fundamental rights also surfaces in other *fora*, including the Council of Europe. The 2018 CEPEJ's **European Ethical Charter on the Use of AI in the Judicial Systems and their Environment** (CEPEJ, 2018) sets out as first principle the *'Principle of respect for fundamental rights'*, described as ensuring *'that the design and implementation of artificial intelligence*

¹⁶⁹ The report published in 2018 by its Ethics Advisory Group (EAG) listed a series of challenges that such Group regarded as relevant for digital ethics, among which was mentioned *'predictive policing'*. The EAG observed *'predictive policing' 'comes with a new set of challenges'*, such as limited predictability of law enforcement decision-making processes, or risks of bias (EAG, 2018, 29). *'Police may risk becoming more interested in the patterns than in the substance, more concerned with the prediction than in observable facts'*, noted the EAG (idem). The report was prepared *'in anticipation of the'* GDPR (ibid., 7), but does not mention the LED, or make any other particular reference to data processing for law enforcement purposes.

tools and services are compatible with fundamental rights' (ibid., 6). The **Guidelines on Artificial Intelligence and Data Protection** published in January 2019 by the Council of Europe's Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108) refer to 'ethical values' as an element of the 'wider view' to be adopted when considering the possible outcomes of data processing. In this sense, the Guidelines state such wider view should '*consider not only human rights and fundamental freedoms but also the functioning of democracies and social and ethical values*' (T-PD(2019)01, 1).

The Preamble to the Council of Europe's *Recommendation on the Human Rights Impacts of Algorithmic Systems* of April 2020 points out that even if 'ongoing public and private sector initiatives intended to develop ethical guidelines and standards for the design, development and ongoing deployment of algorithmic systems' constitute 'a highly welcome recognition of the risks that these systems pose for normative values', these initiatives nevertheless 'do not relieve Council of Europe member States of their obligations as primary guardians' of the ECHR.

5.5. EU-funded support of AI

Much of the EU support of AI and related technologies takes place through funding. EU funds support AI, including law enforcement related AI, through a variety of means – some of them are openly related to research in this area, while others do not primarily have a research component. Investment in AI has been a priority for the European Commission, which **doubled its investments in AI** in Horizon 2020 and had put forward plans for significant investment from Horizon Europe and the Digital Europe Programme (COM(2019) 168 final, 1).¹⁷⁰

Some AI-related funded projects deployed with the support of EU funding have generated controversy. It is the case, for instance, of a 'smart policing' project launched in 2019 by the Greek Police in cooperation with a telecommunications operator, involving the delivery of 'smart devices' with integrated software enabling facial recognition and automated fingerprint identification, among other functionalities, supported by Internal Security Fund (ISF) 2014-2020 of the European Commission (Chelioudakis, 2020). Another example is the City of Marseille's project called '*Big Data de la tranquillité publique*', launched in 2016 to foster the collection of data from institutional partners in order to prevent certain events before they would happen, co-funded by the EU through the European Regional Development Fund (FEDER) scheme¹⁷¹ - developments currently at the centre of local contestation against surveillance, notably organised by the NGO La Quadrature du Net (Meaker 2020).

Current funding of AI research in this field is marked by two issues with potentially problematic fundamental rights implications: first, the **lack of solid specific guidance accompanying the submission and selection** of proposals, and, second, the serious limitations regarding the **transparency of funded research**, including, most notably, information about their compliance with legal and ethics requirements.

Regarding the first point, research proposals are typically submitted, in addition to a scientific evaluation, to an 'Ethics Appraisal Procedure'¹⁷² appraising compliance with '**ethics requirements**'. These requirements refer to a series of issues globally entrenched in research ethics, among which the European Commission places **data protection**. In a gesture that can result in more confusion than

¹⁷⁰ The President of the European Commission stated before her nomination that she would '*make sure that we prioritise investments in Artificial Intelligence, both through the Multiannual Financial Framework and through the increased use of public-private partnerships*' (Von der Leyen, 2019, 13).

¹⁷¹ See page [Le Big Data de la tranquillité publique](#).

¹⁷² This is the procedure applied for Horizon 2020 projects, more information on the relevant [webpage](#).

clarity, 'ethics reviewers' carrying out reviews for the European Commission may decide, during the relevant procedures put in place, to incorporate into the grant agreement to be signed by the relevant consortium a series of 'ethics requirements', some of which might refer, for instance, to the necessity to comply with specific GDPR provisions, despite this being, as such, a legal obligation.

Although some of the ethics reviewers might have been recruited on the grounds of their data protection expertise, not all of them necessarily have precise knowledge on EU data protection law, and in any case they are systematically invited to attempt to agree, if possible, on a consensual decision to pick up certain pre-written statements on '**data protection requirements**'. The wording of these requirements appears to be a legacy of past review activities, and might in many cases not be fully applicable to the case at stake, as the actual obligations of the relevant data controllers and processors are to be analysed taking into account the specific national laws applicable to them. As a result, the utility of the whole exercise is questionable, and can often result in perplexity among the selected applicants, nevertheless faced sometimes with no other option than to attempt to comply with the imposed 'ethics requirements' resulting from this procedure in order to obtain the sought funding.

The most detailed advice made available to applicants suffers from a number of **weaknesses**. Guidance is provided in a document presented as 'reference document' on 'ethics and data protection' at the relevant website, but bearing a disclaimer according to which '*it does not constitute official EU guidance*' (EC, 2018, 1). The document insists on the need to rely on the consent of data subjects to process personal data about them,¹⁷³ which might actually not be the case. Another statement warns that '*if your proposal uses data processed or provided by authorities responsible for preventing, investigating, detecting or prosecuting criminal offences, Directive (EU) 2016/680 may also apply*': actually that would be the case only if said authorities would be, in the context of the research project, carrying out activities for the mentioned purposes – if processing data for research purposes, it is still the GDPR that applies.¹⁷⁴

That document also proclaims that: '*If the goal of the project is to **develop surveillance technologies or techniques for law enforcement purposes**, your proposal should explain why the surveillance can be deemed necessary and proportionate in a democratic society, in accordance with EU values, principles and laws*' (EC, 2018). This issue, however, is not mentioned at all in relevant 'ethics issues check-list', and is also absent from the guidance provided to applicant to self-assess compliance with ethics requirements (EC, *Horizon 2020 Programme Guidance*, 2019). It is thus unclear how often it might be considered, if at all.

Regarding transparency of funded research, it must be noted that generally research proposals foresee the publication of some results, while planning for others a 'confidential' or 'EU restricted' status. Confidentiality of some results is often necessary, especially in the realm of security. The European Commission has however, in principle, the ability to impose that some of the activities and results of the funded project shall be open to public scrutiny. The European Commission has published guidance, prepared by Directorate-General for Migration and Home Affairs (DG HOME), which advised for the **classification of certain types of research and research results** in certain specific fields, such as for instance border security, in Horizon 2020 projects (EC, *H2020 Programme Guidance*, Version 2.2, 2020).

Such guidance does not mention any need to classify deliverables with information regarding legal and ethical standards. It actually also does not recommend classifying research in the domain of

¹⁷³ In this sense: "*Whenever you collect personal data directly from research participants, you must seek their informed consent by means of a procedure that meets the minimum standards of the GDPR*" (EC, 2018, 10). As it is widely known and has recently been again confirmed by the EDPB, the GDPR foresees different legal grounds on which research can be based (EDPB, Guidelines 03/2020, 2020).

¹⁷⁴ See Art. 2(1) of LED.

‘intelligent surveillance’, defined as ‘the use of pattern recognition and other artificial intelligence techniques to analyse data obtained from more conventional security devices, with the aim of identifying behaviour deemed suspicious or anomalous with regard to the given legal and social context’.¹⁷⁵

What is striking insofar as EU-funded research on AI and security is concerned, however, is the lack of public information available on many of the funded projects, including on their **compliance with legal and ethics standards**. In this sense, the NGO Privacy International has argued that it ‘is unclear how some of the funded projects being supported – such as those aimed at monitoring potential migrants to Europe – comply with EU rules, including on data protection’ (Privacy International, 2020a).

A particularly commented case was the Horizon 2020-funded project iBorderCtrl, aiming at creating an automated border security system incorporating ‘automated deception detection’, in spite of concerns about the scientific and ethical opportunity of such an approach. Also commented was the fact that despite the existence of concerns around the project many of its results remained hidden from public scrutiny (Leufer & Jansen, 2020). In May 2020, results of funded activities such as deliverables on the ‘ethics of profiling’ and ‘the risk of stigmatization of individuals’, or a ‘EU wide legal and ethical review’ are described at the project website as submitted to the European Commission but ‘confidential’, and thus not available to other researchers or the public.¹⁷⁶

Lack of transparency can dramatically affect the trust of the population in the role of the EU in supporting AI in law enforcement and criminal justice.

Example 1: Real-time surveillance of public space

The project *Artificial Intelligence and Computer Vision for Real-time Diagnosis of Public Spaces* was funded in 2019 (H2020-SMEInst-2018-2020-1) to bring to the market a solution for **real time diagnosis of public spaces**, including the Urban Security Eye product line, targeting detection of terrorism, suspicious behaviour and insecure spots (EC, 2019, 90). The [website of the coordinator](#) claims that their products are ‘**so anonymous that GDPR becomes irrelevant**’, but does not appear to offer any specific further clarification about legal and ethical implications of their work.

Example 2: Autonomous robots at the borders

The project [ROBORDER](#) aims at developing fully-functional autonomous **border surveillance systems** with **unmanned mobile robots including** aerial, water surface, underwater and ground solutions, with multimodal sensors. An ‘heterogenous robot system’ shall be enhanced with detection capabilities for early identification of criminal activities at border and coastal areas. All deliverables concerned with ethics issues, including an ‘**Ethical Code**’, appear to have been regarded as **fully confidential** and thus not accessible even partially [according to the project website](#).

¹⁷⁵ The document does however note that although ‘(c)lassification is currently not foreseen in this area’, ‘(t)his may change in the future’ (EC, H2020 Programme Guidance, Version 2.2, 2020).

¹⁷⁶ See the list of project deliverables [here](#).

Example 3: Pre-occurring crime prediction and prevention

The InterCONnected NEXt-Generation Immersive IoT Platform of Crime and Terrorism DetectiON, PredictiON, InvestigatiON, and PreventiON Services (CONNEXIONS) project describes itself as a solution encompassing the entire lifecycle of law enforcement operations, from 'pre-occurrence crime prevention and prevention' to 'post-occurrence' investigations. It appears to be concerned *inter alia* with [multilingual automatic speech recognition](#) of online material, and its use cases cover [threats at public events such as football matches and festivals](#). According to its [description](#) at the Community Research and Development Information Service (CORDIS), the project, launched in September 2018, planned to '**adopt ethics and privacy by-design principles**'. [No public deliverable appears to be available online](#) in July 2020.

Example 4: Intelligent dialogue-empowered bots for early terrorist detection

TENSOR was a Horizon 2020 project was about developing a platform to help law enforcement agencies in the early detection of terrorist activities, radicalisation and recruitment – as explained on the [project website](#). The platform was described as integrating 'a set of **automated and semi-automated tools**', including tools for the monitoring of online content but also 'Internet penetration through intelligent dialogue-empowered bots'. A whole funded [Work Package](#) of the funded project was devoted to developing 'the legal and ethical framework that will underpin' such developments; the project's [flyer](#) asserted such work would '*ensure that the solutions are shaped by the privacy and data protection laws that protect the freedom of citizens across Europe in their use of the internet*'. The project ended in November 2019. In July 2020, the project website announced as publicly available a total of four deliverables, none of them about the relevant legal and ethical framework, and **none of them actually downloadable** (see [here](#)).

6. COVID-19 OUTBREAK RESPONSE

The COVID-19 outbreak has had dramatic effects all over the world, and the responses to the crisis have triggered, at least potentially, **a variety of fundamental rights implications**. This section highlights some of the issues that appear as particularly relevant for a reflection on the impact on EU fundamental rights of AI in the field of law enforcement and criminal justice. The section focuses on two main themes: first, data-driven responses which have a major impact on the collection and processing – and potentially, availability - of data about individuals, and, second, initiatives undertaken directly by law enforcement authorities.

Many other issues might nevertheless influence developments in this field. Among them, there are efforts to accelerate the **digitalisation of public services**, and most notably the administration of justice.¹⁷⁷ The work of law enforcement has also been affected, notably by changing the prominence of certain types of crimes (such as, for instance, those related to domestic abuse). According to Interpol, *'cybercriminals are capitalizing on the anxiety caused by COVID-19 through various cyberattacks such as data-harvesting malware, ransomware, online scams and phishing'* (Interpol, 2020).

There has been a general trend to adopt or re-purpose technological solutions, pursuing multiple objectives, and building on quick adaptations by private companies (Gershgorn, 2020). Generally speaking, **the processing of data and surveillance practices** have increased around the world. In China, facial recognition software was adapted to detect the wearing of masks already in February (Jain, 2020);¹⁷⁸ in Moscow, facial recognition would have been used to enforce quarantine obligations since March (Habersetzer, 2020). In the US, the company Athena Security started marketing also in March a 'Fever Detection COVID19 Screening System' allegedly detecting fever 'through AI', although not using facial recognition as such (Cox, 2020).¹⁷⁹ In Singapore, municipal authorities decided to test a four-legged robot marketed by the US robotics design company as a tool to remind park visitors to keep a safe distance from one another, notably by broadcasting pre-recorded messages (Vincent, 2020).

Some measures have directly concerned issues related to **automated decision-making**. In April 2020, US civil society organisations expressed concerns with the use of the risk assessment tool Pattern, originally conceived to attribute scores on 'general recidivism', in order to determine which incarcerated individuals would receive 'priority treatment' in transfer and release decisions in the context of the Covid-19 outbreak (Cook, 2020). The organisations expressed that algorithmic recidivism risk scores should not inform assessment of medical risk, or play any role in determining who receives access to adequate healthcare. They also stressed that the tool had been previously assessed as *'likely to perpetuate racial disparity in decision-making'*, and suffered from *'assumptions built into its design'*.

¹⁷⁷ Already at the beginning of April 2020, the EU Justice ministers agreed to intensify efforts to guarantee the digitalisation of their judicial procedures, as well as a secure digital channel for all judicial cooperation procedures (European Commission, 7 April 2020). Quick moves to online solutions created problems in some Member States; in Italy, for instance, an association of criminal lawyers requested the involvement of the Italian DPA to assess the implications of a forced switch of certain criminal justice procedures to digital issues, in particular through the use of Skype for Business e Teams, both of Microsoft Corporation (Unione Camere Penali Italiane, 2020).

¹⁷⁸ Facial recognition is as such not necessarily prevented by the fact that individuals might wear masks, as technology providers have worked to offer products designed for this scenario (Yang, 2020).

¹⁷⁹ According to the media, Clearview AI would be currently proposing to public authorities in the US 'contact tracing' solutions using images from surveillance cameras, through which it would be able to gather information on the identity of individuals' contacts (Thomsen, 2020; cf. also Grind, McMillan & Mathews, 2020). It is unclear whether what is being offered would be real-time facial recognition technology for purposes of contact tracing in response to the COVID-19 pandemic (Markey, 2020).

encode bias against Black people, Latino people, poor people, unhoused people, and people with mental illness' (ibid., 2).

The possibility of put mobile **applications ('apps')** at use sparked remarkable creativity in many. 'We've had ideas for apps from people as young as 14 or 15, from individuals, from small start-up companies, from huge, globally-based companies', declared on 25 March 2020 the Executive Director of the WHO Health Emergencies Programme, Michael Ryan, before adding: 'It's been the most outstanding and most amazing outpouring of support and collaboration that I have seen in my career' (WHO, 2020, 1). Multiple types of apps have been developed in response to the outbreak, including self-assessment apps, apps for the collection of data for the purposes of research, apps related to the notion of 'immunity certificates', and 'contact tracing' apps, as well as apps combining different of these functions.¹⁸⁰ Deployed apps have been made available to the public accompanied by a variety of measures and approaches to the management of the crisis.¹⁸¹

Systems monitoring the temperature of individuals have proliferated. Italy, for instance, adopted in March 2020 rules foreseeing the real-time measurement of temperature to access a variety of spaces, for workers but also for others.¹⁸²

While data processing and surveillance measures where increasing, some fundamental rights safeguards were put under pressure. In May 2020, the Hungarian government suspended the application of a number of GDPR data subject rights as part of the coronavirus emergency measures, and established time limits for the exercise of remedy rights, including the right to lodge a complaint and the right to an effective judicial remedy as guaranteed by the GDPR. Civil society representatives contacted the EDPB, calling on them to ask the European Commission to launch an infringement procedure against Hungary (Massé, 2020, 16).¹⁸³

Generally speaking, assessments of Covid-19-related developments are cautious, and place emphasis on the need to monitor them closely. Some have warned about the possible strengthening of '*the temptation of adopting technology-enabled mass surveillance to secure society*' (Renda, 2020). In its reaction to the European Commission's *White Paper on AI*, the EDPS has observed that some AI-related measures put forward in the context of the fight against the pandemic represent in themselves, or enable, a kind of 'function creep' which can have '*a chilling effect in democratic societies*' (EDPS, Opinion 4/2020, 20).

¹⁸⁰ In some cases, apps made available were pre-existing apps, re-purposed in the context of the outbreak. In Colombia, the NGO Fundación Karisma analysed the three local apps and discovered an important set of technical vulnerabilities, as well as serious gaps in terms of transparency (Labarthe & Velásquez, 2020b). The NGO pointed out for instance that the most popular app was actually built on an app developed by Brazil to track in 2014 other diseases such as dengue and Zika, and requested data not relevant in the current situation (Labarthe & Velásquez, 2020a).

¹⁸¹ South Korea, for instance, put in place a 'track and trace' model relying on testing and rigorous isolation, as well as public dissemination of person's private lives. Public authorities have indeed publicly shared information on the movements of persons tested infected, presumably to help others assess whether they might be infected. In practice, this has led to problematic disclosures, as well as instances of blackmail (Kim, 2020a). Eventually, media reports linking a series of infections to gay clubs generated fear of a backlash against an already discriminated community (Kim, 2020b).

¹⁸² See: Protocollo condiviso di regolamentazione delle misure per il contrasto e il contenimento della diffusione del virus Covid-19 negli ambienti di lavoro tra Governo e partiti sociali del 14 marzo 2020, section 2.

¹⁸³ The EDPB discussed the issue on 8 May 2020 (EDPB, 8 May 2020), but concluded that further explanation was necessary and has thus requested that the Hungarian Supervisory Authority to provide further details on the scope and the duration of the measures, as well as its opinion on the necessity and proportionality of these measures.

6.1. Data-driven responses

An important dimension of EU's response to the outbreak might be described as data-driven, or data-based (Zanfir-Fortuna, 2020).¹⁸⁴ Concretely, two major types of data-driven measures have been discussed and supported at EU level: the use of **location data**, as made available by telecommunication providers to public authorities, and the use of **ad hoc apps**,¹⁸⁵ notably apps relying on '**contact tracing**' by Bluetooth.

On 24 March 2020, the European Commission revealed that it had explored with telecommunications operators the possibility to collaborate to tackle the coronavirus outbreak. Thierry Breton, the Internal Market Commissioner, had discussed with the heads of Europe's telecommunications companies and the Global System for Mobile Communications association (GSMA) to explore options for sharing geolocation metadata in order to map and forecast the spread of the virus.¹⁸⁶ Many Member States would have taken location-tracking measures in response to the spread of the coronavirus disease, mainly by working with telecommunications companies to map population movements using so-called 'anonymised' and aggregated location data (Dumbrava, 2020).

The Dutch DPA is strongly critical of the very possibility to regard as 'anonymous' the location data at stake. In this sense, it made public in April 2020 a document on the anonymity of aggregated telco location data which concluded that '*(a)nanimisation [sic] is hard. Anonimisation [sic] of location data is almost undoable*' (Autoriteit Persoonsgegevens, 2020). When announcing that the EDPB had published guidance on the matter, the Dutch DPA stressed again the importance it attaches to this message.¹⁸⁷ The concerns of the Dutch DPA with inappropriate anonymisation are *inter alia* connected to **the possible interest in any stored data by the police** and security services (DutchNews, 2020).

On 8 April 2020, the European Commission published a **Recommendation on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data** (C(2020) 2296 final), which argued it '*is therefore necessary to develop a common approach to the use of digital technologies and data in response to the current crisis*' (ibid., 1).

Regarding **location data**, it explained that '*(c)ertain companies including telecommunications providers and major technology platforms have published or made available to public authorities anonymised and aggregated location data*' (ibid., 5). Regarding apps, it noted that, since the beginning of the COVID-19 crisis, a variety of mobile applications had been developed, '*some of them by public authorities*' (ibid., 3), and highlighted that '*w)arning and tracing applications are useful for Member States for contact tracing purposes and can play an important role in containment during de-escalation scenarios*' (ibid., 4). The Recommendation already hinted that '*(i)n order to detect proximity*

¹⁸⁴ An overview of key documents adopted by EU institutions, agencies and bodies on these matters can be found at [VUB's Data Protection Law & Covid-19 Observatory](#); concretely, [here](#).

¹⁸⁵ China had originally deployed a system based on creating a record of 'proximity events' between individuals through mobile apps, and displaying a green, amber or red code with direct implications in terms of movement restrictions (Ferretti et al., 2020, 5). Building on such an initiative, researchers proposed an algorithmic-based model that would enable automated recommendations of '*risk-stratified quarantine and physical distancing measures*', as well as eventual refinements of the algorithm to recommend, for instance, the quarantining of whole households or areas (ibid.). The researchers advocating for this approach put forward a series of possible requirements to obtain the trust of the population, among which can be cited '*the use of a transparent and auditable algorithm*' and '*careful oversight of and effective protections around the uses of data*' (ibid.).

¹⁸⁶ The European Commission's approach to the use of telecommunications data has been described as being 'largely based' on an approach put in place since mid-March by the Belgian 'Data against Corona' taskforce (Mascini, 2020). It relies on telecommunication operators making available location data about mobile phone users in order to elaborate pictures of mobility trends, using data regarded as 'anonymised'.

¹⁸⁷ By linking to its own document in a footnote; see [the announcement](#).

encounters between users of different contact tracing applications (a scenario most likely to happen among people moving across national/regional borders), interoperability between applications should be envisaged' (ibid., 5).¹⁸⁸

For these reasons, the European Commission announced the setting up 'a process for developing a common approach, referred to as a Toolbox, to use digital means to address the crisis', with a focus on a **'pan-European approach for the use of mobile applications, coordinated at Union level', and 'a common scheme for using anonymized and aggregated data on mobility of populations'** (ibid., 7).

The use of contact tracing apps is sometimes referred to as **'digital' contact tracing**, opposed to **'manual' contact tracing**. The latter would describe the practice of re-tracing, with the help of humans, the contacts of infected persons or persons that are believed to have potentially been infected. The former would consist of automatically keeping track of 'proximity events' between devices, that might be regarded as probably implying a contact. Both so-called 'manual' and 'digital' contact tracing trigger significant interferences with the fundamental rights to privacy and data protection, in a number of ways.¹⁸⁹ 'Manual' contact tracing, in this sense, typically implies that the person infected is faced with a set of particularly intrusive questions about who they have met in the days before the testing, and which kind of contacts they have had, if any.

From an EU fundamental rights perspective, the main difference between so-called 'manual' and 'digital' contact tracing is that the first one is based on an intervention that takes place, in principle, only after a person has been tested positive, or presents symptoms, or has been in contact with somebody who has been tested positive or has symptoms – and thus affecting what could be regarded, in principle, a minority of persons.¹⁹⁰ Contact tracing via apps, in contrast, **aims at initiating the collection of data about the relevant contacts of individuals in advance**, targeting a number of persons as broad as possible, in a **preventive approach**, 'just in case' information about such contacts would later be of use. From this standpoint, in principle **the more individuals are engaged in data collection, the better**. Some experts have claimed in this sense that the desirable goal from this perspective would be *'an app that is used by every inhabitant of all the European Member States'* (Mascini, 2020).

Developments at EU followed then quickly, notably with the adoption on 15 April 2020 of a document titled **Common EU Toolbox for Member States: Mobile applications to support contact tracing in the EU's fight against COVID-19** by the eHealth Network (eHealth Network, 15 April 2020).¹⁹¹ On 16 April

¹⁸⁸ The Recommendations specified: *'National health authorities supervising infection transmission chains should be able to exchange interoperable information about users that have tested positive with other Member States or regions in order to address cross-border transmission chains'* (C(2020) 2296 final). There have been concerns that some Member States might require from those wishing to enter into their own territory the use of a certain app, or of a certain types of apps, and which might require that at least nearby Member States offer appropriate solutions. These concerns were for instance voiced at parliamentary discussions in Luxembourg (Chambre des Députés du Grand-Duché de Luxembourg, 2020).

¹⁸⁹ Additionally, one does not exclude the other, and the distinction between them might be occasionally blurred, for instance when public authorities publicly call the population, including healthy individuals, to keep personal diaries of all the persons they encounter. In the EU, DPAs have already provided some input of proposed or deployed apps, as well as on 'manual' contact tracing. The Belgian DPA, in a first analysis of a legislative proposal concerning the creation of a database for public authorities in the context of the fight against the spread of COVID-19, primarily concerned with 'manual' contact tracing but potentially also of use for app-enabled contact tracing, concluded that the proposal did not satisfy the requirements of necessity and proportionality (Belgian DPA, 29 April 2020).

¹⁹⁰ Depending of the level of spread, such minority might constitute a large number of persons; all these persons, nevertheless, would correspond to a category of persons deemed of interest for a special reason.

¹⁹¹ The eHealth Network is a voluntary network providing a platform of Member States' competent authorities dealing with digital health; it was set up under Directive 2011/24/EU, Directive 2011/24/EU of the European Parliament and of the Council of 9 March 2011 on the application of patients' rights in cross-border healthcare, OJ L 88, 4.4.2011, p. 45–65.

2020, the European Commission published its **Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection** (C/2020/2523).

In a **Resolution** adopted on **17 April 2020**, the **European Parliament** made reference to what was then the European Commission's '*plan to call on telecoms providers to hand over anonymised and aggregated data in order to limit the spread of COVID-19, as well as to 'national tracking programmes already in force, and of the introduction of apps allowing authorities to monitor movements, contacts and health data'*' (P9_TA(2020)0054, paragraph 51). The European Parliament stressed the importance of guaranteeing that **any such measures would be compliant with EU data protection law**, and called on the European Commission and the Member States '*to publish the details of these schemes and allow for public scrutiny and full oversight*' by DPAs (paragraph 53). In relation to contact tracing apps, it notably demanded all **storage of data to be decentralised**, and for the functioning of the apps to be transparent, allowing verification of underlying protocols and the code itself (paragraph 52).

On 13 May 2020, the eHealth Network published **Interoperability guidelines for approved contact tracing mobile applications in the EU** (eHealth Network, 13 May 2020), proposing guidelines for cross-border interoperability '*of approved contact tracing mobile apps and associated procedures*' (ibid., 3). The Guidelines also announced that the European Commission was to set up 'a Wiki space (confluence page) also to engage with app developers (ibid., 10).

On 16 June 2020, the European Commission announced that Member States had '*agreed on a set of technical specifications to ensure a safe exchange of information between national contact tracing apps based on a decentralised architecture*' (EC, 16 June 2020). The announcement appears to refer to a document by the eHealth Network addressed to the Member States and the European Commission on **detailed interoperability** elements (eHealth Network, 16 June 2020). The eHealth Network also published another document about interoperability, dated from 12 June 2020 (and titled *Interoperability specifications for cross-border transmission chains between approved apps: Basic interoperability elements between COVID+ Keys driven solutions*) (eHealth Network, 12 June 2020), which notably states that '*(i)n order to reduce the data downloaded by the user's apps, the apps should be able to identify the countries where a user has been during the past 14 day*' (ibid., 6).

The **EDPB** published in June 2020 a *Statement on the data protection impact of the interoperability of contact tracing apps* (EDPB, 16 June 2020). The statement stressed the importance of transparency issues, noting that interoperability leads to additional processing and disclosure of data to additional entities (ibid., 2).

In July 2020, the European Commission published an overview of the situation (EC, July 2020), according to which a **vast majority of Member States** have implemented or are planning to implement voluntary and temporary mobile apps that support contact tracing as part of public health strategies to combat the COVID-19 pandemic (ibid., 4) – relying both on 'decentralised' and 'centralised' approaches. Additionally, many Member States were identified as having implemented or planning to implement other mobile applications such as symptom checkers (idem). Although some Member States have adopted specific legislation related to such apps, some deny the need for such specific legislation (ibid., 8). Regarding interoperability, the overview document referred to the plans by the European Commission to 'set up a gateway service' to facilitate data transfers between national contact tracing apps and servers (ibid. 12).

According to the European Commission, Member States are currently '*working on a common approach for modelling and predicting the evolution of the COVID-19 pandemic through anonymised and aggregated mobility data*' (EC, July 2020).

These developments are of relevance for a discussion of AI in the field of law enforcement and criminal justice primarily because they set in motion novel data practices, potentially encompassing unprecedented availability of certain types of data, which could eventually be made accessible for law enforcement related purposes.

6.2. Law enforcement initiatives

A particularly visible use of technology by law enforcement authorities during the pandemic has concerned drones. In Italy, for instance, some monitoring practices with drones were exceptionally authorised *inter alia* on the grounds that there was a limited risk of hurting individuals, as there were less individuals in public spaces due to lockdown measures (Dotta, 2020). The way in which drones appear to have been used to ensure compliance with movement restriction measures related to Covid-19 has been decried by civil society in some Member States, such as Greece (Chelioudakis, 2020).

In France, the Conseil d'État published on 18 May 2020 a Decision on surveillance by drones, following action by two civil society organisations - La Quadrature du Net, and La Ligue des Droits de l'Homme (Conseil d'État, 2020). The decision concerned measures allowing to capture images by drone, in order to contribute to compliance with lockdown and social distancing measures. The drones at stake are equipped with a camera and a speaker allowing to transmit messages to the public. According to the authorities responsible for their use, the drones did not capture any personal data, notably because the images were only monitored live, in real time, and the team of persons watching them could not identify the individuals being watched. The Conseil d'État, however, observed that regardless of this argumentation, the fact was that the drones' cameras were equipped with a zoom, and that the devices could fly low enough as to allow for obtaining a detailed picture of people, obliged to consider the data potentially collected as personal data (*ibid.*, paragraph 16).

7. POLICY RECOMMENDATIONS

There is broad consensus on the fact that AI systems in the field of law enforcement and criminal justice can have a **particularly serious impact on EU fundamental rights**. In this field, AI systems can touch the very heart of the relation between individuals and public authorities, bringing with them often the participation of private companies, potentially precipitating data flows across jurisdictional borders, and setting in motion a myriad of novel questions on how critical decisions are taken, how can they be explained, and how can they be contested. AI systems in this field may affect basic safeguards designed to guarantee individual freedom and to prevent the abuse of power. It is thus logical that when discussing possible scenarios for a possible regulation of AI, there is often a *'general consensus'* on the need to define at least some *'clear red lines on certain extremely high-risk scenarios'* which would have a significant impact on individuals and society.¹⁹²

In her 2019 political statement, Ursula Von der Leyen proclaimed, referring to the plans to move towards *'A Europe fit for the Digital Age'*, that *'(a)s we increase investment in disruptive research and breakthrough innovation, we must accept that failure will be part of our path'* (Von der Leyen, 2019, 13). In 2017, the European Commission had somehow similarly stated that *'(e)xperimentation is an important part of the exploration of emerging issues in the data economy'* (COM(2017) 9 final, 17). Although these approaches might be of some value from an economic perspective in certain contexts, accepting failure and embracing experimentation cannot be the most appropriate formula when fundamental rights are manifestly at stake.

This study has shown that the progressive deployment of AI in the field of law enforcement and criminal justice, although still embryonic to some extent, has already sparked calls for caution, as well as instances of mistrust and resistance. The study has also shown that although there have been already many EU-level efforts directed at setting up a **trustworthy ecosystem** for the advent of AI, such efforts have not yet been fully informed by a detailed acknowledgement of the specificities of the legal framework in this field.

Taking into account the findings of the study, the recommendations below can be put forward.

7.1. Assess fully the relevance and gaps of the existing legal framework

As described, many discussions on the possible need for a legislative intervention in relation to AI are grounded on interpretations of the data protection safeguards that the GDPR provides to individuals. In general, but especially in relation to personal data processing in the field of law enforcement and criminal justice, it is not sufficient to circumscribe the pertinent assessment to a reading of the GDPR.

It is imperative, first, to consider also EU legal instruments, such as the LED and the EU-DPR. But it is also equally necessary to move beyond the analysis of EU legal instruments, notably towards an assessment of national legislation that either specifies the GDPR or implements the LED, as Member States might have adopted exemptions or derogations that modulate the safeguards eventually available to individuals when data about them are processed for law enforcement or criminal justice purposes. Furthermore, it is equally indispensable to take into account the **enforcement** of relevant provisions, and the experiences of different actors involved.¹⁹³ Finally, it remains to be seen whether

¹⁹² Reference to such a general consensus at a multi-stakeholder workshop on AI and facial recognition celebrated in Brussels in February 2020: Wiewiórowski, 2020.

¹⁹³ There might be significant gap between certain visions of the presumed strengths of the EU data protection legal framework in light of the advent of AI, on the one hand, and the reality as perceived by individuals whose fundamental rights are at stake. In this sense, for instance, although Article 22 of the GDPR is supposed to offer significant safeguards against automated decision-making (when it applies), the latest Eurobarometer study on data protection law revealed that a

any eventual weaknesses can be effectively tackled by means different from a legislative intervention, and whether perhaps strengthening consistency mechanisms might be a sufficient solution.¹⁹⁴

7.2. Adapt initiatives to the specificities of the field

Some general principles commonly regarded as key for a trustworthy ecosystem for AI might face specific challenges in the field of law enforcement and criminal justice. It is the case, for instance, of **transparency**. It is thus important to take into account such specificities since the start, rather than developing a general framework for trustworthy AI, the impact of which in the field of law enforcement and criminal justice would eventually be cancelled, or seriously undermined, by exemptions or derogations adopted in the name of law enforcement or criminal justice needs. This might require, for instance, thinking about specific governance mechanisms that would be efficient enough to compensate other possible limitations of transparency.

The Council of Europe's *Guidelines on Addressing the Human rights Impacts of Algorithmic Systems* explicitly refer to this tension, notably in reference to recommended human rights impact assessments.¹⁹⁵ The Guidelines suggest that such impact assessments '*should be conducted as openly as possible and with the active engagement of affected individuals and groups*', and when high-risk algorithmic systems are concerned '*the results of ongoing human rights impact assessments, identified techniques for risk mitigation, and relevant monitoring and review processes should be made publicly available, without prejudice to secrecy safeguarded by law*'.¹⁹⁶ More specifically, '*(w)hen secrecy rules need to be enforced, any confidential information should be provided in a separate appendix to the assessment report*', which should be **accessible to relevant supervisory authorities**.¹⁹⁷

It is also important to note that any possible limitations of transparency towards data subjects make even more important compliance with other legal principles, such as **legal certainty** and **predictability**.¹⁹⁸

7.3. Consider the need for special rules for certain practices

The many issues triggered by facial recognition, and most notably **live facial recognition in public spaces**, demand urgent attention and possibly a specific response. This should not detract the attention, however, from the possible need to regulate also beyond such specific practices. In this sense, calls for a regulatory framework for algorithmic decision-making for law enforcement purposes appear as very pertinent in light of the findings of this study.¹⁹⁹

majority of individuals in the EU had never heard about their right to have a say when they are subject to automated decisions (Kantar, 2019, 23). The need for more clarity on the interpretation of Article 22 of the GDPR has been already highlighted by many (cf. for instance, Datenethikkommission, 2019, 28, 192).

¹⁹⁴ Such an assessment shall nevertheless also take into account that consistency beyond the GDPR is less formally developed, but also that even under the GDPR consistency mechanisms appear not to be fully efficient at the moment (cf. for instance, noting lack of efficiency in regards to identifying consistent criteria for data protection assessments: Belgian Data Protection Authority (DPA), 2020, 3).

¹⁹⁵ Recommendation CM/Rec(2020)1, Appendix, paragraph B.5.3.

¹⁹⁶ Idem.

¹⁹⁷ Idem.

¹⁹⁸ As stated by Recital (26) of Regulation (EU) 2016/794 in relation to law enforcement cooperation, that is an area '*where data subjects are usually unaware when their personal data are being collected and processed and where the use of personal data may have a very significant impact on the lives and freedoms of individuals*'.

¹⁹⁹ In relation to the UK: Babuta, Oswald and Rinik, 2018, vii.

7.4. Clarify the role of ethics

Ethical frameworks and guidance might be helpful to help delivering a trustworthy ecosystem for AI, but currently there is **much confusion** as to what should be the relation between such ethical frameworks or guidance and fundamental rights safeguards. Such confusion, if further sustained, might be in the end detrimental to the protection of fundamental rights, to the extent that it can divert attention from the necessity of safeguarding certain legal obligations. Most notably, ethical frameworks might not deliver the necessary safeguards in terms of access to remedies.

7.5. Reinforce trust in EU-funded AI efforts

Governments, it has been pointed out, should be *'the first'* to use AI in a manner that safeguards and promotes data protection and fundamental rights in general (Mantelero, 2019, 3); they must uphold particularly high standards because they have a direct obligation to do so, but also because state activity is, in general, expected to set an example for the whole of society (Datenethikkommission, 2019, 212). These reflections can be extended to cover the special responsibility that public authorities – including EU institutions, agencies and bodies – have towards the support of AI research and solutions with public funds.

Transparency, oversight and accountability requirements must be particularly high when EU funds are used to support AI research that can have a serious impact on EU fundamental rights, such as **security-related AI research**. Ethics reviews and assessments must be based on up-to-date, appropriate, pertinent guidance, and be accompanied by clear **transparency requirements**, in order to guarantee that the research funded is appropriately putting forward trustworthy solutions and is duly open to public scrutiny, especially insofar as compliance with legal and 'ethical' standards are concerned.

As research in this area is expected to be increasingly significant in the upcoming years, delivering specific guidance to improve trust in EU-funded research is crucial, and thus progress in terms of openness is essential. The EDPS could play a major role by stepping up its role in the overview of certain data processing operations funded with the support of EU funds. The EDPB could provide better, more detailed guidance, especially for research efforts that cross national borders.²⁰⁰

7.6. Screen developments related to the Covid-19 outbreak

In response to the Covid-19 outbreak, a variety of data-based and technology-driven solutions have been embraced, not necessarily immediately accompanied by the pertinent technical and legal safeguards. This can potentially lead to a situation of increased vulnerability in front of cybersecurity attacks, including threats targeting sensitive data. It can also lead to situations of increased vulnerability due to risks of misuse of power. Even if many of the endorsed initiatives do not explicitly have a law enforcement dimension, they nevertheless enable the processing of data that might, eventually, be made available for law enforcement purposes, thus creating special risks for individuals' fundamental rights. Moreover, due to the virus outbreak and the subsequent management of the crisis, fundamental rights have sometimes been restricted in serious ways. This obliges to be particularly vigilant, in order to prevent weakened fundamental rights from being irreparably damaged by the crisis, but also by its responses.

²⁰⁰ Also stressing the opportunity of concrete guidance on AI-related research, in particular in cross-border projects: Datenethikkommission, 2019, 20.

REFERENCES

- Access Now Europe, [The European human rights agenda in the digital age](#), November 2019.
- Advisory Committee on Equal Opportunities for Women and Men, [Opinion on Artificial Intelligence – opportunities and challenges for gender equality](#), 18 March 2020.
- AI 4 Belgium Coalition, *AI 4 Belgium Report*, 2019.
- AI Now Institute, [AI Now Report 2018](#), December 2018.
- AI Task Force, [Artificial Intelligence in Support of Defence](#), Ministère des Armées (France), September 2019.
- Alfter, B., 'Denmark', in M. Spielkamp, *Automating Society: Taking Stock of Automated Decision-Making in the EU*, AlgorithmWatch and BertelsmannStiftung, January 2019, pp. 46-54.
- Amnesty International, *Trapped in the Matrix: Secrecy, stigma, and bias in the Met's Gangs Database*, May 2018.
- Arroyo, V. and D. Leufer (2020), '[Facial recognition on trial: emotion and gender "detection" under scrutiny in a court case in Brazil](#)', Access Now, 29 June 2020.
- Article 29 Working Party, [Guidelines on Data Protection Impact Assessment \(DPIA\) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679](#), Adopted on 4 April 2017 as last Revised and Adopted on 4 October 2017, WP 248 rev.01.
---, [Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679](#), Adopted on 3 October 2017, as last Revised and Adopted on 6 February 2018, WP251rev.01.
- Association for Computing Machinery (ACM) US Technology Policy Committee (USTPC), [Statement on Principles and Prerequisites for the Development, Evaluation and Use Of Unbiased Facial Recognition Technologies](#), 30 June 2020.
- Austrian Delegation to the Council, [Note to Working Party on Information Exchange and Data Protection \(DAPIX\) on Next generation Prüm \(Prüm.ng\) - Reports from focus groups / Report on face recognition](#), 13356/19, Brussels, 30 October 2019.
- Autoriteit Persoonsgegevens, [On the anonymity of aggregated telco location data](#), April 2020.
- Babuta, A. and M. Oswald, [Data Analytics and Algorithms in Policing in England and Wales Towards A New Policy Framework. Occasional Paper](#), Royal United Services Institute for Defence and Security Studies, February 2020.
- Babuta, A., M. Oswald & C. Rinik, [Machine Learning Algorithms and Police Decision-Making Legal, Ethical and Regulatory Challenges](#), Whitehall Report 3-18, Royal United Services Institute for Defence and Security Studies, September 2018.
- Barik, S., '[Sweden permits police-use of facial recognition technology](#)', *Medianama*, 29 October 2019.
- Belfiore, R., 'The Protection of Personal Data Processed Within the Framework of Police and Judicial Cooperation in Criminal Matters', in S. Ruggeri (ed.), *Transnational Inquiries and the Protection of Fundamental Rights in Criminal Proceedings*, Springer, 2013, pp. 355-370.

- Belgian Data Protection Authority, [Evaluation of the GDPR under Article 97 – Questions To Data Protection Authorities/European Data Protection Board: Answers From The Belgian Supervisory Authority](#), 2020.
 ---, *Objet: Demande d'avis concernant un avant-projet d'arrêté royal n° XXX portant création d'une banque de données auprès de Sciensano dans le cadre de la lutte contre la propagation du coronavirus COVID-19 (CO-A-2020-042), Avis n° 36/2020 du 29 avril 2020.*
- Blatt, J. J., '[Some Observations on the Clearview AI Facial Recognition System – From Someone Who Has Actually Used It ...](#)', 6 May 2020.
- Boddington, P., *Towards a Code of Ethics for Artificial Intelligence*, Springer, 2017.
- Boucher, P., [Why Artificial Intelligence matters \(Briefing\)](#), European Parliamentary Research Service (EPRS), Scientific Foresight Unit (STOA) PE 634.421, March 2019.
- boyd, d., '[Untangling research and practice: What Facebook's "emotional contagion" study teaches us](#)', *Research Ethics*, 12(1), 2016, pp. 4–13.
- Brkan, M., 'Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond', *International Journal of Law and Information Technology*, 27:2, 2019.
- Brundage, M., et al., [The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation](#), February 2018.
- Buolamwini, J. and T. Gebru, '[Gender shades: Intersectional accuracy disparities in commercial gender classification](#)', *Conference on fairness, accountability and transparency*, Proceedings of Machine Learning Research, 81, 2018.
- Campbell, Z. and C. Jones, '[Leaked Reports Show EU Police Are Planning a Pan-European Network of Facial Recognition Databases](#)', *The Intercept*, 21 February 2020.
- Carrera, S. and M. Stefan, [Access to Electronic Data for Criminal Investigations Purposes in the EU](#), No. 2020-01, Center for European Policy Studies (CEPS), February 2020.
- Castro, D., M. McLaughlin & E. Chivot, '[Who Is Winning the AI Race: China, the EU or the United States?](#)', *Centre for Data Innovation*, 19 August 2019.
- Chambre des Députés du Grand-Duché de Luxembourg, [Covid-19 : vers un traçage numérique ?](#), 28 April 2020.
- Chelioudakis, E., '[Greece: Technology-led policing awakens](#)', *About Intel: European Voices on Surveillance*, 29 June 2020.
- Commissioner for Human Rights of the Council of Europe, [Unboxing Artificial Intelligence: 10 steps to protect Human Rights – Recommendation](#), May 2019.
- Committee of Experts on Internet Intermediaries (MSI-NET), *Study on The Human Rights Dimensions of Automated Data Processing Techniques (In Particular Algorithms) and Possible Regulatory Implications*, (Rapporteur: Benjamin Wagner), DGI(2017)12, March 2018.
- Committee of Experts on Quality Journalism in the Digital Age (MSI-JOQ), [Draft Recommendation CM/Rec\(20XX\)XX of the Committee of Ministers to member States on promoting a favourable environment for quality journalism in the digital age: 7th draft as of 26 September 2019](#), Council of Europe, MSI-JOQ(2018)08rev6, 2019.

- Committee on Civil Liberties, Justice and Home Affairs (LIBE) of the European Parliament, [*Draft Report on Artificial Intelligence in criminal law and its use by the police and judicial authorities in criminal matters*](#), Rapporteur: Tudor Ciuhodaru, LIBE, 2020/2016(INI), 8 June 2020.
- Conseil d'État, [*Surveillance par drones \(décision\)*](#), 18 mai 2020.
- Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data ('Convention 108'), [*Practical Guide on the Use of Personal Data in the Police Sector*](#), Strasbourg, 15 February 2018, T-PD(2018)01.
- Cook, S., [*Letter to Attorney General Barr on The use of the PATTERN risk assessment in prioritizing release in response to the COVID-19 pandemic*](#), 3 April 2020.
- Cope, S. and Hussain, S., 'EFF to Court: Social Media Users Have Privacy and Free Speech Interests in Their Public Information', *Electronic Frontier Foundation (EFF)*, 30 June 2020.
- Coudert, F., 'The Europol Regulation and Purpose Limitation: From the 'Silo-Based Approach' to... What Exactly?', *European Data Protection Law (EDPL)*, Vol. 3, Number 3, 2017, pp. 313-324.
- Council of the EU, *Note from Permanent Representatives Committee (Part 1) to Council on Artificial intelligence - b) Conclusions on the coordinated plan on artificial intelligence (Adoption)*, 6177/19, Brussels, 11 February 2019.
- ---, *Note from the Presidency to the Standing Committee on Operational Cooperation on Internal Security (COSI) on The future direction of EU internal security: new technologies and internal security - Preparation of the Council debate*, 12224/19, COSI 184 ENFOPOL 400 CYBER 257 JAI 949, Brussels, 18 September 2019.
- Cox, J., [*Surveillance Company Says It's Deploying 'Coronavirus-Detecting' Cameras in US*](#), *Motherboard: Tech by The Vice*, 17 March 2020,
- Cox, J. and J. Koebler, [*Surveillance Firm Banjo Used a Secret Company and Fake Apps to Scrape Social Media*](#), 9 March 2020, *Motherboard: Tech by The Vice*.
- Data Ethics Commission of the German Federal Government (Datenethikkommission), *Opinion* (English version), Berlin, December 2019.
- Datainspektionen, [*Datainspektionen inleder tillsyn med anledning av Clearview AI*](#), 6 March 2020.
- Datakalab, [*Politique de protection des données personnelles dans le cadre de la mise en œuvre du dispositif de détection de masques de protection respiratoire*](#), April 2020.
- Davies, B., M. Innes, and A. Dawson, *An evaluation of South Wales Police's use of Automated Facial Recognition*, Police Science Institute Crime and Security Research Institute, Cardiff University Institute, September 2018.
- De Brouwer, S., [*European Parliament shows \(again\) its stance on Artificial Intelligence*](#), *Access Now Blog*, 14 February 2020.
- De Capitani, E., [*Complaint and application for a temporary injunction \(Geulen & Klinger\)*](#), 13 May 2019.
- Dodd, V., [*MET removes hundreds from gangs matrix after breaking data laws*](#), *The Guardian*, 15 February 2020.
- Dotta, G., [*ENAC: autorizzati i droni per il monitoraggio*](#), *Punto Informatico*, 24 March 2020,
- Dumbrava, C., [*Tracking mobile devices to fight coronavirus*](#), *European Parliamentary Research Service Blog*, 21 April 2020.

- DutchNews, '[Privacy watchdog slams coronavirus phone data mapping plan](#)', *DutchNews.nl*, 3 July 2020.
- eHealth Network, [Common EU Toolbox for Member States: Mobile applications to support contact tracing in the EU's fight against COVID-19 \(Version 1.0\)](#), 15 April 2020.
 - , [Interoperability guidelines for approved contact tracing mobile applications in the EU](#), 13 May 2020.
 - , [Interoperability specifications for cross-border transmission chains between approved apps - Basic interoperability elements between COVID+ Keys driven solutions](#), V1.0, 12 June 2020.
 - , [To the EU Member States and the European Commission on Interoperability specifications for cross-border transmission chains between approved apps - Detailed interoperability elements between COVID+ Keys driven solutions](#), V1.0, 16 June 2020.
- Ethics Advisory Group (EAG), [Towards a digital ethics](#), European Data Protection Supervisor (EDPS), 2018.
- European Commission (EC), [Community policy on data processing: Communication of the Commission to the Council](#), SEC (73) 4300 final, 21 November 1973.
 - , [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Towards a thriving data-driven economy](#), COM(2014) 442 final, 2 July 2014.
 - , [A Digital Single Market Strategy for Europe](#), COM(2015) 192 final, Brussels, 6 May 2015.
 - , [Communication from the Commission to the European Parliament and the Council: Stronger Smarter Information Systems for Border and Security](#), COM(2016) 205 final, Brussels, 6 April 2016.
 - , [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Building a European Data Economy](#), COM(2017) 9 final, Brussels, 10 January 2017.
 - , [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Commission Work Programme 2018: An agenda for a more united, stronger and more democratic Europe](#), COM(2017) 650 final, Strasbourg, 24 October 2017.
 - , [Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe](#), COM(2018) 237 final, Brussels, 25 April 2018.
 - , [Proposal for a Regulation of the European Parliament and of the Council amending Regulation \(EC\) No 767/2008, Regulation \(EC\) No 810/2009, Regulation \(EU\) 2017/2226, Regulation \(EU\) 2016/399, Regulation XX/2018 \[Interoperability Regulation\] and Decision 2004/512/EC and repealing Council Decision 2008/633/JHA](#), COM/2018/302 final, Brussels, 16 May 2018.
 - , [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Commission Work Programme 2019: Delivering what we promised and preparing for the future](#), COM(2018) 800 final, Strasbourg, 23 October 2018.
 - , [Ethics and data protection](#), 14 November 2018 (EC, 2018).

- , [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Coordinated Plan on Artificial Intelligence Brussels](#), COM(2018) 795 final, 7 December 2018.
- , [Horizon 2020 Programme Guidance: How to complete your ethics self-assessment, Directorate-General for Research & Innovation](#), Version 6.1, 4 February 2019.
- , [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Building Trust in Human-Centric Artificial Intelligence](#), COM(2019) 168 final, Brussels, 8 April 2019.
- , [Communication from the Commission to the European Parliament and the Council - Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union](#), COM(2019) 250 final, Brussels, 29 May 2019.
- , [A Community of Users on Secure, Safe and Resilient Societies: Mapping Horizon H2020 and EU-Funded Capacity-Building Projects under 2016-2018 Programmes](#), October 2019.
- , [H2020 Programme Guidance: Guidelines for the classification of information in research projects](#), Directorate-General for Migration and Home Affairs, Version 2.2, 7 January 2020.
- , [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Commission Work Programme 2020: A Union that strives for more](#), COM(2020) 37 final, Brussels, 29 January 2020.
- , [Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee: Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics](#), COM(2020) 64 final, Brussels, 19 February 2020.
- , [White Paper On Artificial Intelligence -A European approach to excellence and trust](#), COM(2020) 65 final, Brussels, 19 February 2020.
- , [Communication to the European Parliament, to the Council, the European Economic and Social Committee and the Committee of the Regions. A European data strategy](#), COM(2020) 66 final, Brussels, 19 February 2020.
- , [Communication to the European Parliament, to the Council, the European Economic and Social Committee and the Committee of the Regions. Shaping Europe's digital future](#), COM(2020) 67 final, Brussels, 19 February 2020.
- , [Daily News](#), 7 April 2020.
- , [Commission Recommendation on a common Union toolbox for the use of technology and data to combat and exit from the COVID-19 crisis, in particular concerning mobile applications and the use of anonymised mobility data](#), C(2020) 2296 final, Brussels, 8 April 2020.
- , [Guidance on Apps supporting the fight against COVID 19 pandemic in relation to data protection](#), C/2020/2523, 16 April 2020.
- , [Daily News](#), 16 June 2020.
- , [Communication to the European Parliament and the Council, Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition - two years of application of the General Data Protection Regulation](#), COM(2020) 264 final, Brussels, 24 June 2020.

- , [*Communication from the Commission to the European Parliament and the Council: Way forward on aligning the former third pillar acquis with data protection rules*](#), Brussels, COM(2020) 262 final, 24 June 2020.
- , [*Mobile applications to support contact tracing in the EU's fight against COVID-19: Progress reporting June 2020*](#), July 2020.
- European Commission for the Efficiency of Justice (CEPEJ), *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*, adopted by the CEPEJ during its 31st Plenary meeting (Strasbourg, 3-4 December 2018), CEPEJ(2018)14.
- European Data Protection Board (EDPB), [*Facial recognition in school renders Sweden's first GDPR fine*](#), 22 August 2019.
- , [*Guidelines 03/2020 on the processing of data concerning health for the purpose of scientific research in the context of the COVID-19 outbreak*](#), 21 April 2020.
- , [*Draft Agenda 26th session of plenary meeting*](#), 8 May 2020.
- , *Statement on the data protection impact of the interoperability of contact tracing apps*, adopted on 16 June 2020.
- European Data Protection Supervisor (EDPS), [*Annual Report 2019*](#), 2019.
- , [*Opinion 4/2020 on the European Commission's White Paper on Artificial Intelligence – A European approach to excellence and trust*](#), 29 June 2020.
- , [*Outcome of own-initiative investigation into EU institutions' use of Microsoft products and services*](#), 2 July 2020.
- European Migration Network (EMN), [*Annual Report on Migration and Asylum 2018*](#), May 2019.
- European Parliament (EP), [*Resolution on the US NSA surveillance programme, surveillance bodies in various Member States and their impact on EU citizens' fundamental rights and on transatlantic cooperation in Justice and Home Affairs*](#), (2013/2188(INI)), 12 March 2014.
- , [*Resolution on fundamental rights implications of big data: privacy, data protection, non-discrimination, security and law-enforcement \(2016/2225\(INI\)\)*](#), P8_TA(2017)0076, 14 March 2017.
- , [*Resolution on Automated decision-making processes: Ensuring consumer protection, and free movement of goods and services*](#), P9_TA(2020)0032, 12 February 2020.
- , [*Resolution of on EU coordinated action to combat the COVID-19 pandemic and its consequences*](#), P9_TA(2020)0054, 17 April 2020.
- , [*Decision on discharge in respect of the implementation of the general budget of the European Union for the financial year 2018, Section I – European Parliament \(2019/2056\(DEC\)\)*](#), P9_TA(2020)0084, 13 May 2020.
- European Union Agency for the Operational Management of Large-Scale IT Systems in the Area of Freedom, Security and Justice (eu-LISA), [*Testing the borders of the future: Smart Borders Pilot: The results in brief*](#), 2015 (2015a).
- , [*Smart Borders Pilot Project: Technical Report Annexes: Volume 2*](#), (2015b).
- European Union Agency for Fundamental Rights (EU FRA), [*Preventing unlawful profiling today and in the future: A guide*](#), Handbook, 2018 (2018a).
- , [*#BigData: Discrimination in data-supported decision making*](#), FRA Focus, 2018 (2018b).

---, [Facial recognition technology: fundamental rights considerations in the context of law enforcement](#), FRA Focus, November 2019.

---, [AI policy initiatives \(2016-2020\)](#), April 2020.

- Expert Committee on Human Rights Dimensions of Automated Data Processing and Different Forms of Artificial Intelligence (MSI-AUT), *Responsibility and AI: A study of the implications of advanced digital technologies (including AI systems) for the concept of responsibility within a human rights framework* (Rapporteur: Karen Yeung), DGI(2019)05, Council of Europe, September 2019.
- Ferretti, L., et al., '[Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing](#)', *Science*, Vol. 368, Issue 6491, 8 May 2020.
- Fight for the Future, '[Backlash forces UCLA to abandon plans for facial recognition surveillance on campus](#)', *Medium*, 19 February 2020.
- Floridi, L., '[Soft Ethics and the Governance of the Digital](#)', *Philosophy & Technology*, Volume 31, 2018, pp. 1-8.
- Fondazione Giacomo Brodolini (FGB), '[Fundamental rights review of EU data collection instruments and programmes: Final report](#)', 2019.
- Franko Aas, K., *Sentencing in the age of information: from Faust to Macintosh*, Routledge, New York, 2005.
- Fürsteneau, M., '[Germany's facial recognition pilot program divides public](#)', *DW*, 24 August 2017.
- Garvie, C., A. Bedoya, and J. Frankle, '[The Perpetual Line-Up. Unregulated Police Face Recognition in America](#)', *Georgetown Law Center on Privacy & Technology*, 18 October 2016.
- Gates, K. A., *Our biometric future: Facial recognition technology and the culture of surveillance*, New York University Press, 2011.
- General Secretariat of the Council, '[Note to Delegations on Council Conclusions on the implementation of the "Prüm Decisions" ten years after their adoption](#)', 11227/18, Brussels, 18 July 2018.
- Gershgorn, D., '[Facial Recognition Companies See the Coronavirus as a Business Opportunity](#)', *One Zero*, 19 March 2020.
- Gießler, S. and L. Haas, '[Ethics between business lingo and politics: Why bother?](#)', *AI Ethics Guidelines Global Inventory*, April 2020.
- González Fuster, G., '[Feedback to the Consultation on the White Paper on Artificial Intelligence \(AI\) focusing on AI & gender: An EU law perspective](#)', 14 June 2020.
- Grace, J., 'Algorithmic impropriety in UK policing?', *Journal of Information Rights, Policy and Practice*, 3(1), 2019.
- Gräf, E., 'When Automated Profiling Threatens Our Freedom: A Neo-Republican Perspective', *European Data Protection Law (EDPL)*, Vol. 3, Number 4, 2017, pp. 441-451.
- Grind, K., R. McMillan and A. W. Mathews, '[To Track Virus, Governments Weigh Surveillance Tools That Push Privacy Limits](#)', *Wall Street Journal*, 17 March 2020.
- Habersetzer, N., '[Moscow Silently Expands Surveillance of Citizens](#)', *Human Rights Watch*, 25 March 2020.

- Hamon, R., H. Junklewitz & I. Sánchez, *Robustness and Explainability of Artificial Intelligence: From technical to policy solutions*, JRC Technical Report, Joint Research Centre (JRC), EUR 30040 EN, 2020.
- Harcourt, B. E., *Against prediction: Profiling, policing, and punishing in an actuarial age*, University of Chicago Press, 2008.
- Hardyns, W., and A. Rummens, 'Predictive policing as a new tool for law enforcement? Recent developments and challenges', *European Journal on Criminal Policy and Research*, 24(3), 2017, pp. 201-218.
- Hennessy, M., '[Explainer: What happened with the Graham Dwyer data retention case in the Supreme Court?](#)', *The Journal*, 24 February 2020.
- High-Level Expert Group on Artificial Intelligence (AI HLEG), [Ethics Guidelines for Trustworthy Artificial Intelligence \(AI\)](#), 8 April 2019 (2019a).
- ---, [Policy and Investment Recommendations for Trustworthy AI](#), 26 June 2019 (2019b).
- High-Level Group on Fake News and Online Disinformation, [A multi-dimensional approach to disinformation](#), European Commission, March 2018.
- Hill, K., '[The Secretive Company That Might End Privacy as We Know It](#)', *The New York Times*, 18 January 2020.
- Information Commissioner's Office (ICO), [The Office of the Australian Information Commissioner and the UK's Information Commissioner's Office open joint investigation into Clearview AI Inc.](#), Statement, 9 July 2020.
- Innovation Centre of the International Criminal Police Organization (Interpol) and United Nations Interregional Crime and Justice Research Institute (UNICRI), *Artificial Intelligence and Robotics for Law Enforcement*, 2019.
- Interpol, [INTERPOL launches awareness campaign on COVID-19 cyberthreats](#), 6 May 2020.
- IT-Pol, [Danish DPA approves Automated Facial Recognition](#), European Digital Rights Initiative (EDRI), 19 June 2019.
- Jain, R., '[Baidu's Face Detection AI Will Help China Identify People Without Masks](#)', *International Business Times*, 17 February 2020.
- Jelinek, A., *EDPB response to MEPs Sophie in 't Veld, Moritz Körner, Michal Šimečka, Fabienne Keller, Jan-Christoph Oetjen, Anna Donáth, Maite Pagazaurtundúa, Olivier Chastel, concerning the facial recognition app developed by Clearview AI*, 10 June 2020.
- Jobin, A., M. Ienca, and E. Vayena, '[The global landscape of AI ethics guidelines](#)', *Nature Machine Intelligence*, 2019, 1, pp. 389–399.
- Juncker, J. C., [A New Start for Europe: My Agenda for Jobs, Growth, Fairness and Democratic Change](#), Strasbourg, 15 July 2014.
- Kantar, '[Special Eurobarometer 487a – March 2019: The General Data Protection Regulation](#)', March 2019.
- Kayser-Bril, N., '[Unchecked use of computer vision by police carries high risks of discrimination](#)', *AlgorithmWatch*, 28 April 2020.
- Kim, N., '[More Scary Than Coronavirus](#)': South Korea's Health Alerts Expose Private Lives, *The Guardian*, 6 March 2020.

- , [‘Anti-gay backlash feared in South Korea after coronavirus media reports’](#), *The Guardian*, 8 May 2020.
- Koebler, J., E. Maiberg, and J. Cox, [‘This Small Company Is Turning Utah Into a Surveillance Panopticon’](#), *Motherboard: Tech By Vice*, 4 March 2020.
 - Korff, D., [Comments on Selected Topics in the Draft EU Data Protection Regulation](#), London/Cambridge, 2012.
 - Labarthe, S. and A. Velásquez, [CoronApp, Medellín me Cuida y Cali Valle Corona al laboratorio -O cómo se hackea CoronApp sin siquiera intentarlo-](#), 17 April 2020 (2020a).
 - ---, [Covid Apps in Colombia, Karisma’s digital security and privacy evaluation](#), 28 April 2020 (2020b).
 - Lapinski, I., [‘When algorithms decide your rights’](#), *Digital Freedom Fund*, December 2018.
 - La Quadrature du Net, [La reconnaissance faciale des manifestant-e-s est déjà autorisée](#), 18 November 2019.
 - Lee, T. B., [‘Detroit police chief cops to 96-percent facial recognition error rate’](#), *Ars Technica*, 30 June 2020.
 - Leslie, D., *Understanding Artificial Intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*, The Alan Turing Institute, 2019.
 - Leufer & Jansen, [‘The EU is funding dystopian Artificial Intelligence projects’](#), Euractiv, 22 January 2020.
 - Light, F., [‘Russia Is Building One of the World’s Largest Facial Recognition Networks’](#), *The Moscow Times*, 12 November 2019.
 - Lynch, S., [‘The Challenges of Facial Recognition Technologies’](#), *Human-Centered Artificial Intelligence - Stanford University*, 23 April 2020.
 - Mac, R., C. Haskins & L. McDonald, [‘Clearview’s Facial Recognition App Has Been Used By The Justice Department, ICE, Macy’s, Walmart, And The NBA’](#), *Buzzfeed News*, 27 February 2020.
 - Mantelero, A., [Report on Artificial Intelligence - Artificial Intelligence and Data Protection: Challenges and Possible Remedies](#), Consultative Committee of the Convention for the Protection of Individuals with Regard to Automatic Processing of Personal Data (Convention 108), Strasbourg, 25 January 2019, T-PD(2018)09Rev.
 - Markey, E. J., [Letter from Senator Edward J. Markey to Hoan Ton-That, Founder & Chief Executive Officer of Clearview AI](#), 30 April 2020.
 - Mascini, L., [‘EU expert on European response to virus: ‘Telecom data for tracking corona can be made anonymous’](#), *Innovation Origins*, 11 April 2020.
 - Massé, E., *Two Years under the EU GDPR: An Implementation Progress Report*, Access Now, May 2020.
 - Mayor’s Office for Policing and Crime (MOPAC), [Equality Impact Assessment, Gang Violence Matrix, Version 4](#), 2018.
 - McCarthy, O. J., [‘AI & Global Governance: Turning the Tide on Crime with Predictive Policing’](#), *United Nations University, Centre for Policy Research*, 26 February 2019.

- McNamara, Daniel, et al., 'Trade-offs in algorithmic risk assessment: An Australian domestic violence case study' in Daly, A., M. Mann, and S. K. Devitt, *Good Data*, Institute of Network Cultures, Theory on Demand Vol. 29, 2019, pp. 96-116.
- Meaker, M., '[Marseille's fight against AI surveillance](#)', Coda, 26 March 2020.
- Ministry of Finance and Ministry of Industry, Business and Financial Affairs of the Danish Government, *National Strategy for Artificial Intelligence*, March 2019.
- Molnar, P. and L. Gill, [Bots at the Gate: A Human Rights Analysis of Automated Decision-Making in Canada's Immigration and Refugee System](#), International Human Rights Program (Faculty of Law, University of Toronto) and the Citizen Lab (Munk School of Global Affairs and Public Policy, University of Toronto), 2018.
- Monroy, M., '[Prüm Decision: European criminal police offices agree on face recognition system](#)', *Security Architectures and Police Collaboration in the EU*, 11 March 2020.
- Mozur, P., '[One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority](#)', *The New York Times*, 14 April 2019.
- Oswald, M., and J. Grace, 'Intelligence, policing and the use of algorithmic analysis: a freedom of information-based study', *Journal of Information Rights, Policy and Practice*, 1.1, 2016.
- Parliamentary Assembly of the Council of Europe, [Recommendation 2102 \(2017\) on Technological Convergence, Artificial Intelligence and Human Rights](#), adopted on 28 April 2017.
- Peeters, B., '[Facial recognition at Brussels Airport: face down in the mud](#)', *CITIP Blog*, 17 March 2020.
- Penner, K., [Report – Automating Society: Europe](#), 29 January 2019.
- Perry, L. W., et al., [Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations](#), Rand Corporation, 2013.
- Peters, J., '[IBM will no longer offer, develop, or research facial recognition technology](#)', *The Verge*, 8 June 2020.
- Porter, J., '[Facebook and LinkedIn are latest to demand Clearview stop scraping images for facial recognition tech](#)', *The Verge*, 6 February 2020.
- Presidency of the Council, [Note to Working Party on Information Exchange and Data Protection \(DAPIX\) on Next generation Prüm \(Prüm.ng\) - Planning of focus groups, 5556/19](#), Brussels, 15 February 2019.
- ---, [Note to Working Party on Information Exchange and Data Protection \(DAPIX\) on Monitoring the Implementation of Directive \(EU\) 2016/681 on the use of passenger name record \(PNR\) data – State of play and way forward](#), 6300/19, Brussels, 15 February 2019.
- Privacy International, '[MONITORYOU: the MilliONs being spent by the eu on developing surveillance tech to target YOU](#)', 20 January 2020 (2020a).
- ---, '[One Ring to watch them all](#)', 25 January 2020 (2020b).
- ---, '[The SyRI case: a landmark ruling for benefits claimants around the world](#)', 24 February 2020 (2020c).
- ---, '[This UK Government-Funded AI Programme Wants to Make 'Face Recognition Ubiquitous' \(But Sure, We're Probably Being Paranoid About Face Surveillance\)](#)', 3 March 2020 (2020d).

- Renda, A., [‘Will privacy be one of the victims of COVID-19?’](#), Centre for European Policy Studies (CEPS), 23 March 2020.
- Reuters, [‘Germany: Refugees sue the government for invasion of privacy’](#), DW, 5 May 2020.
- Richardson, R., J. Schultz and K. Crawford, [‘Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice’](#), 94 N.Y.U. L. Rev. Online 192, February 2019, pp. 192-233.
- Richardson, R., J. M. Schultz, & V. M. Southerland, [Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems](#), AI Now Institute, September 2019.
- Robinson, D., and L. Koepke, [‘Stuck in a pattern: Early evidence on ‘predictive policing’](#), UpTurn, 2016.
- Ronsin, X. and V. Lamos, *In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data*, Appendix I accompanying the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, adopted by the CEPEJ during its 31st Plenary meeting (Strasbourg, 3-4 December 2018), CEPEJ(2018)14, 2018.
- Rubio, I., [‘Las claves de la polémica por el uso de reconocimiento facial en los supermercados de Mercadona’](#), El País, 7 July 2020.
- Sartor, G. and F. Lagioia (2020), [The impact of the General Data Protection Regulation \(GDPR\) on Artificial Intelligence](#), European Parliamentary Research Service (EPRS), Scientific Foresight Unit (STOA) PE 641.530 – June 2020.
- Samoili, S. et al., [AI Watch - Defining Artificial Intelligence: Towards an operational definition and taxonomy of artificial intelligence](#), JRC Technical Reports, European Union, 2020.
- Secretariat of the Council of Europe, *Submission to the Consultation on the “White Paper on Artificial Intelligence – a European approach to excellence and trust”*, 19 June 2020.
- Scott, M., [‘Europe is fighting tech battle with one hand tied behind its back’](#), Politico, 24 February 2020.
- Singer, N. and M. Isaac, [‘Facebook to Pay \\$550 Million to Settle Facial Recognition Suit’](#), The New York Times, 29 January 2020.
- Temme, M., ‘Algorithms and Transparency in View of the of GDPR’, *European Data Protection Law (EDPL)*, Vol. 3, Number 4, 2017, pp. 473-485.
- Thomsen, M., [‘Federal government and three states are in talks with controversial facial recognition company Clearview AI to track coronavirus patients and see who they have been in contact with’](#), Daily Mail, 2 May 2020.
- Turß, D., [European Court of Justice to decide on mass processing of passenger data](#), 22 January 2020.
- Unione Camere Penali Italiane, [Processo penale da remoto e protezione dei dati personali. L'Unione scrive al Garante](#), 14 April 2020.
- United Nations (UN) Independent Human Rights Experts Press Service, [Landmark ruling by Dutch court stops government attempts to spy on the poor – UN expert](#), Press release, 5 February 2020.
- United Nations (UN) Special Rapporteur on Extreme Poverty and Human Rights, [Report of the Special Rapporteur on extreme poverty and human rights](#), A/74/493, 2019.

- United States (US) Department of Defense, [*Summary of the 2018 National Defense Strategy of The United States of America: Sharpening the American Military's Competitive Edge*](#), Washington, D.C., 2018.
- , [*DOD Adopts Ethical Principles for Artificial Intelligence*](#), Press Release, 24 February 2020.
- United States (US) Department of Justice, *Predictive Analytics in Law Enforcement: A Report by the Department of Justice*, EPIC-16-06-15-DOJ-FOIA-20200319-Settlement-Production-pt1, November 2014.
- Vaas, L., '[Boston bans government use of facial recognition](#)', *Naked Security* by Sophos, 6 July 2020.
- Van Brakel, R. (2019), 'Belgium', in M. Spielkamp, *Automating Society: Taking Stock of Automated Decision-Making in the EU*, AlgorithmWatch and BertelsmannStiftung, January 2019, pp. 39-44.
- Van Hoboken, J., et al., [*The legal framework on the dissemination of disinformation through Internet services and the regulation of political advertising: Final report*](#), Institute for Information Law (IViR), University of Amsterdam, 2019.
- Van Roy, V., [*AI Watch - National strategies on Artificial Intelligence: A European perspective in 2019*](#), EUR 30102 EN, Publications Office of the European Union, Luxembourg, 2020.
- Vincent, J., [Spot the robot is reminding parkgoers in Singapore to keep their distance from one another](#), *The Verge*, 8 May 2020.
- Von der Leyen, U., '[A Union that strives for more: My agenda for Europe – Political guidelines for the next European Commission 2019-2024](#)', 2019.
- Wagner, J. and A. Benecke, 'National Legislation within the Framework of the GDPR', *European Data Protection Law (EDPL)*, Vol. 2, Number 3, 2016, pp. 353-361.
- Wiewiórowski, W., 2020, '[AI and Facial Recognition: Challenges and Opportunities](#)', *European Data Protection Supervisor (EDPS) Blog*, 21 February 2020.
- Wilson, D., 'Algorithmic patrol: The futures of predictive policing', in A. Završnik (ed.), *Big Data, Crime and Social Control*, Routledge, London and New York, 2017, pp. 108-127.
- World Commission on the Ethics of Scientific Knowledge and Technology (COMEST), *Preliminary Study on The Ethics Of Artificial Intelligence*, UNESCO, 26 February 2019.
- World Economic Forum (WEF), *A Framework for Responsible Limits on Facial Recognition Use Case: Flow Management, Pilot project*, White Paper, February 2020.
- World Health Organisation, *COVID-19 virtual press conference - 25 March 2020*, 2020.
- Yang, Y., '[How China built facial recognition for people wearing masks](#)', *Ars Technica*, 18 March 2020.
- Yun Chee, F., '[EU drops idea of facial recognition ban in public areas: paper](#)', *Reuters*, 30 January 2020.
- Zanfir-Fortuna, G., '[European Union's Data-Based Policy Against the Pandemic, Explained](#)', *Future of Privacy Forum*, 30 April 2020,
- Završnik, A., 'Big data: What is it and why does it matter for crime and social control?' in A. Završnik (ed.), *Big Data, Crime and Social Control*, Routledge, London and New York, 2017, pp. 3-28. (2017a)
- , Završnik, A., 'Algorithmic crime control' in A. Završnik (ed.), *Big Data, Crime and Social Control*, Routledge, London and New York, 2017, pp. 131-153 (2017b).

ANNEX – A COMPARISON BETWEEN ‘ALGORITHMIC DATA RIGHTS’ IN THE GDPR AND THE LED

This Annex offers a succinct review of the **rights granted to data subjects** by the GDPR and by the LED **in relation to profiling and automated decision-making**. These rights might be referred to as ‘algorithmic data rights’ to the extent that they are the ones most directly concerned with algorithmic data processing. It is necessary to underline that this does not mean that other data subject rights might not be relevant too. It is also important to stress that these rights might not systematically apply when processing related to law enforcement and criminal justice are at stake, as both the GDPR and the LED allow for restrictions and derogations. The main purpose of this review is to illustrate that these rights, which are **at the core of the intersection between EU data protection law and AI**, have been given substance **differently in the GDPR and the LED**.

Representatives of EU DPAs, gathered as Article 29 Working Party, elaborated in 2017 *Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679* which were later endorsed by the EDPB (A29WP, 2017/2018). There exist no equivalent guidelines about the LED. The Guidelines emphasise in any case that ‘*profiling and automated decision-making can pose significant risks for individuals’ rights and freedoms which require appropriate safeguards*’ (ibid., 5).

The risks identified evolve around the **opacity** and **discrimination**. **Opacity** might affect the very fact that one is being profiled, and the nature of such profiling (idem). Concerns about opacity are primarily addressed in the GDPR through the principle of transparency, which takes precisely as starting point that ‘*the process of profiling is often invisible to the data subject*’ (ibid., 9). Additionally, opacity relates to the fact that ‘*(i)ndividuals have differing levels of comprehension and may find it challenging to understand the complex techniques involved in profiling and automated decision-making processes*’ (idem). **Discrimination** concerns are multiple: profiling can perpetuate existing stereotypes and segregation, locking a person into a specific category, and restricting them to suggested preferences. Also, being placed in the wrong category might lead to unjustified discrimination (ibid., 5).

Concerns about discrimination are primarily addressed in the GDPR through the **principle of fairness**, also present in Article 4(1)(a) of the LED. The Guidelines stress that ‘*(p)rofiling may be unfair and create discrimination*’ (ibid., 10). The GDPR has a number of provisions which aim at ensuring that profiling and automated individual decision-making (whether or not it includes profiling) are not used in ways that have an unjustified impact on individuals’ rights, most notably, in addition to specific transparency and fairness requirements, accountability obligations; specified legal bases for the processing; rights for individuals to oppose profiling and specifically profiling for marketing purposes; and, if certain conditions are met, the obligation to carry out a data protection impact assessment (ibid., 6). Recital (26) of the LED points out: ‘*The data protection principle of fair processing is a distinct notion from the right to a fair trial as defined in Article 47 of the Charter and in Article 6 of the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR)*’.

According to the Guidelines, ‘*(g)iven the core **principle of transparency** underpinning the **GDPR**, controllers must ensure they explain clearly and simply to individuals how the profiling or automated decision-making process works*’ (ibid., 16). ‘*In particular*’, they specify, ‘*where the processing involves profiling-based decision making (irrespective of whether it is caught by Article 22 provisions), then the fact that the processing is for the purposes of both (a) profiling and (b) making a decision based on the profile generated, must be made clear to the data subject*’ (idem). Recital (60) of the GDPR states indeed that ‘*the data subject should **be informed of the existence of profiling and the consequences** of such profiling*’, as part of the GDPR principle of transparency (ibid., 17).

The **principle of transparency** is **not explicitly mentioned** as such in the LED. It does not appear in Article 4, which lists the Principles relating to processing of personal data: Article 4(1)(a) merely states that *'Member States shall provide for personal data to be (...) processed lawfully and fairly'*. There is a reference to transparency in the Recitals of the LED, concretely at Recital (26), which states that *'(a)ny processing of personal data must be lawful, fair and transparent in relation to the natural persons concerned, and only processed for specific purposes laid down by law'*. There is not, in the LED, any explicit reference to the fact the data subject should be informed of the existence of profiling and the consequences of such profiling.

In addition to a right to be informed, there exists a **right of access** – also mentioned in Article 8 of the EU Charter, together with the right to rectification. The Guidelines state that Article 15 of the GDPR, on the right of access, *'gives the data subject the right to obtain details of **any personal data used for profiling**, including the categories of data used to construct a profile'*, and specified that *'(i)n addition to general information about the processing, pursuant to Article 15(3), the controller has a duty to make available the data used as input to create the profile as well as access to **information on the profile** and details of **which segments** the data subject has been placed into'* (ibid., 17).

Said Article 15 of the GDPR refers to the obligation imposed on data controllers to *'provide **a copy** of the personal data undergoing processing'*. The right of access under the LED does not foresee an equivalent obligation imposed on data controllers to provide a copy of the data at stake. Recital (43) of the LED indicates that data controllers might comply with the right of access by providing to the data subject *'a full summary of those data'*.

Data subjects have the right to **rectification** and **erasure**. The Guidelines note that in the GDPR *'(t)he rights to rectification and erasure apply to both the 'input personal data' (the personal data used to create the profile) and the 'output data' (the profile itself or 'score' assigned to the person)'* (ibid., 18). The right to rectification and to erasure are also present in the LED, although the provision establishing them also indicates that in certain cases a refusal by the data controller is permissible. Member States must make sure that when such is the case the data controller informs the data subject in writing about the refusal, although Member States may adopt legislative measures restricting, wholly or partly, the obligation to provide such information under certain conditions (Article 16(4) of the LED).

Data subjects have a right to **object**. Under Article 21 of the GDPR, the right to object applies to the cases where processing of personal data is *'based on point (e) or (f) of Article 6(1), including **profiling** based on those provisions'*, that is, when the processing is 'necessary for the performance of a task carried out in the **public interest** or in the exercise of official authority vested in the controller' (Article 6(1)(e) of the GDPR) or *'processing is necessary for the purposes of the **legitimate interests** pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child'* (Article 6(1)(f) of the GDPR). There is no equivalent right to object under the LED.

Article 22 of the GDPR, still in the GDPR Chapter about data subject rights, is titled *Automated individual decision-making, including profiling*, and its first paragraph states that *'The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her'*. The Guidelines take the view that the use of the term 'right' here does not imply that Article 22(1) applies only when actively invoked by the data subject. Instead, Article 22(1) of the GDPR would establish **a general prohibition for decision-making based solely on automated processing**, applying nevertheless in a limited number of circumstances (the decision based solely on automated processing, and has a legal effect on or similarly significantly affects someone), and subject to exceptions. According to the Guidelines, Article 22 of the GDPR provides that under the GDPR *'as a rule, there is a general prohibition on fully automated individual decision-making, including profiling that has a legal or similarly significant*

effect, although *'there are exceptions to the rule'*, even if *'where one of these exceptions applies, there must be measures in place to safeguard the data subject's rights and freedoms and legitimate interests'* (ibid., 19).

Article 22(1) of the GDPR applies to decisions based solely on automated processing, including profiling, which produce legal effects concerning or similarly significantly affect the data subject. The Guidelines clarify that *'(p)rocessing that might have little impact on individuals generally may in fact have a significant effect for certain groups of society, such as minority groups or vulnerable adults'* (ibid., 22).

The safeguards to be applied when automated decision-making based solely on automated processing, which produces legal effects or similarly significantly affects the data subject is allowed encompass a **right to be informed** (specifically, to be provided with meaningful information about the logic involved, as well as the significance and envisaged consequences for the data subject), and other safeguards such as **the right to obtain human intervention** and **the right to challenge the decision** (addressed in Article 22(3)) (ibid., 20).

In the LED, there is a reference for Member States having to establish that decisions based solely on automated processing, including profiling, which produce an **adverse legal effect** concerning the data subject or significantly affects them, shall be prohibited unless authorised by an EU or Member State law to which the controller is subject *'and which provides appropriate safeguards for the rights and freedoms of the data subject, at least the right to obtain human intervention on the part of the controller'* (Article 11(1) of the LED). Automated decisions producing adverse legal effects are thus **not fully prohibited** under the LED, but conditioned to the existence of a legal provision providing for some safeguards. Article 11 of the LED, contrary to Article 22 of the GDPR, appears outside of the Chapter on data subject rights. There is no explicit reference to the need to foresee a right to challenge the decision, although there is a mention to this issue in Recital (38), which portrays the possibility to challenge a decision as a suitable safeguard when evaluating personal aspects of an individual is based solely on automated processing and which produces adverse legal effects, or significantly affects them.

In relation to these safeguards, the Guidelines noted that *'(a)ny processing likely to result in a high risk to data subjects requires the controller to carry out a **Data Protection Impact Assessment (DPIA)**'*, which might be *'particularly useful for controllers who are unsure whether their proposed activities will fall within the Article 22(1) definition, and, if allowed by an identified exception, what safeguarding measures must be applied'* (ibid., 20). Both the GDPR and the LED foresee that DPIAs are necessary in some cases (cf. Article 35 of the GDPR, and Article 27 of the LED). The requirements for DPIAs are however not the same. The GDPR notably establishes that *'(w)here appropriate, the controller shall seek the views of data subjects or their representatives on the intended processing'* (Article 35(9) of the GDPR), whereas the LED does not have an equivalent provision.

Under the GDPR, it is possible to exceptionally undertake the processing described in Article 22(1) where the decision is (a) necessary for the performance of or entering into a contract, (b) authorised by EU or Member State law to which the controller is subject laying down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or (c) based on the data subject's explicit consent. Regarding (b), Recital (71) of the GDPR notes that this might refer to *'fraud and tax-evasion monitoring and prevention purposes conducted in accordance with the regulations, standards and recommendations of Union institutions or national oversight bodies and to ensure the security and reliability of a service provided by the controller'*.

In line with Article 22(4) of the GDPR, automated decision-making covered by Article 22(1) of the GDPR that involves special categories of personal data is only allowed under the following cumulative conditions of falling under an applicable Article 22(2) exemption; and falling under point (a) or (g) of Article 9(2), which refer to the need of explicit consent of the data subject; or necessity for reasons of substantial public interest, on the basis of EU or Member State law which shall be proportionate to the

aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and interests of the data subject. In both cases, the controller must put in place suitable measures to safeguard the data subject's rights and freedoms and legitimate interests. In Article 11 of the LED there is no reference to consent. Under this provision, decisions covered by Article 11 of the LED shall not be based on special categories of personal data, *'unless suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place'*.

Moving specifically to the rights of the data subject related to decisions falling under Article 22 of the GDPR, the Guidelines noted that *'(g)iven the potential risks and interference that profiling caught by Article 22 poses to the rights of data subjects, data controllers should be **particularly mindful of their transparency obligations**'* (ibid., 24). Articles 13(2)(f) and 14(2)(g) of the GDPR require data controllers to provide information about *'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'*.

As explained by the Article 29 Working Party, this means that under the GDPR data controllers must **inform the data subject** that they are engaging in this type of activity, provide meaningful information about the logic involved, and explain the significance and envisaged consequences of the processing. According to the Guidelines, even when automated decision-making and profiling do not fall under Article 22(1) of the GDPR, *'it is nevertheless good practice to provide the above information'* (AWP29, 2017/2018, 25). Moreover, it adds that in any event data controllers *'must provide sufficient information to the data subject to make the processing fair'* (idem, in reference to Recital (60)).

Under the LED, there are **no equivalent information obligations**. The provision detailing the 'Information to be made available or given to the data subject', that is, Article 13 of the LED, does not include any reference to profiling or automated decision-making. The only explicit reference to profiling in the Recitals of the LED is the allusion to the fact that profiling that results in discrimination against natural persons on the basis of personal data which are by their nature particularly sensitive in relation to fundamental rights and freedoms *'should be prohibited under the conditions laid down in Articles 21 and 52 of the Charter'* (Recital (38) of the LED).

Regarding the relation between the decisions falling under Article 22 of the GDPR and the **right of access**, Article 15(1)(g) establishes that data subjects exercising their right of access shall have the right to receive information on *'the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject'*. There is no reference here to the need to provide a detailed explanation as to why a specific decision was taken.

The Guidelines noted, in reference to the GDPR, that *'by exercising their Article 15 rights, the data subject can become aware of a decision made concerning him or her, including one based on profiling'* (ibid., 27). In that case, *'(t)he controller should provide the data subject with general information (notably, on factors taken into account for the decision-making process, and on their respective 'weight' on an aggregate level) which is also useful for him or her to challenge the decision'* (idem). This refers to instrumentality of transparency obligations and data subject rights in EU data protection law: it is by being informed about certain data processing practices that the data subject will be in a position to further exercise their rights.

Regarding suitable safeguards, the Guidelines pointed out in relation to the GDPR that the requirement of 'human intervention' relates to the fact that *'(a)ny review must be carried out by someone who has the appropriate authority and capability to change the decision'*, and that *'(t)he reviewer should undertake a*

thorough assessment of all the relevant data, including any additional information provided by the data subject' (ibid., 27).

Suitable safeguards under the GDPR include, in addition to the **right to obtain human intervention**, the **right for the data subject to express their point of view** and the **right to contest the decision**. The Guidelines stated that the controller '*must provide a simple way for the data subject to exercise these rights*', and that they emphasise the 'need for transparency about the processing', as '*the data subject will only be able to challenge a decision or express their view if they fully understand how it has been made and on what basis*' (ibid., 27).

Other safeguards highlighted by the Guidelines include safeguards aimed at mitigating the risk of errors or bias in collected or shared data, or in the automated decision-making process, that can result in '*incorrect classifications; and assessments based on imprecise projections; that impact negatively on individuals*' (ibid., 27). These safeguards can include '*carrying out frequent assessments on the data sets they process to check for any bias, and develop ways to address any prejudicial elements, including any over-reliance on correlations*' or '*(s)ystems that audit algorithms and regular reviews of the accuracy and relevance of automated decision-making including profiling*' (ibid., 28). In addition, controllers should introduce appropriate procedures and measures to prevent errors, inaccuracies or discrimination on the basis of special categories of data, measures which should be used on a cyclical basis: not only at the design stage, but also continuously and making sure the outcome of such testing feeds back into the system design (idem). All this is connected to **Recital (71) of the GDPR**, which mentions the need to use appropriate mathematical or statistical procedures.

This study, commissioned by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs at the request of the LIBE Committee, examines the impact on fundamental rights of Artificial Intelligence in the field of law enforcement and criminal justice, from a European Union perspective. It presents the applicable legal framework (notably in relation to data protection), and analyses major trends and key policy discussions. The study also considers developments following the Covid-19 outbreak. It argues that the seriousness and scale of challenges may require intervention at EU level, based on the acknowledgement of the area's specificities.

PE 656.295
IP/C/LIBE/IC/2019-083

Print ISBN 978-92-846-6938-7 | doi: 10.2861/138135 | QA-04-20-418-EN -C
PDF ISBN 978-92-846-6939-4 | doi:10.2861/ 880024 | QA-04-20-418-EN -N



Brussels, 8.4.2019
COM(2019) 168 final

**COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL
COMMITTEE AND THE COMMITTEE OF THE REGIONS**

Building Trust in Human-Centric Artificial Intelligence

1. INTRODUCTION — THE EUROPEAN AI STRATEGY

Artificial intelligence (AI) has the potential to transform our world for the better: it can improve healthcare, reduce energy consumption, make cars safer, and enable farmers to use water and natural resources more efficiently. AI can be used to predict environmental and climate change, improve financial risk management and provides the tools to manufacture, with less waste, products tailored to our needs. AI can also help to detect fraud and cybersecurity threats, and enables law enforcement agencies to fight crime more efficiently.

AI can benefit the whole of society and the economy. It is a strategic technology that is now being developed and used at a rapid pace across the world. Nevertheless, AI also brings with it new challenges for the future of work, and raises legal and ethical questions.

To address these challenges and make the most of the opportunities which AI offers, the Commission published a European strategy¹ in April 2018. The strategy places people at the centre of the development of AI — **human-centric AI**. It is a three-pronged approach to boost the EU's technological and industrial capacity and AI uptake across the economy, prepare for socio-economic changes, and ensure an appropriate ethical and legal framework.

To deliver on the AI strategy, **the Commission developed together with Member States a coordinated plan on AI²**, which it presented in December 2018, to create synergies, pool data — the raw material for many AI applications — and increase joint investments. The aim is to foster cross-border cooperation and mobilise all players to increase public and private investments to **at least EUR 20 billion** annually over the next decade³. The Commission doubled its investments in AI in Horizon 2020 and plans to invest EUR 1 billion annually from Horizon Europe and the Digital Europe Programme, in support notably of common data spaces in health, transport and manufacturing, and large experimentation facilities such as smart hospitals and infrastructures for automated vehicles and a strategic research agenda.

To implement such a common strategic research, innovation and deployment agenda the Commission has intensified its **dialogue with all relevant stakeholders** from industry, research institutes and public authorities. The new Digital Europe programme will also be crucial in helping to make AI available to small and medium-size enterprises across all Member States through digital innovation hubs, strengthened testing and experimentation facilities, data spaces and training programmes.

Building on its reputation for safe and high-quality products, Europe's ethical approach to AI strengthens citizens' trust in the digital development and aims at building a competitive advantage for European AI companies. The purpose of this Communication is to launch a comprehensive piloting phase involving stakeholders on the widest scale in order to test the practical implementation of ethical guidance for AI development and use.

2. BUILDING TRUST IN HUMAN-CENTRIC AI

The European AI strategy and the coordinated plan make clear that **trust is a prerequisite to ensure a human-centric approach to AI**: AI is not an end in itself, but a tool that has to serve people with the ultimate aim of increasing human well-being. To achieve this, the

¹ COM(2018) 237.

² COM(2018) 795.

³ To help reach this goal, the Commission proposed, under the next programming period 2021-2027, that the Union allocates at least EUR 1 billion per year in funding from the Horizon Europe and Digital Europe programmes to invest in AI.

trustworthiness of AI should be ensured. The values on which our societies are based need to be fully integrated in the way AI develops.

The Union is founded on **the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights**, including the rights of persons belonging to minorities⁴. These values are common to the societies of all Member States in which pluralism, non-discrimination, tolerance, justice, solidarity and equality prevail. In addition, the **EU Charter of Fundamental Rights** brings together – in a single text – the personal, civic, political, economic and social rights enjoyed by people within the EU.

The EU has a **strong regulatory framework** that will set the global standard for human-centric AI. The General Data Protection Regulation ensures a high standard of protection of personal data, and requires the implementation of measures to ensure data protection by design and by default⁵. The Free Flow of Non-Personal Data Regulation removes barriers to the free movement of non-personal data and ensures the processing of all categories of data anywhere in Europe. The recently adopted Cybersecurity Act will help to strengthen trust in the online world, and the proposed ePrivacy Regulation⁶ also aims at this goal.

Nevertheless, AI brings new challenges because it enables machines to “learn” and to take and implement decisions without human intervention. Before long, this kind of functionality will become standard in many types of goods and services, from smart phones to automated cars, robots and online applications. Yet, decisions taken by algorithms could result from data that is incomplete and therefore not reliable, they may be tampered with by cyber-attackers, or they may be biased or simply mistaken. Unreflectively applying the technology as it develops would therefore lead to problematic outcomes as well as reluctance by citizens to accept or use it.

Instead, AI technology should be developed in a way that puts people at its centre and is thus worthy of the public’s trust. This implies that AI applications should not only be consistent with the law, but also adhere to ethical principles and ensure that their implementations avoid unintended harm. Diversity in terms of gender, racial or ethnic origin, religion or belief, disability and age should be ensured at every stage of AI development. AI applications should empower citizens and respect their fundamental rights. They should aim to enhance people’s abilities, not replace them, and also enable access by people with disabilities.

Therefore, there is a need for **ethics guidelines** that build on the existing regulatory framework and that should be applied by developers, suppliers and users of AI in the internal market, establishing an ethical level playing field across all Member States. This is why the Commission has set up a **high-level expert group on AI**⁷ representing a wide range of stakeholders and has tasked it with drafting AI ethics guidelines as well as preparing a set of recommendations for broader AI policy. At the same time, the **European AI Alliance**⁸, an open multi-stakeholder platform with over 2700 members, was set up to provide broader input for the work of the AI high-level expert group.

⁴ In addition, the EU is a party to the UN Convention on the Rights of persons with disabilities.

⁵ Regulation (EU) 2016/679. The General Data Protection Regulation (GDPR) guarantees the free flow of personal data within the Union. It contains provisions on decision-making based solely on automated processing, including profiling. The individuals concerned have the right to be informed about the existence of automated-decision making and to receive meaningful information about the logic involved in the automated decision-making and about the significance and envisaged consequences of the processing for them. They also have the right in such cases to obtain human intervention, to express their point of view and to contest the decision.

⁶ COM(2017) 10.

⁷ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>

⁸ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

The AI high-level expert group published a first draft of the ethics guidelines in December 2018. Following a **stakeholder consultation**⁹ and **meetings with representatives from Member States**¹⁰, the AI expert group has delivered a revised document to the Commission in March 2019. In their feedback so far, stakeholders overall have welcomed the practical nature of the guidelines and the concrete guidance they offer to developers, suppliers and users of AI on how to ensure trustworthiness.

2.1. Guidelines for trustworthy AI drafted by the AI high-level expert group

The guidelines drafted by the AI high-level expert group, to which this Communication refers¹¹, build in particular on the work done by the European Group on Ethics in Science and New Technologies and the Fundamental Rights Agency.

The guidelines postulate that in order to achieve ‘trustworthy AI’, three components are necessary: (1) it should comply with the law, (2) it should fulfil ethical principles and (3) it should be robust.

Based on these three components and the European values set out in section 2, the guidelines identify seven key requirements that AI applications should respect to be considered trustworthy. The guidelines also include an assessment list to help check whether these requirements are fulfilled.

The seven key requirements are:

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

While these requirements are intended to apply to all AI systems in different settings and industries, the specific context in which they are applied should be taken into account for their concrete and proportionate implementation, taking an impact-based approach. For illustration, AI application suggesting an unsuitable book to read is much less perilous than misdiagnosing a cancer and could therefore be subject to less stringent supervision.

The guidelines drafted by the AI high-level expert group are non-binding and as such do not create any new legal obligations. However, many existing (and often use- or domain-specific)

⁹ This consultation resulted in comments from 511 organisations, associations, companies, research institutes, individuals and others. A summary of the feedback received is available at: https://ec.europa.eu/futurium/en/system/files/ged/consultation_feedback_on_draft_ai_ethics_guidelines_4.pdf

¹⁰ The work of the expert group was positively received by Member States, with the Council conclusions adopted on 18 February 2019 *inter alia* taking note of the forthcoming publication of the ethics guidelines and supporting the Commission's effort to bring an EU ethical approach to the global stage: <https://data.consilium.europa.eu/doc/document/ST-6177-2019-INIT/en/pdf>

¹¹ <https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines#Top>

provisions of Union law of course already reflect one or several of these key requirements, for example safety, personal data protection, privacy or environmental protection rules.

The Commission welcomes the work of the AI high-level expert group and considers it valuable input for its policy-making.

2.2. Key requirements for trustworthy AI

The Commission supports the following key requirements for trustworthy AI, which are based on European values. It encourages stakeholders to apply the requirements and to test the assessment list that operationalises them in order to create the right environment of trust for the successful development and use of AI. The Commission welcomes feedback from stakeholders to evaluate whether this assessment list provided in the guidelines requires further adjustment.

I. Human agency and oversight

AI systems should support individuals in making better, more informed choices in accordance with their goals. They should act as enablers to a flourishing and equitable society by supporting human agency and **fundamental rights**, and not decrease, limit or misguide human autonomy. The overall **wellbeing of the user** should be central to the system's functionality.

Human oversight helps ensuring that an AI system does not undermine human autonomy or causes other adverse effects. Depending on the specific AI-based system and its application area, the appropriate degrees of **control measures**, including the adaptability, accuracy and explainability of AI-based systems, should be ensured¹². **Oversight** may be achieved through governance mechanisms such as ensuring a human-in-the-loop, human-on-the-loop, or human-in-command approach.¹³ It must be ensured that public authorities have the ability to exercise their oversight powers in line with their mandates. All other things being equal, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance is required.

II. Technical robustness and safety

Trustworthy AI requires algorithms to be secure, reliable and robust enough to deal with errors or inconsistencies during all life cycle phases of the AI system, and to adequately cope with erroneous outcomes. AI systems need to be **reliable**, secure enough to be

¹² The General Data Protection Regulation gives individuals the right not to be subject to a decision based solely on automated processing when this produces legal effects on users or similarly significantly affects them (Article 22 GDPR).

¹³ Human-in-the-loop (HITL) refers to the human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. Human-on-the-loop (HOTL) refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. Human-in-command (HIC) refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by the system.

resilient against both overt attacks and more subtle attempts to manipulate data or algorithms themselves, and they must ensure a **fall-back plan** in case of problems. Their decisions must be **accurate**, or at least correctly reflect their level of accuracy, and their outcomes should be **reproducible**.

In addition, AI systems should integrate safety and security-by-design mechanisms to ensure that they are **verifiably safe** at every step, taking at heart the physical and mental safety of all concerned. This includes the minimisation and where possible the reversibility of unintended consequences or errors in the system's operation. Processes to clarify and assess potential risks associated with the use of AI systems, across various application areas, should be put in place.

III. Privacy and Data Governance

Privacy and **data protection** must be guaranteed at **all stages** of the AI system's life cycle. Digital records of human behaviour may allow AI systems to infer not only individuals' preferences, age and gender but also their sexual orientation, religious or political views. To allow individuals to trust the data processing, it must be ensured that they have full control over their own data, and that data concerning them will not be used to harm or discriminate against them.

In addition to safeguarding privacy and personal data, requirements must be fulfilled to ensure high quality AI systems. The quality of the data sets used is paramount to the performance of AI systems. When data is gathered, it may reflect socially constructed biases, or contain inaccuracies, errors and mistakes. This needs to be addressed prior to training an AI system with any given data set. In addition, the **integrity** of the data must be ensured. Processes and data sets used must be tested and documented at each step such as planning, training, testing and deployment. This should also apply to AI systems that were not developed in-house but acquired elsewhere. Finally, the **access** to data must be adequately governed and controlled.

IV. Transparency

The **traceability** of AI systems should be ensured; it is important to log and document both the decisions made by the systems, as well as the entire process (including a description of data gathering and labelling, and a description of the algorithm used) that yielded the decisions. Linked to this, **explainability** of the algorithmic decision-making process, adapted to the persons involved, should be provided to the extent possible. Ongoing research to develop explainability mechanisms should be pursued. In addition, explanations of the degree to which an AI system influences and shapes the organisational decision-making process, design choices of the system, as well as the rationale for deploying it, should be available (hence ensuring not just data and system transparency, but also business model transparency).

Finally, it is important to adequately **communicate** the AI system's capabilities and limitations to the different stakeholders involved in a manner appropriate to the use case at hand. Moreover, AI systems should be identifiable as such, ensuring that users know they are interacting with an AI system and which persons are responsible for it.

V. Diversity, non-discrimination and fairness

Data sets used by AI systems (both for training and operation) may suffer from the inclusion of inadvertent historic bias, incompleteness and bad governance models. The continuation of such biases could lead to (in)direct discrimination. Harm can also result from the intentional exploitation of (consumer) biases or by engaging in unfair competition. Moreover, the way in which AI systems are developed (e.g. the way in which the programming code of an algorithm is written) may also suffer from bias. Such concerns should be tackled from the beginning of the system's development.

Establishing **diverse design teams** and setting up mechanisms ensuring **participation**, in particular of citizens, in AI development can also help to address these concerns. It is advisable to consult stakeholders who may directly or indirectly be affected by the system throughout its life cycle. AI systems should consider the whole range of human abilities, skills and requirements, and ensure accessibility through a universal design approach to strive to achieve equal access for persons with disabilities.

VI. Societal and environmental well-being

For AI to be trustworthy, its impact on the **environment and other sentient beings** should be taken into account. Ideally, all humans, including future generations, should benefit from biodiversity and a habitable environment. Sustainability and **ecological responsibility** of AI systems should hence be encouraged. The same applies to AI solutions addressing areas of global concern, such as for instance the UN Sustainable Development Goals.

Furthermore, the impact of AI systems should be considered not only from an individual perspective, but also from the perspective of **society as a whole**. The use of AI systems should be given careful consideration particularly in situations relating to the democratic process, including opinion-formation, political decision-making or electoral contexts. Moreover, AI's **social impact** should be considered. While AI systems can be used to enhance social skills, they can equally contribute to their deterioration.

VII. Accountability

Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their implementation. **Auditability** of AI systems is key in this regard, as the assessment of AI systems by internal and external auditors, and the availability of such evaluation reports, strongly contributes to the trustworthiness of the technology. External auditability should especially be ensured in applications affecting fundamental rights, including safety-critical applications.

Potential negative impacts of AI systems should be identified, assessed, documented and minimised. The use of impact assessments facilitates this process. These assessments should be proportionate to the extent of the risks that the AI systems pose. **Trade-offs** between the requirements – which are often unavoidable – should be addressed in a rational and methodological manner, and should be accounted for. Finally, when unjust adverse impact occurs, accessible mechanisms should be foreseen that ensure **adequate redress**.

2.3. Next steps: a piloting phase involving stakeholders on the widest scale

Reaching consensus on these key requirements for AI systems is a first important milestone towards guidelines for ethical AI. As a next step, the Commission will ensure that this guidance can be tested and implemented in practice.

To this end, the Commission will now launch a targeted piloting phase designed to obtain structured feedback from stakeholders. This exercise will focus in particular on the assessment list which the high-level expert group has drawn up for each of the key requirements.

This work will have two strands: (i) a piloting phase for the guidelines involving stakeholders who develop or use AI, including public administrations, and (ii) a continued stakeholder consultation and awareness-raising process across Member States and different groups of stakeholders, including industry and service sectors:

- (i) Starting in June 2019, all stakeholders and individuals will be invited to test the assessment list and provide feedback on how to improve it. In addition, the AI high-level expert group will set up an in-depth review with stakeholders from the private and the public sector to gather more detailed feedback on how the guidelines can be implemented in a wide range of application domains. All feedback on the guidelines' workability and feasibility will be evaluated by the end of 2019.
- (ii) In parallel, the Commission will organise further outreach activities, giving representatives of the AI high-level expert group the opportunity to present the guidelines to relevant stakeholders in the Member States, including industry and service sectors, and providing these stakeholders with an additional opportunity to comment on and contribute to the AI guidelines.

The Commission will take into account the work of the group of experts on ethics for connected and automated driving¹⁴ and work with EU-funded research projects on AI and with relevant public-private partnerships on implementing the key requirements¹⁵. For example, the Commission will support, in coordination with Member States, the development of a common database of health images initially dedicated to the most common forms of cancer, so that algorithms can be trained to diagnose symptoms with very high accuracy. Similarly, the cooperation of the Commission and Member States enables an increasing number of cross-border corridors for testing connected and automated vehicles. The guidelines should be applied in these projects and tested, and the results will feed into the evaluation process.

The piloting phase and the stakeholder consultation will benefit from the contribution of the European AI Alliance and AI4EU, the AI on-demand platform. The AI4EU project¹⁶, launched in January 2019, brings together algorithms, tools, datasets and services to help organisations, in particular small and medium-size enterprises, to implement AI solutions. The European AI Alliance, together with AI4EU, will continue to mobilise the AI ecosystem across Europe, also in view of piloting the AI ethics guidelines and promoting the respect for human-centric AI.

At the beginning of 2020, building on the evaluation of feedback received during the piloting phase, the AI high-level expert group will review and update the guidelines. Based on the

¹⁴ See the Commission's Communication on connected and automated mobility, COM(2018) 283.

¹⁵ In the Framework of the European Defence Fund, the Commission will also develop specific ethical guidance for the evaluation of project proposals in the area of AI for defence.

¹⁶ <https://ec.europa.eu/digital-single-market/en/news/artificial-intelligence-ai4eu-project-launches-1-january-2019>

review and on the experience acquired, **the Commission will evaluate the outcome and propose any next steps.**

Ethical AI is a win-win proposition. Guaranteeing the respect for fundamental values and rights is not only essential in itself, it also facilitates acceptance by the public and increases the competitive advantage of European AI companies by establishing a brand of human-centric, trustworthy AI known for ethical and secure products. This builds more generally on the strong reputation of European companies for providing safe and secure products of high quality. The pilot phase will help to ensure that AI products fulfil this promise.

2.4. Towards international AI ethics guidelines

International discussions on AI ethics have intensified after Japan's G7 Presidency put the topic high on the agenda in 2016. Given the international interlinkages of AI development in terms of data circulation, algorithmic development and research investments, **the Commission will continue its efforts to bring the Union's approach to the global stage and build a consensus on a human-centric AI¹⁷.**

The work done by the AI high-level expert group, and more specifically the list of requirements and the engagement process with stakeholders, provides the Commission with additional valuable input for contributing to the international discussions. The European Union can have a leadership role in developing international AI guidelines and, if possible, a related assessment mechanism.

Therefore, the Commission will:

Strengthen cooperation with like-minded partners:

- exploring the extent to which convergence can be achieved with third countries' draft ethics guidelines (e.g. Japan, Canada, Singapore) and, building on this group of like-minded countries, to prepare for a broader discussion, supported by actions implementing the Partnership Instrument for cooperation with Third Countries¹⁸; and
- exploring how companies from non-EU countries and international organisations can contribute to the 'pilot phase' of the guidelines through testing and validation.

Continue to play an active role in international discussions and initiatives:

- contributing to multilateral fora such as the G7 and G20;
- engaging in dialogues with non-EU countries and organising bilateral and multilateral meetings to build a consensus on human-centric AI;

¹⁷ The High Representative of the Union for Foreign Affairs and Security Policy will, with the support of the Commission, build on consultations in the United Nations, the Global Tech Panel, and other multilateral fora, and in particular coordinate proposals for addressing the complex security challenges involved.

¹⁸ Regulation (EU) No 234/2014 of the European Parliament and of the Council of 11 March 2014 establishing a Partnership Instrument for cooperation with third countries (OJ L 77, 15.3.2014, p. 77). For instance the planned project on 'An international alliance for a human-centric approach to artificial intelligence' will facilitate joint initiatives with like-minded partners, in order to promote ethical guidelines and to adopt common principles and operational conclusions. It will enable the EU and like-minded countries to discuss operational conclusions resulting from the ethical guidelines on AI proposed by the expert group in order to reach a common approach. Moreover, it will provide for monitoring the uptake of AI technology globally. Finally, the project plans to organise public diplomacy activities accompanying international events e.g. by the G7, G20 and the Organisation for Economic Cooperation and Development.

- contributing to relevant standardisation activities in international standards development organisations to promote this vision; and
- strengthening the collection and diffusion of insights on public policies, working jointly with relevant international organisations.

3. CONCLUSIONS

The EU is founded on a set of fundamental values and has constructed a strong and balanced regulatory framework on these foundations. Building on this existing regulatory framework, there is a need for ethics guidelines for the development and use of AI due to its novelty and the specific challenges this technology brings. Only if AI is developed and used in a way that respects widely-shared ethical values, it can be considered trustworthy.

With a view to this objective, the Commission welcomes the input prepared by the AI high-level expert group. Based on the key requirements for AI to be considered trustworthy, the Commission will now launch a targeted piloting phase to ensure that the resulting ethical guidelines for AI development and use can be implemented in practice. The Commission will also work to forge a broad societal consensus on human-centric AI, including with all involved stakeholders and our international partners.

The ethical dimension of AI is not a luxury feature or an add-on: it needs to be an integral part of AI development. By striving towards human-centric AI based on trust, we safeguard the respect for our core societal values and carve out a distinctive trademark for Europe and its industry as a leader in cutting-edge AI that can be trusted throughout the world.

To ensure the ethical development of AI in Europe in its wider context, the Commission is pursuing a comprehensive approach including in particular the following lines of action to be implemented by the third quarter of 2019:

- It will start launching a set of **networks of AI research excellence** centres through Horizon 2020. It will select up to four networks, focusing on scientific or technological major challenges such as explainability and advanced human-machine interaction, which are key ingredients for trustworthy AI.
- It will begin setting up **networks of digital innovation hubs**¹⁹ focussing on AI in manufacturing and on big data.
- Together with Member States and stakeholders, the Commission will start preparatory discussions to develop and implement **a model for data sharing and making best use of common data spaces**, with a focus notably on transport, healthcare and industrial manufacturing.²⁰

In addition, the Commission is working on a report on the challenges posed by AI to the safety and liability frameworks and a guidance document on the implementation of the Product Liability Directive²¹. At the same time, the European High-Performance Computing Joint Undertaking (EuroHPC)²² will develop the next generation of supercomputers because

¹⁹ <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs>

²⁰ The necessary resources will be mobilised from Horizon 2020 (under which close to 1.5 billion EUR are dedicated to AI for the period 2018-2020) and its planned successor Horizon Europe, the Digital part of the Connecting Europe Facility and especially the future Digital Europe Programme. Projects will also draw on resources from the private sector and Member State programmes.

²¹ See the Commission's Communication Artificial Intelligence for Europe, COM (2018) 237.

²² <https://eurohpc-ju.europa.eu>

computing capacity is essential for processing data and training AI and Europe needs to master the full digital value chain. The ongoing partnership with Member States and industry on microelectronic components and systems (ECSEL)²³ as well as the European Processor Initiative²⁴ will contribute to the development of low-power processor technology for trustworthy and secure high-performance and edge computing.

Just like the work on ethical guidelines for AI, all these initiatives build on **close cooperation of all concerned stakeholders**, Member States, industry, societal actors and citizens. Overall, Europe's approach to Artificial Intelligence shows how economic competitiveness and societal trust must start from the same fundamental values and mutually reinforce each other.

²³ www.ecsel.eu

²⁴ www.european-processor-initiative.eu



INDEPENDENT

HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE

SET UP BY THE EUROPEAN COMMISSION



ETHICS GUIDELINES FOR TRUSTWORTHY AI

ETHICS GUIDELINES FOR TRUSTWORTHY AI

High-Level Expert Group on Artificial Intelligence

This document was written by the High-Level Expert Group on AI (AI HLEG). The members of the AI HLEG named in this document support the overall framework for Trustworthy AI put forward in these Guidelines, although they do not necessarily agree with every single statement in the document.

The Trustworthy AI assessment list presented in Chapter III of this document will undergo a piloting phase by stakeholders to gather practical feedback. A revised version of the assessment list, taking into account the feedback gathered through the piloting phase, will be presented to the European Commission in early 2020.

The AI HLEG is an independent expert group that was set up by the European Commission in June 2018.

Contact Nathalie Smuha - AI HLEG Coordinator
E-mail CNECT-HLG-AI@ec.europa.eu

European Commission
B-1049 Brussels

Document made public on 8 April 2019.

A first draft of this document was released on 18 December 2018 and was subject to an open consultation which generated feedback from more than 500 contributors. We wish to explicitly and warmly thank all those who contributed their feedback on the document's first draft, which was considered in the preparation of this revised version.

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information. The contents of this working document are the sole responsibility of the High-Level Expert Group on Artificial Intelligence (AI HLEG). Although Commission staff facilitated the preparation of the Guidelines, the views expressed in this document reflect the opinion of the AI HLEG and may not in any circumstances be regarded as reflecting an official position of the European Commission.

More information on the High-Level Expert Group on Artificial Intelligence is available online (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

The reuse policy of European Commission documents is regulated by Decision 2011/833/EU (OJ L 330, 14.12.2011, p.39). For any use or reproduction of photos or other material that is not under the EU copyright, permission must be sought directly from the copyright holders.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	2
A. INTRODUCTION	4
B. A FRAMEWORK FOR TRUSTWORTHY AI	6
I. Chapter I: Foundations of Trustworthy AI	9
II. Chapter II: Realising Trustworthy AI	14
1. Requirements of Trustworthy AI	14
2. Technical and non-technical methods to realise Trustworthy AI	22
III. Chapter III: Assessing Trustworthy AI	24
C. EXAMPLES OF OPPORTUNITIES AND CRITICAL CONCERNS RAISED BY AI	32
D. CONCLUSION	35
GLOSSARY	36

EXECUTIVE SUMMARY

The aim of the Guidelines is to promote Trustworthy AI. Trustworthy AI has **three components**, which should be met throughout the system's entire life cycle: (1) it should be **lawful**, complying with all applicable laws and regulations (2) it should be **ethical**, ensuring adherence to ethical principles and values and (3) it should be **robust**, both from a technical and social perspective since, even with good intentions, AI systems can cause unintentional harm. Each component in itself is necessary but not sufficient for the achievement of Trustworthy AI. Ideally, all three components work in harmony and overlap in their operation. If, in practice, tensions arise between these components, society should endeavour to align them.

These Guidelines set out a **framework for achieving Trustworthy AI**. The framework does not explicitly deal with Trustworthy AI's first component (lawful AI).¹ Instead, it aims to offer guidance on the second and third components: fostering and securing ethical and robust AI. Addressed to all stakeholders, these Guidelines seek to go beyond a list of ethical principles, by providing guidance on how such principles can be operationalised in socio-technical systems. Guidance is provided in three layers of abstraction, from the most abstract in Chapter I to the most concrete in Chapter III, closing with examples of opportunities and critical concerns raised by AI systems.

- I. Based on an approach founded on fundamental rights, **Chapter I** identifies the **ethical principles** and their correlated values that must be respected in the development, deployment and use of AI systems.

Key guidance derived from Chapter I:

- ✓ Develop, deploy and use AI systems in a way that adheres to the ethical principles of: *respect for human autonomy, prevention of harm, fairness and explicability*. Acknowledge and address the potential tensions between these principles.
- ✓ Pay particular attention to situations involving more vulnerable groups such as children, persons with disabilities and others that have historically been disadvantaged or are at risk of exclusion, and to situations which are characterised by asymmetries of power or information, such as between employers and workers, or between businesses and consumers.²
- ✓ Acknowledge that, while bringing substantial benefits to individuals and society, AI systems also pose certain risks and may have a negative impact, including impacts which may be difficult to anticipate, identify or measure (e.g. on democracy, the rule of law and distributive justice, or on the human mind itself.) Adopt adequate measures to mitigate these risks when appropriate, and proportionately to the magnitude of the risk.

- II. Drawing upon Chapter I, **Chapter II** provides guidance on how Trustworthy AI can be realised, by listing **seven requirements** that AI systems should meet. Both technical and non-technical methods can be used for their implementation.

Key guidance derived from Chapter II:

- ✓ Ensure that the development, deployment and use of AI systems meets the seven key requirements for Trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability.
- ✓ Consider technical and non-technical methods to ensure the implementation of those requirements.

¹ All normative statements in this document aim to reflect guidance towards achieving the second and third component of trustworthy AI (ethical and robust AI). These statements are hence not meant to provide legal advice or to offer guidance on compliance with applicable laws, though it is acknowledged that many of these statements are to some extent already reflected in existing laws. In this regard, see §21 and following.

² See articles 24 to 27 of the Charter of Fundamental Rights of the EU (EU Charter), dealing with the rights of the child and the elderly, the integration of persons with disabilities and workers' rights. See also article 38 dealing with consumer protection.

- ✓ Foster research and innovation to help assess AI systems and to further the achievement of the requirements; disseminate results and open questions to the wider public, and systematically train a new generation of experts in AI ethics.
- ✓ Communicate, in a clear and proactive manner, information to stakeholders about the AI system's capabilities and limitations, enabling realistic expectation setting, and about the manner in which the requirements are implemented. Be transparent about the fact that they are dealing with an AI system.
- ✓ Facilitate the traceability and auditability of AI systems, particularly in critical contexts or situations.
- ✓ Involve stakeholders throughout the AI system's life cycle. Foster training and education so that all stakeholders are aware of and trained in Trustworthy AI.
- ✓ Be mindful that there might be fundamental tensions between different principles and requirements. Continuously identify, evaluate, document and communicate these trade-offs and their solutions.

III. **Chapter III** provides a concrete and non-exhaustive Trustworthy AI assessment list aimed at operationalising the key requirements set out in Chapter II. This **assessment list** will need to be tailored to the specific use case of the AI system.³

Key guidance derived from Chapter III:

- ✓ Adopt a Trustworthy AI assessment list when developing, deploying or using AI systems, and adapt it to the specific use case in which the system is being applied.
- ✓ Keep in mind that such an assessment list will never be exhaustive. Ensuring Trustworthy AI is not about ticking boxes, but about continuously identifying and implementing requirements, evaluating solutions, ensuring improved outcomes throughout the AI system's lifecycle, and involving stakeholders in this.

A final section of the document aims to concretise some of the issues touched upon throughout the framework, by offering examples of beneficial opportunities that should be pursued, and critical concerns raised by AI systems that should be carefully considered.

While these Guidelines aim to offer guidance for AI applications in general by building a horizontal foundation to achieve Trustworthy AI, different situations raise different challenges. It should therefore be explored whether, in addition to this horizontal framework, a sectorial approach is needed, given the context-specificity of AI systems.

These Guidelines do not intend to substitute any form of current or future policymaking or regulation, nor do they aim to deter the introduction thereof. They should be seen as a living document to be reviewed and updated over time to ensure their continuous relevance as the technology, our social environments, and our knowledge evolve. This document is a starting point for the discussion about "Trustworthy AI for Europe".⁴

Beyond Europe, the Guidelines also aim to foster research, reflection and discussion on an ethical framework for AI systems at a global level.

³ In line with the scope of the framework, this assessment list does not provide any advice on ensuring legal compliance (lawful AI), but limits itself to offering guidance on meeting the second and third components of trustworthy AI (ethical and robust AI).

⁴ This ideal is intended to apply to AI systems developed, deployed and used in the Member States of the European Union (EU), as well as to systems developed or produced elsewhere but deployed and used in the EU. When referring to "Europe" in this document, we mean this to encompass the EU Member States. However, these Guidelines also aspire to be relevant outside the EU. In this regard, it can also be noted that both Norway and Switzerland are part of the Coordinated Plan on AI agreed and published in December 2018 by the Commission and Member States.

A. INTRODUCTION

In its Communication of 25 April 2018 and 7 December 2018, the European Commission set out its vision for artificial intelligence (AI), which supports “ethical, secure and cutting-edge AI made in Europe”.⁵ Three pillars underpin the Commission’s vision: (i) increasing public and private investments in AI to boost its uptake, (ii) preparing for socio-economic changes, and (iii) ensuring an appropriate ethical and legal framework to strengthen European values.

To support the implementation of this vision, the Commission established the High-Level Expert Group on Artificial Intelligence (AI HLEG), an independent group mandated with the drafting of two deliverables: (1) AI Ethics Guidelines and (2) Policy and Investment Recommendations.

This document contains the AI Ethics Guidelines, which have been revised following further deliberation by our Group in light of feedback received from the public consultation on the draft published on 18 December 2018. It builds on the work of the European Group on Ethics in Science and New Technologies⁶ and takes inspiration from other similar efforts.⁷

Over the past months, the 52 of us met, discussed and interacted, committed to the European motto: united in diversity. We believe that AI has the potential to significantly transform society. AI is not an end in itself, but rather a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation. In particular, AI systems can help to facilitate the achievement of the UN’s Sustainable Development Goals, such as promoting gender balance and tackling climate change, rationalising our use of natural resources, enhancing our health, mobility and production processes, and supporting how we monitor progress against sustainability and social cohesion indicators.

To do this, AI systems⁸ need to be **human-centric**, resting on a commitment to their use in the service of humanity and the common good, with the goal of improving human welfare and freedom. While offering great opportunities, AI systems also give rise to certain risks that must be handled appropriately and proportionately. We now have an important window of opportunity to shape their development. We want to ensure that we can trust the socio-technical environments in which they are embedded. We also want producers of AI systems to get a competitive advantage by embedding Trustworthy AI in their products and services. This entails seeking to **maximise the benefits of AI systems** while at the same time **preventing and minimising their risks**.

In a context of rapid technological change, we believe it is essential that trust remains the bedrock of societies, communities, economies and sustainable development. We therefore identify **Trustworthy AI as our foundational ambition**, since human beings and communities will only be able to have confidence in the technology’s development and its applications when a clear and comprehensive framework for achieving its trustworthiness is in place.

This is the path that we believe Europe should follow to become the home and leader of cutting-edge and ethical technology. It is through Trustworthy AI that we, as European citizens, will seek to reap its benefits in a way that is aligned with our foundational values of respect for human rights, democracy and the rule of law.

Trustworthy AI

Trustworthiness is a prerequisite for people and societies to develop, deploy and use AI systems. Without AI systems – and the human beings behind them – being demonstrably worthy of trust, unwanted consequences may ensue and their uptake might be hindered, preventing the realisation of the potentially vast social and economic

⁵ COM(2018)237 and COM(2018)795. Note that the term “made in Europe” is used throughout the Commission’s communication. The scope of these Guidelines however aims to encompass not only those AI systems made in Europe, but also those developed elsewhere and deployed or used in Europe. Throughout this document, we hence aim to promote trustworthy AI “for” Europe.

⁶ The European Group on Ethics in Science and New Technologies (EGE) is an advisory group of the Commission.

⁷ See Section 3.3 of COM(2018)237.

⁸ The Glossary at the end of this document provides a definition of AI systems for the purpose of this document. This definition is further elaborated on in a dedicated document prepared by the AI HLEG that accompanies these Guidelines, titled “A definition of AI: Main capabilities and scientific disciplines”.

benefits that they can bring. To help Europe realise those benefits, our vision is to ensure and scale Trustworthy AI.

Trust in the development, deployment and use of AI systems concerns not only the technology's inherent properties, but also the qualities of the socio-technical systems involving AI applications.⁹ Analogous to questions of (loss of) trust in aviation, nuclear power or food safety, it is not simply components of the AI system but the system in its overall context that may or may not engender trust. Striving towards Trustworthy AI hence concerns not only the trustworthiness of the AI system itself, but requires a holistic and systemic approach, encompassing the trustworthiness of all actors and processes that are part of the system's socio-technical context throughout its entire life cycle.

Trustworthy AI has **three components**, which should be met throughout the system's entire life cycle:

1. it should be **lawful**, complying with all applicable laws and regulations;
2. it should be **ethical**, ensuring adherence to ethical principles and values; and
3. it should be **robust**, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm.

Each of these three components is necessary but not sufficient in itself to achieve Trustworthy AI.¹⁰ Ideally, all three work in harmony and overlap in their operation. In practice, however, there may be tensions between these elements (e.g. at times the scope and content of existing law might be out of step with ethical norms). It is our individual and collective responsibility as a society to work towards ensuring that all three components help to secure Trustworthy AI.¹¹

A trustworthy approach is key to enabling “responsible competitiveness”, by providing the foundation upon which all those affected by AI systems can trust that their design, development and use are lawful, ethical and robust. These Guidelines are intended to foster responsible and sustainable AI innovation in Europe. They seek to make ethics a core pillar for developing a unique approach to AI, one that aims to benefit, empower and protect both individual human flourishing and the common good of society. We believe that this will enable Europe to position itself as a global leader in cutting-edge AI worthy of our individual and collective trust. Only by ensuring trustworthiness will European individuals fully reap AI systems' benefits, secure in the knowledge that measures are in place to safeguard against their potential risks.

Just as the use of AI systems does not stop at national borders, neither does their impact. Global solutions are therefore required for the global opportunities and challenges that AI systems bring forth. We therefore encourage all stakeholders to work towards a global framework for Trustworthy AI, building international consensus while promoting and upholding our fundamental rights-based approach.

Audience and Scope

These guidelines are addressed to all AI stakeholders designing, developing, deploying, implementing, using or being affected by AI, including but not limited to companies, organisations, researchers, public services, government agencies, institutions, civil society organisations, individuals, workers and consumers. Stakeholders committed towards achieving Trustworthy AI can **voluntarily** opt to use these Guidelines as a method to operationalise their commitment, in particular by using the practical assessment list of Chapter III when developing, deploying or using AI systems. This assessment list can also complement – and hence be incorporated in – existing assessment processes.

The Guidelines aim to provide guidance for AI applications in general, building a horizontal foundation to achieve Trustworthy AI. However, **different situations raise different challenges**. AI music recommendation systems do not

⁹ These systems comprise humans, state actors, corporations, infrastructure, software, protocols, standards, governance, existing laws, oversight mechanisms, incentive structures, auditing procedures, best practices reporting and others.

¹⁰ This does not exclude the fact that additional conditions may be(come) necessary.

¹¹ This also means that the legislature or policy-makers may need to review the adequacy of existing law where these might be out of step with ethical principles.

raise the same ethical concerns as AI systems proposing critical medical treatments. Likewise, different opportunities and challenges arise from AI systems used in the context of business-to-consumer, business-to-business, employer-to-employee and public-to-citizen relationships, or more generally, in different sectors or use cases. Given the context-specificity of AI systems, the implementation of these Guidelines needs to be adapted to the particular AI-application. Moreover, the necessity of an additional sectorial approach, to complement the more general horizontal framework proposed in this document, should be explored.

To gain a better understanding of how this guidance can be implemented at a horizontal level, and of those matters that require a sectorial approach, we invite all stakeholders to pilot the Trustworthy AI assessment list (Chapter III) that operationalises this framework and to provide us feedback. Based on the feedback gathered through this piloting phase, we will revise the assessment list of these Guidelines by early 2020. The piloting phase will be launched by the summer of 2019 and last until the end of the year. All interested stakeholders will be able to participate by indicating their interest through the European AI Alliance.

B. A FRAMEWORK FOR TRUSTWORTHY AI

These Guidelines articulate a framework for achieving Trustworthy AI based on fundamental rights as enshrined in the Charter of Fundamental Rights of the European Union (EU Charter), and in relevant international human rights law. Below, we briefly touch upon Trustworthy AI's three components.

Lawful AI

AI systems do not operate in a lawless world. A number of legally binding rules at European, national and international level already apply or are relevant to the development, deployment and use of AI systems today. Legal sources include, but are not limited to: EU primary law (the Treaties of the European Union and its Charter of Fundamental Rights), EU secondary law (such as the General Data Protection Regulation, the Product Liability Directive, the Regulation on the Free Flow of Non-Personal Data, anti-discrimination Directives, consumer law and Safety and Health at Work Directives), the UN Human Rights treaties and the Council of Europe conventions (such as the European Convention on Human Rights), and numerous EU Member State laws. Besides horizontally applicable rules, various domain-specific rules exist that apply to particular AI applications (such as for instance the Medical Device Regulation in the healthcare sector).

The law provides both positive and negative obligations, which means that it should not only be interpreted with reference to what *cannot* be done, but also with reference to what *should* be done and what *may* be done. The law not only prohibits certain actions but also enables others. In this regard, it can be noted that the EU Charter contains articles on the 'freedom to conduct a business' and the 'freedom of the arts and sciences', alongside articles addressing areas that we are more familiar with when looking to ensure AI's trustworthiness, such as for instance data protection and non-discrimination.

The Guidelines do not explicitly deal with the first component of Trustworthy AI (lawful AI), but instead aim to offer guidance on fostering and securing the second and third components (ethical and robust AI). While the two latter are to a certain extent often already reflected in existing laws, their full realisation may go beyond existing legal obligations.

Nothing in this document shall be construed or interpreted as providing legal advice or guidance concerning how compliance with any applicable existing legal norms and requirements can be achieved. Nothing in this document shall create legal rights nor impose legal obligations towards third parties. We however recall that it is the duty of any natural or legal person to comply with laws – whether applicable today or adopted in the future according to the development of AI. **These Guidelines proceed on the assumption that all legal rights and obligations that apply to the processes and activities involved in developing, deploying and using AI systems remain mandatory and must be duly observed.**

Ethical AI

Achieving Trustworthy AI requires not only compliance with the law, which is but one of its three components. Laws

are not always up to speed with technological developments, can at times be out of step with ethical norms or may simply not be well suited to addressing certain issues. For AI systems to be trustworthy, they should hence also be ethical, ensuring alignment with ethical norms.

Robust AI

Even if an ethical purpose is ensured, individuals and society must also be confident that AI systems will not cause any unintentional harm. Such systems should perform in a safe, secure and reliable manner, and safeguards should be foreseen to prevent any unintended adverse impacts. It is therefore important to ensure that AI systems are robust. This is needed both from a technical perspective (ensuring the system's technical robustness as appropriate in a given context, such as the application domain or life cycle phase), and from a social perspective (in due consideration of the context and environment in which the system operates).

Ethical and robust AI are hence closely intertwined and complement each other. The principles put forward in Chapter I, and the requirements derived from these principles in Chapter II, address both components.

The framework

The Guidance in this document is provided in three chapters, from most abstract in Chapter I to most concrete in Chapter III:

- **Chapter I – Foundations of Trustworthy AI:** sets out the foundations of Trustworthy AI by laying out its fundamental-rights¹² based approach. It identifies and describes the ethical principles that must be adhered to in order to ensure ethical and robust AI.
- **Chapter II – Realising Trustworthy AI:** translates these ethical principles into seven key requirements that AI systems should implement and meet throughout their entire life cycle. In addition, it offers both technical and non-technical methods that can be used for their implementation.
- **Chapter III – Assessing Trustworthy AI:** sets out a concrete and non-exhaustive Trustworthy AI assessment list to operationalise the requirements of Chapter II, offering AI practitioners practical guidance. This assessment should be tailored to the particular system's application.

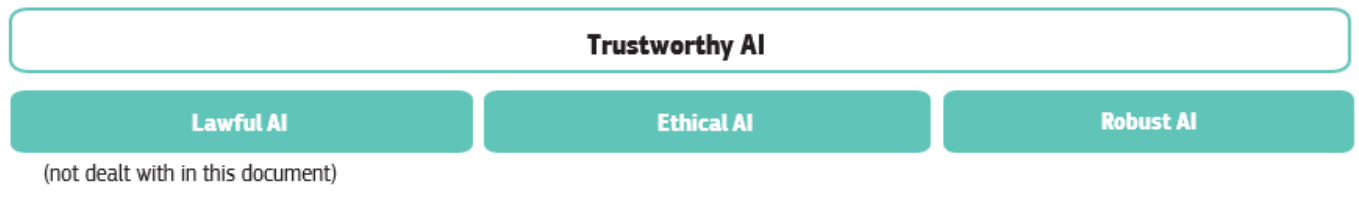
The document's final section lists examples of beneficial opportunities and critical concerns raised by AI systems, which should serve to stimulate further debate.

The Guidelines' structure is illustrated in *Figure 1* below

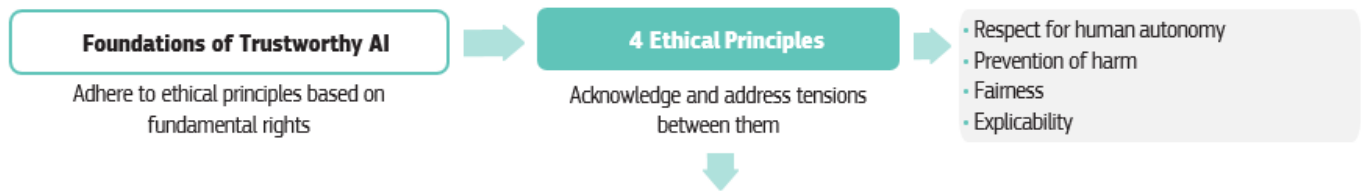
¹² Fundamental rights lie at the foundation of both international and EU human rights law and underpin the legally enforceable rights guaranteed by the EU Treaties and the EU Charter. Being legally binding, compliance with fundamental rights hence falls under trustworthy AI's first component (lawful AI). Fundamental rights can however also be understood as reflecting special moral entitlements of all individuals arising by virtue of their humanity, regardless of their legally binding status. In that sense, they hence also form part of the second component of trustworthy AI (ethical AI).

Framework for Trustworthy AI

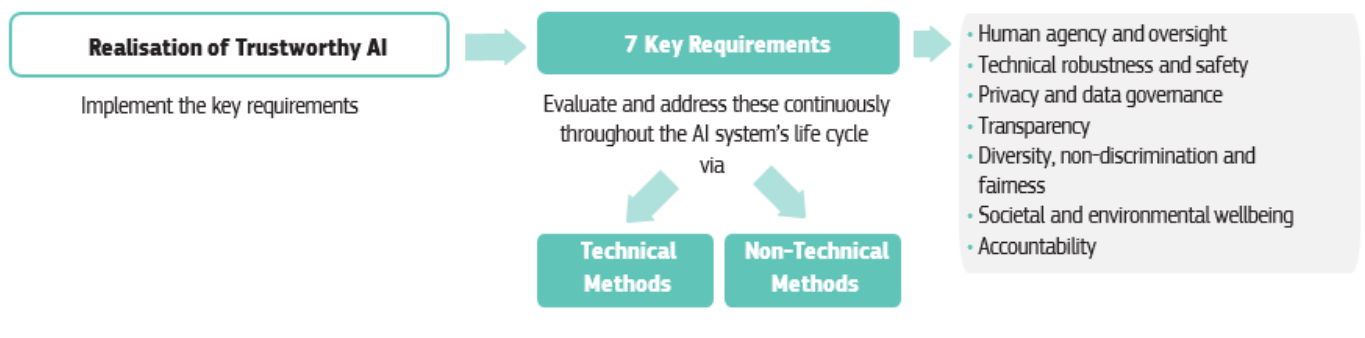
INTRODUCTION



CHAPTER I



CHAPTER II



CHAPTER III



Figure 1: The Guidelines as a framework for Trustworthy AI

I. Chapter I: Foundations of Trustworthy AI

This Chapter sets out the foundations of Trustworthy AI, grounded in fundamental rights and reflected by four ethical principles that should be adhered to in order to ensure ethical and robust AI. It draws heavily on the field of ethics.

AI ethics is a sub-field of applied ethics, focusing on the ethical issues raised by the development, deployment and use of AI. Its central concern is to identify how AI can advance or raise concerns to the good life of individuals, whether in terms of quality of life, or human autonomy and freedom necessary for a democratic society.

Ethical reflection on AI technology can serve multiple purposes. First, it can stimulate reflection on the need to protect individuals and groups at the most basic level. Second, it can stimulate new kinds of innovations that seek to foster ethical values, such as those helping to achieve the UN Sustainable Development Goals¹³, which are firmly embedded in the forthcoming EU Agenda 2030.¹⁴ While this document mostly concerns itself with the first purpose mentioned, the importance that ethics could have in the second should not be underestimated. Trustworthy AI can improve individual flourishing and collective wellbeing by generating prosperity, value creation and wealth maximization. It can contribute to achieving a fair society, by helping to increase citizens' health and well-being in ways that foster equality in the distribution of economic, social and political opportunity.

It is therefore imperative that we understand how to best support AI development, deployment and use to ensure that everyone can thrive in an AI-based world, and to build a better future while at the same time being globally competitive. As with any powerful technology, the use of AI systems in our society raises several ethical challenges, for instance relating to their impact on people and society, decision-making capabilities and safety. If we are increasingly going to use the assistance of or delegate decisions to AI systems, we need to make sure these systems are fair in their impact on people's lives, that they are in line with values that should not be compromised and able to act accordingly, and that suitable accountability processes can ensure this.

Europe needs to define what normative vision of an AI-immersed future it wants to realise, and understand which notion of AI should be studied, developed, deployed and used in Europe to achieve this vision. With this document, we intend to contribute to this effort by introducing the notion of Trustworthy AI, which we believe is the right way to build a future with AI. A future where democracy, the rule of law and fundamental rights underpin AI systems and where such systems continuously improve and defend democratic culture will also enable an environment where innovation and responsible competitiveness can thrive.

A domain-specific ethics code – however consistent, developed and fine-grained future versions of it may be – can never function as a substitute for ethical reasoning itself, which must always remain sensitive to contextual details that cannot be captured in general Guidelines. Beyond developing a set of rules, ensuring Trustworthy AI requires us to build and maintain an ethical culture and mind-set through public debate, education and practical learning.

1. Fundamental rights as moral and legal entitlements

We believe in an approach to AI ethics based on the fundamental rights enshrined in the EU Treaties,¹⁵ the EU Charter and international human rights law.¹⁶ Respect for fundamental rights, within a framework of democracy and the rule of law, provides the most promising foundations for identifying abstract ethical principles and values, which can be operationalised in the context of AI.

The EU Treaties and the EU Charter prescribe a series of fundamental rights that EU member states and EU institutions are legally obliged to respect when implementing EU law. These rights are described in the EU Charter

¹³ https://ec.europa.eu/commission/publications/reflection-paper-towards-sustainable-europe-2030_en

¹⁴ <https://sustainabledevelopment.un.org/?menu=1300>

¹⁵ The EU is based on a constitutional commitment to protect the fundamental and indivisible rights of human beings, to ensure respect for the rule of law, to foster democratic freedom and promote the common good. These rights are reflected in Articles 2 and 3 of the Treaty on European Union, and in the Charter of Fundamental Rights of the EU.

¹⁶ Other legal instruments reflect and provide further specification of these commitments, such as for instance the Council of Europe's European Social Charter or specific legislation such as the EU's General Data Protection Regulation.

by reference to dignity, freedoms, equality and solidarity, citizens' rights and justice. The common foundation that unites these rights can be understood as rooted in respect for human dignity – thereby reflecting what we describe as a “human-centric approach” in which the human being enjoys a unique and inalienable moral status of primacy in the civil, political, economic and social fields.¹⁷

While the rights set out in the EU Charter are legally binding,¹⁸ it is important to recognise that fundamental rights do not provide comprehensive legal protection in every case. For the EU Charter, for instance, it is important to underline that its field of application is limited to areas of EU law. International human rights law and in particular the European Convention on Human Rights are legally binding on EU Member States, including in areas that fall outside the scope of EU law. At the same time, fundamental rights are also bestowed on individuals and (to a certain degree) groups by virtue of their moral status as human beings, independently of their legal force. Understood as legally enforceable rights, fundamental rights therefore fall under the first component of Trustworthy AI (lawful AI), which safeguards compliance with the law. Understood as the rights of everyone, rooted in the inherent moral status of human beings, they also underpin the second component of Trustworthy AI (ethical AI), dealing with ethical norms that are not necessarily legally binding yet crucial to ensure trustworthiness. Since this document does not aim to offer guidance on the former component, for the purpose of these non-binding guidelines, references to fundamental rights reflect the latter component.

2. From fundamental rights to ethical principles

2.1 Fundamental rights as a basis for Trustworthy AI

Among the comprehensive set of indivisible rights set out in international human rights law, the EU Treaties and the EU Charter, the below families of fundamental rights are particularly apt to cover AI systems. Many of these rights are, in specified circumstances, legally enforceable in the EU so that compliance with their terms is legally obligatory. But even after compliance with legally enforceable fundamental rights has been achieved, ethical reflection can help us understand how the development, deployment and use of AI systems may implicate fundamental rights and their underlying values, and can help provide more fine-grained guidance when seeking to identify what we *should* do rather than what we (currently) *can* do with technology.

Respect for human dignity. Human dignity encompasses the idea that every human being possesses an “intrinsic worth”, which should never be diminished, compromised or repressed by others – nor by new technologies like AI systems.¹⁹ In this context, respect for human dignity entails that all people are treated with respect due to them as moral *subjects*, rather than merely as *objects* to be sifted, sorted, scored, herded, conditioned or manipulated. AI systems should hence be developed in a manner that respects, serves and protects humans' physical and mental integrity, personal and cultural sense of identity, and satisfaction of their essential needs.²⁰

Freedom of the individual. Human beings should remain free to make life decisions for themselves. This entails freedom from sovereign intrusion, but also requires intervention from government and non-governmental organisations to ensure that individuals or people at risk of exclusion have equal access to AI's benefits and opportunities. In an AI context, freedom of the individual for instance requires mitigation of (in)direct illegitimate coercion, threats to mental autonomy and mental health, unjustified surveillance, deception and unfair manipulation. In fact, freedom of the individual means a commitment to enabling individuals to wield even higher control over their lives, including (among other rights) protection of the freedom to conduct a business, the freedom of the arts and science, freedom of expression, the right to private life and privacy, and freedom of

¹⁷ It should be noted that a commitment to human-centric AI and its anchoring in fundamental rights requires collective societal and constitutional foundations in which individual freedom and respect for human dignity is both practically possible and meaningful, rather than implying an unduly individualistic account of the human.

¹⁸ Pursuant to Article 51 of the Charter, it applies to EU Institutions and to EU member states when implementing EU law.

¹⁹ C. McCrudden, Human Dignity and Judicial Interpretation of Human Rights, *EJIL*, 19(4), 2008.

²⁰ For an understanding of “human dignity” along these lines see E. Hilgendorf, Problem Areas in the Dignity Debate and the Ensemble Theory of Human Dignity, in: D. Grimm, A. Kemmerer, C. Möllers (eds.), *Human Dignity in Context. Explorations of a Contested Concept*, 2018, pp. 325 ff.

assembly and association.

Respect for democracy, justice and the rule of law. All governmental power in constitutional democracies must be legally authorised and limited by law. AI systems should serve to maintain and foster democratic processes and respect the plurality of values and life choices of individuals. AI systems must not undermine democratic processes, human deliberation or democratic voting systems. AI systems must also embed a commitment to ensure that they do not operate in ways that undermine the foundational commitments upon which the rule of law is founded, mandatory laws and regulation, and to ensure due process and equality before the law.

Equality, non-discrimination and solidarity - including the rights of persons at risk of exclusion. Equal respect for the moral worth and dignity of all human beings must be ensured. This goes beyond non-discrimination, which tolerates the drawing of distinctions between dissimilar situations based on objective justifications. In an AI context, equality entails that the system's operations cannot generate unfairly biased outputs (e.g. the data used to train AI systems should be as inclusive as possible, representing different population groups). This also requires adequate respect for potentially vulnerable persons and groups,²¹ such as workers, women, persons with disabilities, ethnic minorities, children, consumers or others at risk of exclusion.

Citizens' rights. Citizens benefit from a wide array of rights, including the right to vote, the right to good administration or access to public documents, and the right to petition the administration. AI systems offer substantial potential to improve the scale and efficiency of government in the provision of public goods and services to society. At the same time, citizens' rights could also be negatively impacted by AI systems and should be safeguarded. When the term "citizens' rights" is used here, this is not to deny or neglect the rights of third-country nationals and irregular (or illegal) persons in the EU who also have rights under international law, and – therefore – in the area of AI systems.

2.2 Ethical Principles in the Context of AI Systems²²

Many public, private, and civil organizations have drawn inspiration from fundamental rights to produce ethical frameworks for AI systems.²³ In the EU, the European Group on Ethics in Science and New Technologies ("EGE") proposed a set of 9 basic principles, based on the fundamental values laid down in the EU Treaties and Charter.²⁴ We build further on this work, recognising most of the principles hitherto propounded by various groups, while clarifying the ends that all principles seek to nurture and support. These ethical principles can inspire new and specific regulatory instruments, can help interpreting fundamental rights as our socio-technical environment evolves over time, and can guide the rationale for AI systems' development, deployment and use – adapting dynamically as society itself evolves.

AI systems should improve individual and collective wellbeing. This section lists **four ethical principles**, rooted in fundamental rights, which must be respected in order to ensure that AI systems are developed, deployed and used in a trustworthy manner. They are specified as **ethical imperatives**, such that AI practitioners should always strive to adhere to them. Without imposing a hierarchy, we list the principles here below in manner that mirrors the order of appearance of the fundamental rights upon which they are based in the EU Charter.²⁵

²¹ For a description of the term as used throughout this document, see the Glossary.

²² These principles also apply to the development, deployment and use of other technologies, and hence are not specific to AI systems. In what follows, we have aimed to set out their relevance specifically in an AI-related context.

²³ Reliance on fundamental rights also helps to limit regulatory uncertainty as it can build on the basis of decades of practice of fundamental rights protection in the EU, thereby offering clarity, readability and foreseeability.

²⁴ More recently, the AI4People's taskforce has surveyed the aforementioned EGE principles as well as 36 other ethical principles put forward to date and subsumed them under four overarching principles: L. Floridi, J. COWLS, M. Beltrametti, R. Chatila, P. Chazerand, V. Dignum, C. Luetge, R. Madelin, U. Pagallo, F. Rossi, B. Schafer, P. Valcke, E. J. M. Vayena (2018), "AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations", *Minds and Machines* 28(4): 689-707.

²⁵ Respect for human autonomy is strongly associated with the right to human dignity and liberty (reflected in Articles 1 and 6 of the Charter). The prevention of harm is strongly linked to the protection of physical or mental integrity (reflected in Article 3).

These are the principles of:

- (i) Respect for human autonomy
- (ii) Prevention of harm
- (iii) Fairness
- (iv) Explicability

Many of these are to a large extent already reflected in existing legal requirements for which mandatory compliance is required and hence also fall within the scope of lawful AI, which is Trustworthy AI's first component.²⁶ Yet, as set out above, while many legal obligations reflect ethical principles, adherence to ethical principles goes beyond formal compliance with existing laws.²⁷

- *The principle of respect for human autonomy*

The fundamental rights upon which the EU is founded are directed towards ensuring respect for the freedom and autonomy of human beings. Humans interacting with AI systems must be able to keep full and effective self-determination over themselves, and be able to partake in the democratic process. AI systems should not unjustifiably subordinate, coerce, deceive, manipulate, condition or herd humans. Instead, they should be designed to augment, complement and empower human cognitive, social and cultural skills. The allocation of functions between humans and AI systems should follow human-centric design principles and leave meaningful opportunity for human choice. This means securing human oversight²⁸ over work processes in AI systems. AI systems may also fundamentally change the work sphere. It should support humans in the working environment, and aim for the creation of meaningful work.

- *The principle of prevention of harm*

AI systems should neither cause nor exacerbate harm²⁹ or otherwise adversely affect human beings.³⁰ This entails the protection of human dignity as well as mental and physical integrity. AI systems and the environments in which they operate must be safe and secure. They must be technically robust and it should be ensured that they are not open to malicious use. Vulnerable persons should receive greater attention and be included in the development, deployment and use of AI systems. Particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information, such as between employers and employees, businesses and consumers or governments and citizens. Preventing harm also entails consideration of the natural environment and all living beings.

- *The principle of fairness*

The development, deployment and use of AI systems must be fair. While we acknowledge that there are many different interpretations of fairness, we believe that fairness has both a substantive and a procedural dimension. The substantive dimension implies a commitment to: ensuring equal and just distribution of both benefits and costs, and ensuring that individuals and groups are free from unfair bias, discrimination and stigmatisation. If unfair biases can be avoided, AI systems could even increase societal fairness. Equal opportunity in terms of access to education, goods, services and technology should also be fostered. Moreover, the use of AI systems should never lead to people being deceived or unjustifiably impaired in their freedom of choice. Additionally, fairness implies that AI practitioners should respect the principle of proportionality between means and ends, and consider carefully how to

Fairness is closely linked to the rights to Non-discrimination, Solidarity and Justice (reflected in Articles 21 and following). Explicability and Responsibility are closely linked to the rights relating to Justice (as reflected in Article 47).

²⁶ Think for instance of the GDPR or EU consumer protection regulations.

²⁷ For further reading on this subject, see for instance L. Floridi, Soft Ethics and the Governance of the Digital, *Philosophy & Technology*, March 2018, Volume 31, Issue 1, pp 1–8.

²⁸ The concept of human oversight is further developed as one of the key requirements set out in Chapter II here below.

²⁹ Harms can be individual or collective, and can include intangible harm to social, cultural and political environments.

³⁰ This also encompasses the way of living of individuals and social groups, avoiding for instance cultural harm.

balance competing interests and objectives.³¹ The procedural dimension of fairness entails the ability to contest and seek effective redress against decisions made by AI systems and by the humans operating them.³² In order to do so, the entity accountable for the decision must be identifiable, and the decision-making processes should be explicable.

- *The principle of explicability*

Explicability is crucial for building and maintaining users' trust in AI systems. This means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected. Without such information, a decision cannot be duly contested. An explanation as to why a model has generated a particular output or decision (and what combination of input factors contributed to that) is not always possible. These cases are referred to as 'black box' algorithms and require special attention. In those circumstances, other explicability measures (e.g. traceability, auditability and transparent communication on system capabilities) may be required, provided that the system as a whole respects fundamental rights. The degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate.³³

2.3 Tensions between the principles

Tensions may arise between the above principles, for which there is no fixed solution. In line with the EU fundamental commitment to democratic engagement, due process and open political participation, methods of accountable deliberation to deal with such tensions should be established. For instance, in various application domains, *the principle of prevention of harm* and *the principle of human autonomy* may be in conflict. Consider as an example the use of AI systems for 'predictive policing', which may help to reduce crime, but in ways that entail surveillance activities that impinge on individual liberty and privacy. Furthermore, AI systems' overall benefits should substantially exceed the foreseeable individual risks. While the above principles certainly offer guidance towards solutions, they remain abstract ethical prescriptions. AI practitioners can hence not be expected to find the right solution based on the principles above, yet they should approach ethical dilemmas and trade-offs via reasoned, evidence-based reflection rather than intuition or random discretion.

There may be situations, however, where no ethically acceptable trade-offs can be identified. Certain fundamental rights and correlated principles are absolute and cannot be subject to a balancing exercise (e.g. human dignity).

Key guidance derived from Chapter I:

- ✓ Develop, deploy and use AI systems in a way that adheres to the ethical principles of: *respect for human autonomy, prevention of harm, fairness and explicability*. Acknowledge and address the potential tensions between these principles.
- ✓ Pay particular attention to situations involving more vulnerable groups such as children, persons with disabilities and others that have historically been disadvantaged or are at risk of exclusion, and to situations which are characterised by asymmetries of power or information, such as between employers and workers, or between businesses and consumers.³⁴

³¹ This relates to the principle of proportionality (reflected in the maxim that one should not 'use a sledge hammer to crack a nut'). Measures taken to achieve an end (e.g. the data extraction measures implemented to realise the AI optimisation function) should be limited to what is strictly necessary. It also entails that when several measures compete for the satisfaction of an end, preference should be given to the one that is least adverse to fundamental rights and ethical norms (e.g. AI developers should always prefer public sector data to personal data). Reference can also be made to the proportionality between user and deployer, considering the rights of companies (including intellectual property and confidentiality) on the one hand, and the rights of the user on the other.

³² Including by using their right of association and to join a trade union in a working environment, as provided for by Article 12 of the EU Charter of fundamental rights.

³³ For example, little ethical concern may flow from inaccurate shopping recommendations generated by an AI system, in contrast to AI systems that evaluate whether an individual convicted of a criminal offence should be released on parole.

³⁴ See articles 24 to 27 of the EU Charter, dealing with the rights of the child and the elderly, the integration of persons with disabilities and workers' rights. See also article 38 dealing with consumer protection.

- ✓ Acknowledge that, while bringing substantial benefits to individuals and society, AI systems also pose certain risks and may have a negative impact, including impacts which may be difficult to anticipate, identify or measure (e.g. on democracy, the rule of law and distributive justice, or on the human mind itself.) Adopt adequate measures to mitigate these risks when appropriate, and proportionately to the magnitude of the risk.

II. Chapter II: Realising Trustworthy AI

This Chapter offers guidance on the implementation and realisation of Trustworthy AI, via a list of seven requirements that should be met, building on the principles outlined in Chapter I. In addition, available technical and non-technical methods are introduced for the implementation of these requirements throughout the AI system's life cycle.

1. Requirements of Trustworthy AI

The principles outlined in Chapter I must be translated into concrete requirements to achieve Trustworthy AI. These requirements are applicable to different stakeholders partaking in AI systems' life cycle: developers, deployers and end-users, as well as the broader society. By developers, we refer to those who research, design and/or develop AI systems. By deployers, we refer to public or private organisations that use AI systems within their business processes and to offer products and services to others. End-users are those engaging with the AI system, directly or indirectly. Finally, the broader society encompasses all others that are directly or indirectly affected by AI systems.

Different groups of stakeholders have different roles to play in ensuring that the requirements are met:

- a. Developers should implement and apply the requirements to design and development processes;
- b. Deployers should ensure that the systems they use and the products and services they offer meet the requirements;
- c. End-users and the broader society should be informed about these requirements and able to request that they are upheld.

The below list of requirements is non-exhaustive.³⁵ It includes systemic, individual and societal aspects:

1 Human agency and oversight

Including fundamental rights, human agency and human oversight

2 Technical robustness and safety

Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility

3 Privacy and data governance

Including respect for privacy, quality and integrity of data, and access to data

4 Transparency

Including traceability, explainability and communication

5 Diversity, non-discrimination and fairness

Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation

6 Societal and environmental wellbeing

Including sustainability and environmental friendliness, social impact, society and democracy

7 Accountability

Including auditability, minimisation and reporting of negative impact, trade-offs and redress.

³⁵ Without imposing a hierarchy, we list the principles here below in manner that mirrors the order of appearance of the principles and rights to which they relate in the EU Charter.



Figure 2: Interrelationship of the seven requirements: all are of equal importance, support each other, and should be implemented and evaluated throughout the AI system's lifecycle

While all requirements are of equal importance, context and potential tensions between them will need to be taken into account when applying them across different domains and industries. Implementation of these requirements should occur throughout an AI system's entire life cycle and depends on the specific application. While most requirements apply to all AI systems, special attention is given to those directly or indirectly affecting individuals. Therefore, for some applications (for instance in industrial settings), they may be of lesser relevance.

The above requirements include elements that are in some cases already reflected in existing laws. We reiterate that – in line with Trustworthy AI's first component – it is the responsibility of AI practitioners to ensure that they comply with their legal obligations, both as regards horizontally applicable rules as well as domain-specific regulation.

In the following paragraphs, each requirement is explained in more detail.

1.1 Human agency and oversight

AI systems should support human autonomy and decision-making, as prescribed by the principle of *respect for human autonomy*. This requires that AI systems should both act as enablers to a democratic, flourishing and equitable society by supporting the user's agency and foster fundamental rights, and allow for human oversight.

Fundamental rights. Like many technologies, AI systems can equally enable and hamper fundamental rights. They can benefit people for instance by helping them track their personal data, or by increasing the accessibility of education, hence supporting their right to education. However, given the reach and capacity of AI systems, they can also negatively affect fundamental rights. In situations where such risks exist, a fundamental rights impact assessment should be undertaken. This should be done prior to the system's development and include an evaluation of whether those risks can be reduced or justified as necessary in a democratic society in order to respect the rights and freedoms of others. Moreover, mechanisms should be put into place to receive external feedback

regarding AI systems that potentially infringe on fundamental rights.

Human agency. Users should be able to make informed autonomous decisions regarding AI systems. They should be given the knowledge and tools to comprehend and interact with AI systems to a satisfactory degree and, where possible, be enabled to reasonably self-assess or challenge the system. AI systems should support individuals in making better, more informed choices in accordance with their goals. AI systems can sometimes be deployed to shape and influence human behaviour through mechanisms that may be difficult to detect, since they may harness sub-conscious processes, including various forms of unfair manipulation, deception, herding and conditioning, all of which may threaten individual autonomy. The overall principle of user autonomy must be central to the system's functionality. Key to this is the right not to be subject to a decision based solely on automated processing when this produces legal effects on users or similarly significantly affects them.³⁶

Human oversight. Human oversight helps ensuring that an AI system does not undermine human autonomy or causes other adverse effects. Oversight may be achieved through governance mechanisms such as a human-in-the-loop (HITL), human-on-the-loop (HOTL), or human-in-command (HIC) approach. HITL refers to the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable. HOTL refers to the capability for human intervention during the design cycle of the system and monitoring the system's operation. HIC refers to the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system. Moreover, it must be ensured that public enforcers have the ability to exercise oversight in line with their mandate. Oversight mechanisms can be required in varying degrees to support other safety and control measures, depending on the AI system's application area and potential risk. All other things being equal, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance is required.

1.2 Technical robustness and safety

A crucial component of achieving Trustworthy AI is technical robustness, which is closely linked to the *principle of prevention of harm*. Technical robustness requires that AI systems be developed with a preventative approach to risks and in a manner such that they reliably behave as intended while minimising unintentional and unexpected harm, and preventing unacceptable harm. This should also apply to potential changes in their operating environment or the presence of other agents (human and artificial) that may interact with the system in an adversarial manner. In addition, the physical and mental integrity of humans should be ensured.

Resilience to attack and security. AI systems, like all software systems, should be protected against vulnerabilities that can allow them to be exploited by adversaries, e.g. hacking. Attacks may target the data (data poisoning), the model (model leakage) or the underlying infrastructure, both software and hardware. If an AI system is attacked, e.g. in adversarial attacks, the data as well as system behaviour can be changed, leading the system to make different decisions, or causing it to shut down altogether. Systems and data can also become corrupted by malicious intention or by exposure to unexpected situations. Insufficient security processes can also result in erroneous decisions or even physical harm. For AI systems to be considered secure,³⁷ possible unintended applications of the AI system (e.g. dual-use applications) and potential abuse of the system by malicious actors should be taken into account, and steps should be taken to prevent and mitigate these.³⁸

Fallback plan and general safety. AI systems should have safeguards that enable a fallback plan in case of problems.

³⁶ Reference can be made to Article 22 of the GDPR where this right is already enshrined.

³⁷ See e.g. considerations under 2.7 of the European Union's Coordinated Plan on Artificial Intelligence.

³⁸ There may be a strong imperative to develop a virtuous circle in research and development between understanding of attacks, development of adequate protection, and improvement of evaluation methodologies. To achieve this, convergence between the AI community and the security community should be promoted. In addition, it is the responsibility of all relevant actors to create common cross-border safety and security norms and to establish an environment of mutual trust, fostering international collaboration. For possible measures, see *Malicious Use of AI*, Avin S., Brundage M. et. al., 2018.

This can mean that AI systems switch from a statistical to rule-based procedure, or that they ask for a human operator before continuing their action.³⁹ It must be ensured that the system will do what it is supposed to do without harming living beings or the environment. This includes the minimisation of unintended consequences and errors. In addition, processes to clarify and assess potential risks associated with the use of AI systems, across various application areas, should be established. The level of safety measures required depends on the magnitude of the risk posed by an AI system, which in turn depends on the system's capabilities. Where it can be foreseen that the development process or the system itself will pose particularly high risks, it is crucial for safety measures to be developed and tested proactively.

Accuracy. Accuracy pertains to an AI system's ability to make correct judgements, for example to correctly classify information into the proper categories, or its ability to make correct predictions, recommendations, or decisions based on data or models. An explicit and well-formed development and evaluation process can support, mitigate and correct unintended risks from inaccurate predictions. When occasional inaccurate predictions cannot be avoided, it is important that the system can indicate how likely these errors are. A high level of accuracy is especially crucial in situations where the AI system directly affects human lives.

Reliability and Reproducibility. It is critical that the results of AI systems are reproducible, as well as reliable. A reliable AI system is one that works properly with a range of inputs and in a range of situations. This is needed to scrutinise an AI system and to prevent unintended harms. Reproducibility describes whether an AI experiment exhibits the same behaviour when repeated under the same conditions. This enables scientists and policy makers to accurately describe what AI systems do. Replication files⁴⁰ can facilitate the process of testing and reproducing behaviours.

1.3 Privacy and data governance

Closely linked to the *principle of prevention of harm* is privacy, a fundamental right particularly affected by AI systems. Prevention of harm to privacy also necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy.

Privacy and data protection. AI systems must guarantee privacy and data protection throughout a system's entire lifecycle.⁴¹ This includes the information initially provided by the user, as well as the information generated about the user over the course of their interaction with the system (e.g. outputs that the AI system generated for specific users or how users responded to particular recommendations). Digital records of human behaviour may allow AI systems to infer not only individuals' preferences, but also their sexual orientation, age, gender, religious or political views. To allow individuals to trust the data gathering process, it must be ensured that data collected about them will not be used to unlawfully or unfairly discriminate against them.

Quality and integrity of data. The quality of the data sets used is paramount to the performance of AI systems. When data is gathered, it may contain socially constructed biases, inaccuracies, errors and mistakes. This needs to be addressed prior to training with any given data set. In addition, the integrity of the data must be ensured. Feeding malicious data into an AI system may change its behaviour, particularly with self-learning systems. Processes and data sets used must be tested and documented at each step such as planning, training, testing and deployment. This should also apply to AI systems that were not developed in-house but acquired elsewhere.

Access to data. In any given organisation that handles individuals' data (whether someone is a user of the system or not), data protocols governing data access should be put in place. These protocols should outline who can access data and under which circumstances. Only duly qualified personnel with the competence and need to access individual's data should be allowed to do so.

³⁹ Scenarios where human intervention would not immediately be possible should also be considered.

⁴⁰ This concerns files that will replicate each step of the AI system's development process, from research and initial data collection to the results.

⁴¹ Reference can be made to existing privacy laws, such as the GDPR or the forthcoming ePrivacy Regulation.

1.4 Transparency

This requirement is closely linked with the *principle of explicability* and encompasses transparency of elements relevant to an AI system: the data, the system and the business models.

Traceability. The data sets and the processes that yield the AI system's decision, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard to allow for traceability and an increase in transparency. This also applies to the decisions made by the AI system. This enables identification of the reasons why an AI-decision was erroneous which, in turn, could help prevent future mistakes. Traceability facilitates auditability as well as explainability.

Explainability. Explainability concerns the ability to explain both the technical processes of an AI system and the related human decisions (e.g. application areas of a system). Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. Moreover, trade-offs might have to be made between enhancing a system's explainability (which may reduce its accuracy) or increasing its accuracy (at the cost of explainability). Whenever an AI system has a significant impact on people's lives, it should be possible to demand a suitable explanation of the AI system's decision-making process. Such explanation should be timely and adapted to the expertise of the stakeholder concerned (e.g. layperson, regulator or researcher). In addition, explanations of the degree to which an AI system influences and shapes the organisational decision-making process, design choices of the system, and the rationale for deploying it, should be available (hence ensuring business model transparency).

Communication. AI systems should not represent themselves as humans to users; humans have the right to be informed that they are interacting with an AI system. This entails that AI systems must be identifiable as such. In addition, the option to decide against this interaction in favour of human interaction should be provided where needed to ensure compliance with fundamental rights. Beyond this, the AI system's capabilities and limitations should be communicated to AI practitioners or end-users in a manner appropriate to the use case at hand. This could encompass communication of the AI system's level of accuracy, as well as its limitations.

1.5 Diversity, non-discrimination and fairness

In order to achieve Trustworthy AI, we must enable inclusion and diversity throughout the entire AI system's life cycle. Besides the consideration and involvement of all affected stakeholders throughout the process, this also entails ensuring equal access through inclusive design processes as well as equal treatment. This requirement is closely linked with *the principle of fairness*.

Avoidance of unfair bias. Data sets used by AI systems (both for training and operation) may suffer from the inclusion of inadvertent historic bias, incompleteness and bad governance models. The continuation of such biases could lead to unintended (in)direct prejudice and discrimination⁴² against certain groups or people, potentially exacerbating prejudice and marginalisation. Harm can also result from the intentional exploitation of (consumer) biases or by engaging in unfair competition, such as the homogenisation of prices by means of collusion or a non-transparent market.⁴³ Identifiable and discriminatory bias should be removed in the collection phase where possible. The way in which AI systems are developed (e.g. algorithms' programming) may also suffer from unfair bias. This could be counteracted by putting in place oversight processes to analyse and address the system's purpose, constraints, requirements and decisions in a clear and transparent manner. Moreover, hiring from diverse backgrounds, cultures and disciplines can ensure diversity of opinions and should be encouraged.

Accessibility and universal design. Particularly in business-to-consumer domains, systems should be user-centric and designed in a way that allows all people to use AI products or services, regardless of their age, gender, abilities or characteristics. Accessibility to this technology for persons with disabilities, which are present in all societal

⁴² For a definition of direct and indirect discrimination, see for instance Article 2 of Council Directive 2000/78/EC of 27 November 2000 establishing a general framework for equal treatment in employment and occupation. See also Article 21 of the Charter of Fundamental Rights of the EU.

⁴³ See the EU Agency for Fundamental Rights' paper: "BigData: Discrimination in data-supported decision making", 2018, <http://fra.europa.eu/en/publication/2018/big-data-discrimination>.

groups, is of particular importance. AI systems should not have a one-size-fits-all approach and should consider Universal Design⁴⁴ principles addressing the widest possible range of users, following relevant accessibility standards.⁴⁵ This will enable equitable access and active participation of all people in existing and emerging computer-mediated human activities and with regard to assistive technologies.⁴⁶

Stakeholder Participation. In order to develop AI systems that are trustworthy, it is advisable to consult stakeholders who may directly or indirectly be affected by the system throughout its life cycle. It is beneficial to solicit regular feedback even after deployment and set up longer term mechanisms for stakeholder participation, for example by ensuring workers information, consultation and participation throughout the whole process of implementing AI systems at organisations.

1.6 Societal and environmental well-being

In line with the *principles of fairness* and *prevention of harm*, the broader society, other sentient beings and the environment should be also considered as stakeholders throughout the AI system's life cycle. Sustainability and ecological responsibility of AI systems should be encouraged, and research should be fostered into AI solutions addressing areas of global concern, such as for instance the Sustainable Development Goals. Ideally, AI systems should be used to benefit all human beings, including future generations.

Sustainable and environmentally friendly AI. AI systems promise to help tackling some of the most pressing societal concerns, yet it must be ensured that this occurs in the most environmentally friendly way possible. The system's development, deployment and use process, as well as its entire supply chain, should be assessed in this regard, e.g. via a critical examination of the resource usage and energy consumption during training, opting for less harmful choices. Measures securing the environmental friendliness of AI systems' entire supply chain should be encouraged.

Social impact. Ubiquitous exposure to social AI systems⁴⁷ in all areas of our lives (be it in education, work, care or entertainment) may alter our conception of social agency, or impact our social relationships and attachment. While AI systems can be used to enhance social skills,⁴⁸ they can equally contribute to their deterioration. This could also affect people's physical and mental wellbeing. The effects of these systems must therefore be carefully monitored and considered.

Society and Democracy. Beyond assessing the impact of an AI system's development, deployment and use on individuals, this impact should also be assessed from a societal perspective, taking into account its effect on institutions, democracy and society at large. The use of AI systems should be given careful consideration particularly in situations relating to the democratic process, including not only political decision-making but also electoral contexts.

1.7 Accountability

The requirement of accountability complements the above requirements, and is closely linked to the *principle of fairness*. It necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use.

Auditability. Auditability entails the enablement of the assessment of algorithms, data and design processes. This does not necessarily imply that information about business models and intellectual property related to the AI

⁴⁴ Article 42 of the Public Procurement Directive requires technical specifications to consider accessibility and 'design for all'.

⁴⁵ For instance EN 301 549.

⁴⁶ This requirement links to the United Nations Convention on the Rights of Persons with Disabilities.

⁴⁷ This denotes AI systems communicating and interacting with humans by simulating sociality in human robot interaction (embodied AI) or as avatars in virtual reality. By doing so, those systems have the potential to change our socio-cultural practices and the fabric of our social life.

⁴⁸ See for instance the EU-funded project developing AI-based software that enables robots to interact more effectively with autistic children in human-led therapy sessions, helping to improve their social and communication skills: http://ec.europa.eu/research/infocentre/article_en.cfm?id=/research/headlines/news/article_19_03_12_en.html?infocentre&item=Infocentre&artid=49968

system must always be openly available. Evaluation by internal and external auditors, and the availability of such evaluation reports, can contribute to the trustworthiness of the technology. In applications affecting fundamental rights, including safety-critical applications, AI systems should be able to be independently audited.

Minimisation and reporting of negative impacts. Both the ability to report on actions or decisions that contribute to a certain system outcome, and to respond to the consequences of such an outcome, must be ensured. Identifying, assessing, documenting and minimising the potential negative impacts of AI systems is especially crucial for those (in)directly affected. Due protection must be available for whistle-blowers, NGOs, trade unions or other entities when reporting legitimate concerns about an AI system. The use of impact assessments (e.g. red teaming or forms of Algorithmic Impact Assessment) both prior to and during the development, deployment and use of AI systems can be helpful to minimise negative impact. These assessments must be proportionate to the risk that the AI systems pose.

Trade-offs. When implementing the above requirements, tensions may arise between them, which may lead to inevitable trade-offs. Such trade-offs should be addressed in a rational and methodological manner within the state of the art. This entails that relevant interests and values implicated by the AI system should be identified and that, if conflict arises, trade-offs should be explicitly acknowledged and evaluated in terms of their risk to ethical principles, including fundamental rights. In situations in which no ethically acceptable trade-offs can be identified, the development, deployment and use of the AI system should not proceed in that form. Any decision about which trade-off to make should be reasoned and properly documented. The decision-maker must be accountable for the manner in which the appropriate trade-off is being made, and should continually review the appropriateness of the resulting decision to ensure that necessary changes can be made to the system where needed.⁴⁹

Redress. When unjust adverse impact occurs, accessible mechanisms should be foreseen that ensure adequate redress.⁵⁰ Knowing that redress is possible when things go wrong is key to ensure trust. Particular attention should be paid to vulnerable persons or groups.

2. Technical and non-technical methods to realise Trustworthy AI

To implement the above requirements, both technical and non-technical methods can be employed. These encompass all stages of an AI system's life cycle. An evaluation of the methods employed to implement the requirements, as well as reporting and justifying⁵¹ changes to the implementation processes, should occur on an ongoing basis. AI systems are continuously evolving and acting in a dynamic environment. The realisation of Trustworthy AI is therefore a continuous process, as depicted in Figure 3 here below.

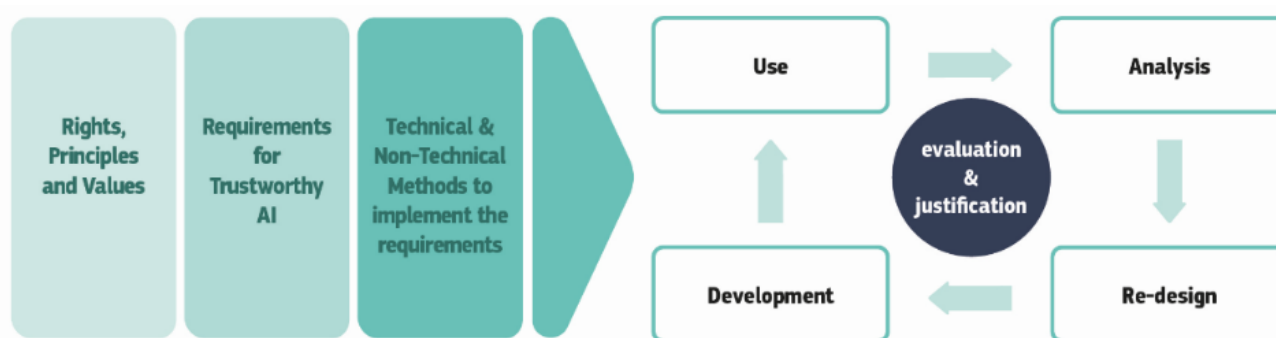


Figure 3: Realising Trustworthy AI throughout the system's entire life cycle

⁴⁹ Different governance models can help achieving this. E.g. the presence of an internal and/or external ethical (and sector specific) expert or board might be useful to highlight areas of potential conflict and suggest ways in which that conflict might best be resolved. Meaningful consultation and discussion with stakeholders, including those at risk of being adversely affected by an AI system is useful too. European universities should take a leading role in training the ethics experts needed.

⁵⁰ See also the European Union Agency for Fundamental Rights' Opinion on 'Improving access to remedy in the area of business and human rights at the EU level', 2017, <https://fra.europa.eu/en/opinion/2017/business-human-rights>.

⁵¹ This entails e.g. justification of the choices in the system's design, development and deployment to implement the requirements.

The following methods can be either complementary or alternative to each other, since different requirements – and different sensitivities – may raise the need for different methods of implementation. This overview is neither meant to be comprehensive or exhaustive, nor mandatory. Rather, its aim is to offer a list of suggested methods that may help to implement Trustworthy AI.

2.1. Technical methods

This section describes technical methods to ensure Trustworthy AI that can be incorporated in the design, development and use phases of an AI system. The methods listed below vary in level of maturity.⁵²

- *Architectures for Trustworthy AI*

Requirements for Trustworthy AI should be “translated” into procedures and/or constraints on procedures, which should be anchored in the AI system’s architecture. This could be accomplished through a set of “white list” rules (behaviours or states) that the system should always follow, “black list” restrictions on behaviours or states that the system should never transgress, and mixtures of those or more complex provable guarantees regarding the system’s behaviour. Monitoring of the system’s compliance with these restrictions during operations may be achieved by a separate process.

AI systems with learning capabilities that can dynamically adapt their behaviour can be understood as non-deterministic systems possibly exhibiting unexpected behaviour. These are often considered through the theoretical lens of a “sense-plan-act” cycle. Adapting this architecture to ensure Trustworthy AI requires the requirements’ integration at all three steps of the cycle: (i) at the “sense”-step, the system should be developed such that it recognises all environmental elements necessary to ensure adherence to the requirements; (ii) at the “plan”-step, the system should only consider plans that adhere to the requirements; (iii) at the “act”-step, the system’s actions should be restricted to behaviours that realise the requirements.

The architecture as sketched above is generic and only provides an imperfect description for most AI systems. Nevertheless, it gives anchor points for constraints and policies that should be reflected in specific modules to result in an overall system that is trustworthy and perceived as such.

- *Ethics and rule of law by design (X-by-design)*

Methods to ensure values-by-design provide precise and explicit links between the abstract principles which the system is required to respect and the specific implementation decisions. The idea that compliance with norms can be implemented into the design of the AI system is key to this method. Companies are responsible for identifying the impact of their AI systems from the very start, as well as the norms their AI system ought to comply with to avert negative impacts. Different “by-design” concepts are already widely used, e.g. *privacy-by-design* and *security-by-design*. As indicated above, to earn trust AI needs to be secure in its processes, data and outcomes, and should be designed to be robust to adversarial data and attacks. It should implement a mechanism for fail-safe shutdown and enable resumed operation after a forced shut-down (such as an attack).

- *Explanation methods*

For a system to be trustworthy, we must be able to understand why it behaved a certain way and why it provided a given interpretation. A whole field of research, Explainable AI (XAI) tries to address this issue to better understand the system’s underlying mechanisms and find solutions. Today, this is still an open challenge for AI systems based on neural networks. Training processes with neural nets can result in network parameters set to numerical values that are difficult to correlate with results. Moreover, sometimes small changes in data values might result in dramatic changes in interpretation, leading the system to e.g. confuse a school bus with an ostrich. This vulnerability can also be exploited during attacks on the system. Methods involving XAI research are vital not only to explain the system’s

⁵² While some of these methods are already available today, others still require more research. Those areas where further research is needed will also inform the AI HLEG’s second deliverable, i.e. the Policy and Investment Recommendations.

behaviour to users, but also to deploy reliable technology.

- *Testing and validating*

Due to the non-deterministic and context-specific nature of AI systems, traditional testing is not enough. Failures of the concepts and representations used by the system may only manifest when a programme is applied to sufficiently realistic data. Consequently, to verify and validate processing of data, the underlying model must be carefully monitored during both training and deployment for its stability, robustness and operation within well-understood and predictable bounds. It must be ensured that the outcome of the planning process is consistent with the input, and that the decisions are made in a way allowing validation of the underlying process.

Testing and validation of the system should occur as early as possible, ensuring that the system behaves as intended throughout its entire life cycle and especially after deployment. It should include all components of an AI system, including data, pre-trained models, environments and the behaviour of the system as a whole. The testing processes should be designed and performed by an as diverse group of people as possible. Multiple metrics should be developed to cover the categories that are being tested for different perspectives. Adversarial testing by trusted and diverse “red teams” deliberately attempting to “break” the system to find vulnerabilities, and “bug bounties” that incentivise outsiders to detect and responsibly report system errors and weaknesses, can be considered. Finally, it must be ensured that the outputs or actions are consistent with the results of the preceding processes, comparing them to the previously defined policies to ensure that they are not violated.

- *Quality of Service Indicators*

Appropriate quality of service indicators can be defined for AI systems to ensure that there is a baseline understanding as to whether they have been tested and developed with security and safety considerations in mind. These indicators could include measures to evaluate the testing and training of algorithms as well as traditional software metrics of functionality, performance, usability, reliability, security and maintainability.

2.2. Non-technical methods

This section describes a variety of non-technical methods that can serve a valuable role in securing and maintaining Trustworthy AI. These too should be evaluated on an **ongoing basis**.

- *Regulation*

As mentioned above, regulation to support AI’s trustworthiness already exists today – think of product safety legislation and liability frameworks. To the extent we consider that regulation may need to be revised, adapted or introduced, both as a safeguard and as an enabler, this will be raised in our second deliverable, consisting of AI Policy and Investment Recommendations.

- *Codes of conduct*

Organisations and stakeholders can sign up to the Guidelines and adapt their charter of corporate responsibility, Key Performance Indicators (“KPIs”), their codes of conduct or internal policy documents to add the striving towards Trustworthy AI. An organisation working on or with AI systems can, more generally, document its intentions, as well as underwrite them with standards of certain desirable values such as fundamental rights, transparency and the avoidance of harm.

- *Standardisation*

Standards, for example for design, manufacturing and business practices, can function as a quality management system for AI users, consumers, organisations, research institutions and governments by offering the ability to recognise and encourage ethical conduct through their purchasing decisions. Beyond conventional standards, co-regulatory approaches exist: accreditation systems, professional codes of ethics or standards for fundamental rights compliant design. Current examples are e.g. ISO Standards or the IEEE P7000 standards series, but in the future a possible ‘Trustworthy AI’ label might be suitable, confirming by reference to specific technical standards that the system, for instance, adheres to safety, technical robustness and transparency.

- *Certification*

As it cannot be expected that everyone is able to fully understand the workings and effects of AI systems, consideration can be given to organisations that can attest to the broader public that an AI system is transparent, accountable and fair.⁵³ These certifications would apply standards developed for different application domains and AI techniques, appropriately aligned with the industrial and societal standards of different contexts. Certification can however never replace responsibility. It should hence be complemented by accountability frameworks, including disclaimers as well as review and redress mechanisms.⁵⁴

- *Accountability via governance frameworks*

Organisations should set up governance frameworks, both internal and external, ensuring accountability for the ethical dimensions of decisions associated with the development, deployment and use of AI systems. This can, for instance, include the appointment of a person in charge of ethics issues relating to AI systems, or an internal/external ethics panel or board. Amongst the possible roles of such a person, panel or board, is to provide oversight and advice. As set out above, certification specifications and bodies can also play a role to this end. Communication channels should be ensured with industry and/or public oversight groups, sharing best practices, discussing dilemmas or reporting emerging issues of ethical concerns. Such mechanisms can complement but cannot replace legal oversight (e.g. in the form of the appointment of a data protection officer or equivalent measures, legally required under data protection law).

- *Education and awareness to foster an ethical mind-set*

Trustworthy AI encourages the informed participation of all stakeholders. Communication, education and training play an important role, both to ensure that knowledge of the potential impact of AI systems is widespread, and to make people aware that they can participate in shaping the societal development. This includes all stakeholders, e.g. those involved in making the products (the designers and developers), the users (companies or individuals) and other impacted groups (those who may not purchase or use an AI system but for whom decisions are made by an AI system, and society at large). Basic AI literacy should be fostered across society. A prerequisite for educating the public is to ensure the proper skills and training of ethicists in this space.

- *Stakeholder participation and social dialogue*

The benefits of AI systems are many, and Europe needs to ensure that they are available to all. This requires an open discussion and the involvement of social partners and stakeholders, including the general public. Many organisations already rely on stakeholder panels to discuss the use of AI systems and data analytics. These panels include various members, such as legal experts, technical experts, ethicists, consumer representatives and workers. Actively seeking participation and dialogue on the use and impact of AI systems supports the evaluation of results and approaches, and can particularly be helpful in complex cases.

- *Diversity and inclusive design teams*

Diversity and inclusion play an essential role when developing AI systems that will be employed in the real world. It is critical that, as AI systems perform more tasks on their own, the teams that design, develop, test and maintain, deploy and procure these systems reflect the diversity of users and of society in general. This contributes to objectivity and consideration of different perspectives, needs and objectives. Ideally, teams are not only diverse in terms of gender, culture, age, but also in terms of professional backgrounds and skill sets.

⁵³ As advocated by e.g. the IEEE Ethically Aligned Design Initiative: <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>.

⁵⁴ For more on the limitations of certification, see: https://ainowinstitute.org/AI_Now_2018_Report.pdf.

Key guidance derived from Chapter II:

- ✓ Ensure that the AI system's entire life cycle meets the seven key requirements for Trustworthy AI: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, non-discrimination and fairness, (6) environmental and societal well-being and (7) accountability.
- ✓ Consider technical and non-technical methods to ensure the implementation of those requirements.
- ✓ Foster research and innovation to help assessing AI systems and to further the achievement of the requirements; disseminate results and open questions to the wider public, and systematically train a new generation of experts in AI ethics.
- ✓ Communicate, in a clear and proactive manner, information to stakeholders about the AI system's capabilities and limitations, enabling realistic expectation setting, and about the manner in which the requirements are implemented. Be transparent about the fact that they are dealing with an AI system.
- ✓ Facilitate the traceability and auditability of AI systems, particularly in critical contexts and situations.
- ✓ Involve stakeholders throughout the AI system's life cycle. Foster training and education so that all stakeholders are aware of and trained in Trustworthy AI.
- ✓ Be mindful that there might be fundamental tensions between different principles and requirements. Continuously identify, evaluate, document and communicate these trade-offs and their solutions.

III. Chapter III: Assessing Trustworthy AI

Based on the key requirements of Chapter II, this Chapter sets out a non-exhaustive **Trustworthy AI assessment list** (pilot version) to **operationalise Trustworthy AI**. It particularly applies to AI systems that directly interact with users, and is primarily addressed to developers and deployers of AI systems (whether self-developed or acquired from third parties). This assessment list does not address the operationalisation of the first component of Trustworthy AI (lawful AI). Compliance with this assessment list is not evidence of legal compliance, nor is it intended as guidance to ensure compliance with applicable law. Given the application-specificity of AI systems, the assessment list will need to be tailored to the specific use case and context in which the system operates. In addition, this chapter offers a general recommendation on how to implement the assessment list for Trustworthy AI through a governance structure embracing both operational and management level.

The assessment list and governance structure will be developed in close collaboration with stakeholders across the public and private sector. The process will be driven as a piloting process, allowing for extensive feedback from two parallel processes:

- a) a qualitative process, ensuring representability, where a small selection of companies, organisations and institutions (from different sectors and of different sizes) will sign up to pilot the assessment list and the governance structure in practice and to provide in-depth feedback;
- b) a quantitative process where all interested stakeholders can sign up to pilot the assessment list and provide feedback through an open consultation.

After the piloting phase, we will integrate the results from the feedback process into the assessment list and prepare a revised version in early 2020. The aim is to achieve a framework that can be horizontally used across all applications and hence offer a foundation for ensuring Trustworthy AI in all domains. Once such foundation has been established, a sectorial or application-specific framework could be developed.

▪ *Governance*

Stakeholders may wish to consider how the Trustworthy AI assessment list can be implemented in their organisation. This can be done by incorporating the assessment process into existing governance mechanisms, or by implementing new processes. This choice will depend on the internal structure of the organisation as well as its size and available resources.

Research demonstrates that management attention at the highest level is essential to achieve change.⁵⁵ It also demonstrates that involving all stakeholders in a company, organisation or institution fosters the acceptance and the relevance of the introduction of any new process (whether or not technological).⁵⁶ Therefore, we recommend implementing a process that embraces both the involvement of operational level as well as top management level.

Level	Relevant roles (depending on the organisation)
Management and Board	Top management discusses and evaluates the AI systems' development, deployment or procurement and serves as an escalation board for evaluating all AI innovations and uses, when critical concerns are detected. It involves those impacted by the possible introduction of AI systems (e.g. workers) and their representatives throughout the process via information, consultation and participation procedures.
Compliance/Legal department/Corporate responsibility department	The responsibility department monitors the use of the assessment list and its necessary evolution to meet the technological or regulatory changes. It updates the standards or internal policies on AI systems and ensures that the use of such systems complies with the current legal and regulatory framework and to the values of the organisation.
Product and Service Development or equivalent	The Product and Service Development department uses the assessment list to evaluate AI-based products and services and logs all the results. These results are discussed at management level, which ultimately approves the new or revised AI-based applications.
Quality Assurance	The Quality Assurance department (or equivalent) ensures and checks the results of the assessment list and takes action to escalate an issue higher up if the result is not satisfactory or if unforeseen results are detected.
HR	The HR department ensures the right mix of competences and diversity of profiles for developers of AI systems. It ensures that the appropriate level of training is delivered on Trustworthy AI inside the organisation.
Procurement	The procurement department ensures that the process to procure AI-based products or services includes a check of Trustworthy AI.
Day-to-day Operations	Developers and project managers include the assessment list in their daily work and document the results and outcomes of the assessment.

▪ *Using the Trustworthy AI assessment list*

When using the assessment list in practice, we recommend paying attention not only to the areas of concern but also to the questions that cannot be (easily) answered. One potential problem might be the lack of diversity of skills and competences in the team developing and testing the AI system, and therefore it might be necessary to involve other stakeholders inside or outside the organisation. It is strongly recommended to log all results both in technical terms and in management terms, ensuring that the problem solving can be understood at all levels in the governance structure.

This assessment list is meant to guide AI practitioners to achieve Trustworthy AI. The assessment should be tailored to the specific use case in a proportionate way. During the piloting phase, specific sensitive areas might be revealed and the need for further specifications in such cases will be evaluated in the next steps. While this

⁵⁵ <https://www.mckinsey.com/business-functions/operations/our-insights/secrets-of-successful-change-implementation>

⁵⁶ See for instance A. Bryson, E. Barth and H. Dale-Olsen, The Effects of Organisational change on worker well-being and the moderating role of trade unions, *ILRRReview*, 66(4), July 2013; Jirjahn, U. and Smith, S.C. (2006). 'What Factors Lead Management to Support or Oppose Employee Participation—With and Without Works Councils? Hypotheses and Evidence from Germany's Industrial Relations, 45(4), 650–680; Michie, J. and Sheehan, M. (2003). 'Labour market deregulation, "flexibility" and innovation', *Cambridge Journal of Economics*, 27(1), 123–143.

assessment list does not provide concrete answers to address the raised questions, it encourages reflection on how Trustworthy AI can be operationalised, and on the potential steps that should be taken in this regard.

- *Relation to existing law and processes*

It is also important for AI practitioners to recognise that there are various existing laws mandating particular processes or prohibiting particular outcomes, which may overlap and coincide with some of the measures listed in the assessment list. For example, data protection law sets out a series of legal requirements that must be met by those engaged in the collection and processing of personal data. Yet, because Trustworthy AI also requires the ethical handling of data, internal procedures and policies aimed at securing compliance with data protection laws might also help to facilitate ethical data handling and can hence complement existing legal processes. Compliance with this assessment list is *not*, however, evidence of legal compliance, nor is it intended as guidance to ensure compliance with applicable laws.

Moreover, many AI practitioners already have existing assessment tools and software development processes in place to ensure compliance also with non-legal standards. The below assessment should not necessarily be carried out as a stand-alone exercise, but can be incorporated into such existing practices.

TRUSTWORTHY AI ASSESSMENT LIST (PILOT VERSION)

1. Human agency and oversight

Fundamental rights:

- ✓ Did you carry out a fundamental rights impact assessment where there could be a negative impact on fundamental rights? Did you identify and document potential trade-offs made between the different principles and rights?
- ✓ Does the AI system interact with decisions by human (end) users (e.g. recommended actions or decisions to take, presenting of options)?
 - Could the AI system affect human autonomy by interfering with the (end) user's decision-making process in an unintended way?
 - Did you consider whether the AI system should communicate to (end) users that a decision, content, advice or outcome is the result of an algorithmic decision?
 - In case of a chat bot or other conversational system, are the human end users made aware that they are interacting with a non-human agent?

Human agency:

- ✓ Is the AI system implemented in work and labour process? If so, did you consider the task allocation between the AI system and humans for meaningful interactions and appropriate human oversight and control?
 - Does the AI system enhance or augment human capabilities?
 - Did you take safeguards to prevent overconfidence in or overreliance on the AI system for work processes?

Human oversight:

- ✓ Did you consider the appropriate level of human control for the particular AI system and use case?
 - Can you describe the level of human control or involvement?
 - Who is the "human in control" and what are the moments or tools for human intervention?
 - Did you put in place mechanisms and measures to ensure human control or oversight?
 - Did you take any measures to enable audit and to remedy issues related to governing AI autonomy?
- ✓ Is there is a self-learning or autonomous AI system or use case? If so, did you put in place more specific mechanisms of control and oversight?
 - Which detection and response mechanisms did you establish to assess whether something could go wrong?

- Did you ensure a stop button or procedure to safely abort an operation where needed? Does this procedure abort the process entirely, in part, or delegate control to a human?

2. Technical robustness and safety

Resilience to attack and security:

- ✓ Did you assess potential forms of attacks to which the AI system could be vulnerable?
 - Did you consider different types and natures of vulnerabilities, such as data pollution, physical infrastructure, cyber-attacks?
- ✓ Did you put measures or systems in place to ensure the integrity and resilience of the AI system against potential attacks?
- ✓ Did you verify how your system behaves in unexpected situations and environments?
- ✓ Did you consider to what degree your system could be dual-use? If so, did you take suitable preventative measures against this case (including for instance not publishing the research or deploying the system)?

Fallback plan and general safety:

- ✓ Did you ensure that your system has a sufficient fallback plan if it encounters adversarial attacks or other unexpected situations (for example technical switching procedures or asking for a human operator before proceeding)?
- ✓ Did you consider the level of risk raised by the AI system in this specific use case?
 - Did you put any process in place to measure and assess risks and safety?
 - Did you provide the necessary information in case of a risk for human physical integrity?
 - Did you consider an insurance policy to deal with potential damage from the AI system?
 - Did you identify potential safety risks of (other) foreseeable uses of the technology, including accidental or malicious misuse? Is there a plan to mitigate or manage these risks?
- ✓ Did you assess whether there is a probable chance that the AI system may cause damage or harm to users or third parties? Did you assess the likelihood, potential damage, impacted audience and severity?
 - Did you consider the liability and consumer protection rules, and take them into account?
 - Did you consider the potential impact or safety risk to the environment or to animals?
 - Did your risk analysis include whether security or network problems such as cybersecurity hazards could pose safety risks or damage due to unintentional behaviour of the AI system?
- ✓ Did you estimate the likely impact of a failure of your AI system when it provides wrong results, becomes unavailable, or provides societally unacceptable results (for example discrimination)?
 - Did you define thresholds and did you put governance procedures in place to trigger alternative/fallback plans?
 - Did you define and test fallback plans?

Accuracy

- ✓ Did you assess what level and definition of accuracy would be required in the context of the AI system and use case?
 - Did you assess how accuracy is measured and assured?
 - Did you put in place measures to ensure that the data used is comprehensive and up to date?
 - Did you put in place measures in place to assess whether there is a need for additional data, for example to improve accuracy or to eliminate bias?
- ✓ Did you verify what harm would be caused if the AI system makes inaccurate predictions?
- ✓ Did you put in place ways to measure whether your system is making an unacceptable amount of inaccurate predictions?
- ✓ Did you put in place a series of steps to increase the system's accuracy?

Reliability and reproducibility:

- ✓ Did you put in place a strategy to monitor and test if the AI system is meeting the goals, purposes and intended applications?
 - Did you test whether specific contexts or particular conditions need to be taken into account to ensure reproducibility?
 - Did you put in place verification methods to measure and ensure different aspects of the system's reliability and reproducibility?
 - Did you put in place processes to describe when an AI system fails in certain types of settings?
 - Did you clearly document and operationalise these processes for the testing and verification of the reliability of AI systems?
 - Did you establish mechanisms of communication to assure (end-)users of the system's reliability?

3. Privacy and data governance

Respect for privacy and data Protection:

- ✓ Depending on the use case, did you establish a mechanism allowing others to flag issues related to privacy or data protection in the AI system's processes of data collection (for training and operation) and data processing?
- ✓ Did you assess the type and scope of data in your data sets (for example whether they contain personal data)?
- ✓ Did you consider ways to develop the AI system or train the model without or with minimal use of potentially sensitive or personal data?
- ✓ Did you build in mechanisms for notice and control over personal data depending on the use case (such as valid consent and possibility to revoke, when applicable)?
- ✓ Did you take measures to enhance privacy, such as via encryption, anonymisation and aggregation?
- ✓ Where a Data Privacy Officer (DPO) exists, did you involve this person at an early stage in the process?

Quality and integrity of data:

- ✓ Did you align your system with relevant standards (for example ISO, IEEE) or widely adopted protocols for daily data management and governance?
- ✓ Did you establish oversight mechanisms for data collection, storage, processing and use?
- ✓ Did you assess the extent to which you are in control of the quality of the external data sources used?
- ✓ Did you put in place processes to ensure the quality and integrity of your data? Did you consider other processes? How are you verifying that your data sets have not been compromised or hacked?

Access to data:

- ✓ What protocols, processes and procedures did you follow to manage and ensure proper data governance?
 - Did you assess who can access users' data, and under what circumstances?
 - Did you ensure that these persons are qualified and required to access the data, and that they have the necessary competences to understand the details of data protection policy?
 - Did you ensure an oversight mechanism to log when, where, how, by whom and for what purpose data was accessed?

4. Transparency

Traceability:

- ✓ Did you establish measures that can ensure traceability? This could entail documenting the following methods:
 - Methods used for designing and developing the algorithmic system:
 - Rule-based AI systems: the method of programming or how the model was built;
 - Learning-based AI systems; the method of training the algorithm, including which input data was gathered and selected, and how this occurred.

- Methods used to test and validate the algorithmic system:
 - Rule-based AI systems; the scenarios or cases used in order to test and validate;
 - Learning-based model: information about the data used to test and validate.
- Outcomes of the algorithmic system:
 - The outcomes of or decisions taken by the algorithm, as well as potential other decisions that would result from different cases (for example, for other subgroups of users).

Explainability:

- ✓ Did you assess:
 - to what extent the decisions and hence the outcome made by the AI system can be understood?
 - to what degree the system's decision influences the organisation's decision-making processes?
 - why this particular system was deployed in this specific area?
 - what the system's business model is (for example, how does it create value for the organisation)?
- ✓ Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?
- ✓ Did you design the AI system with interpretability in mind from the start?
 - Did you research and try to use the simplest and most interpretable model possible for the application in question?
 - Did you assess whether you can analyse your training and testing data? Can you change and update this over time?
 - Did you assess whether you can examine interpretability after the model's training and development, or whether you have access to the internal workflow of the model?

Communication:

- ✓ Did you communicate to (end-)users – through a disclaimer or any other means – that they are interacting with an AI system and not with another human? Did you label your AI system as such?
- ✓ Did you establish mechanisms to inform (end-)users on the reasons and criteria behind the AI system's outcomes?
 - Did you communicate this clearly and intelligibly to the intended audience?
 - Did you establish processes that consider users' feedback and use this to adapt the system?
 - Did you communicate around potential or perceived risks, such as bias?
 - Depending on the use case, did you consider communication and transparency towards other audiences, third parties or the general public?
- ✓ Did you clarify the purpose of the AI system and who or what may benefit from the product/service?
 - Did you specify usage scenarios for the product and clearly communicate these to ensure that it is understandable and appropriate for the intended audience?
 - Depending on the use case, did you think about human psychology and potential limitations, such as risk of confusion, confirmation bias or cognitive fatigue?
- ✓ Did you clearly communicate characteristics, limitations and potential shortcomings of the AI system?
 - In case of the system's development: to whoever is deploying it into a product or service?
 - In case of the system's deployment: to the (end-)user or consumer?

5. Diversity, non-discrimination and fairness

Unfair bias avoidance:

- ✓ Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?
 - Did you assess and acknowledge the possible limitations stemming from the composition of the used data sets?
 - Did you consider diversity and representativeness of users in the data? Did you test for specific populations or problematic use cases?
 - Did you research and use available technical tools to improve your understanding of the data, model and performance?

- Did you put in place processes to test and monitor for potential biases during the development, deployment and use phase of the system?
- ✓ Depending on the use case, did you ensure a mechanism that allows others to flag issues related to bias, discrimination or poor performance of the AI system?
 - Did you establish clear steps and ways of communicating on how and to whom such issues can be raised?
 - Did you consider others, potentially indirectly affected by the AI system, in addition to the (end)-users?
- ✓ Did you assess whether there is any possible decision variability that can occur under the same conditions?
 - If so, did you consider what the possible causes of this could be?
 - In case of variability, did you establish a measurement or assessment mechanism of the potential impact of such variability on fundamental rights?
- ✓ Did you ensure an adequate working definition of “fairness” that you apply in designing AI systems?
 - Is your definition commonly used? Did you consider other definitions before choosing this one?
 - Did you ensure a quantitative analysis or metrics to measure and test the applied definition of fairness?
 - Did you establish mechanisms to ensure fairness in your AI systems? Did you consider other potential mechanisms?

Accessibility and universal design:

- ✓ Did you ensure that the AI system accommodates a wide range of individual preferences and abilities?
 - Did you assess whether the AI system usable by those with special needs or disabilities or those at risk of exclusion? How was this designed into the system and how is it verified?
 - Did you ensure that information about the AI system is accessible also to users of assistive technologies?
 - Did you involve or consult this community during the development phase of the AI system?
- ✓ Did you take the impact of your AI system on the potential user audience into account?
 - Did you assess whether the team involved in building the AI system is representative of your target user audience? Is it representative of the wider population, considering also of other groups who might tangentially be impacted?
 - Did you assess whether there could be persons or groups who might be disproportionately affected by negative implications?
 - Did you get feedback from other teams or groups that represent different backgrounds and experiences?

Stakeholder participation:

- ✓ Did you consider a mechanism to include the participation of different stakeholders in the AI system’s development and use?
- ✓ Did you pave the way for the introduction of the AI system in your organisation by informing and involving impacted workers and their representatives in advance?

6. Societal and environmental well-being

Sustainable and environmentally friendly AI:

- ✓ Did you establish mechanisms to measure the environmental impact of the AI system’s development, deployment and use (for example the type of energy used by the data centres)?
- ✓ Did you ensure measures to reduce the environmental impact of your AI system’s life cycle?

Social impact:

- ✓ In case the AI system interacts directly with humans:
 - Did you assess whether the AI system encourages humans to develop attachment and empathy towards the system?
 - Did you ensure that the AI system clearly signals that its social interaction is simulated and that it

has no capacities of “understanding” and “feeling”?

- ✓ Did you ensure that the social impacts of the AI system are well understood? For example, did you assess whether there is a risk of job loss or de-skilling of the workforce? What steps have been taken to counteract such risks?

Society and democracy:

- ✓ Did you assess the broader societal impact of the AI system’s use beyond the individual (end-)user, such as potentially indirectly affected stakeholders?

7. Accountability

Auditability:

- ✓ Did you establish mechanisms that facilitate the system’s auditability, such as ensuring traceability and logging of the AI system’s processes and outcomes?
- ✓ Did you ensure, in applications affecting fundamental rights (including safety-critical applications) that the AI system can be audited independently?

Minimising and reporting negative Impact:

- ✓ Did you carry out a risk or impact assessment of the AI system, which takes into account different stakeholders that are (in)directly affected?
- ✓ Did you provide training and education to help developing accountability practices?
 - Which workers or branches of the team are involved? Does it go beyond the development phase?
 - Do these trainings also teach the potential legal framework applicable to the AI system?
 - Did you consider establishing an ‘ethical AI review board’ or a similar mechanism to discuss overall accountability and ethics practices, including potentially unclear grey areas?
- ✓ Did you foresee any kind of external guidance or put in place auditing processes to oversee ethics and accountability, in addition to internal initiatives?
- ✓ Did you establish processes for third parties (e.g. suppliers, consumers, distributors/vendors) or workers to report potential vulnerabilities, risks or biases in the AI system?

Documenting trade-offs:

- ✓ Did you establish a mechanism to identify relevant interests and values implicated by the AI system and potential trade-offs between them?
- ✓ How do you decide on such trade-offs? Did you ensure that the trade-off decision was documented?

Ability to redress:

- ✓ Did you establish an adequate set of mechanisms that allows for redress in case of the occurrence of any harm or adverse impact?
- ✓ Did you put mechanisms in place both to provide information to (end-)users/third parties about opportunities for redress?

We invite all stakeholders to pilot this Assessment List in practice and to provide feedback on its implementability, completeness, relevance for the specific AI application or domain, as well as overlap or complementarity with existing compliance or assessment processes. Based on this feedback, a revised version of the Trustworthy AI assessment list will be proposed to the Commission in early 2020

Key guidance derived from Chapter III:

- ✓ Adopt a Trustworthy AI **assessment list** when developing, deploying or using AI systems, and adapt it to the specific use case in which the system is being applied.
- ✓ Keep in mind that such assessment list will **never be exhaustive**. Ensuring Trustworthy AI is not about ticking boxes, but about continuously identifying requirements, evaluating solutions and ensuring improved outcomes throughout the AI system’s lifecycle, and involving stakeholders therein.

C. EXAMPLES OF OPPORTUNITIES AND CRITICAL CONCERNS RAISED BY AI

In the following section, we provide examples of AI development and use that should be encouraged, as well as examples of where AI development, deployment or use can run counter to our values and may raise specific concerns. A balance must be struck between what should and what can be done with AI, and due care must be given to what should not be done with AI.

1. Examples of Trustworthy AI's opportunities

Trustworthy AI can represent a great opportunity to support the mitigation of pressing challenges facing society such as an ageing population, growing social inequality and environmental pollution. This potential is also reflected globally, such as with the UN Sustainable Development Goals.⁵⁷ The following section looks at how to encourage a European AI strategy that tackles some of these challenges.

▪ *Climate action and sustainable infrastructure*

While tackling climate change should be a top priority for policy-makers across the world, digital transformation and Trustworthy AI have a great potential to reduce humans' impact on the environment and enable the efficient and effective use of energy and natural resources.⁵⁸ Trustworthy AI can, for instance, be coupled to big data in order to detect energy needs more accurately, resulting in more efficient energy infrastructure and consumption.⁵⁹

Looking at sectors like public transportation, AI systems for intelligent transport systems⁶⁰ can be used to minimise queuing, optimise routing, allow vision impaired people to be more independent,⁶¹ optimise energy efficient engines and thereby enhance decarbonisation efforts and reduce the environmental footprint, for a greener society. Currently, worldwide, one human dies every 23 seconds in a car accident.⁶² AI systems could help to reduce the number fatalities significantly, for instance through better reaction times and better adherence to rules.⁶³

▪ *Health and well-being*

Trustworthy AI technologies can be used – and are already being used – to render treatment smarter and more targeted, and to help preventing life-threatening diseases.⁶⁴ Doctors and medical professionals can potentially perform a more accurate and detailed analysis of a patient's complex health data, even before people get sick, and provide tailored preventive treatment.⁶⁵ In the context of Europe's ageing population, AI technologies and robotics can be valuable tools to assist caregivers, support elderly care,⁶⁶ and monitor patients' conditions on a real time

⁵⁷ <https://sustainabledevelopment.un.org/?menu=1300>

⁵⁸ A number of EU projects aim for the development of Smart Grids and Energy Storage, which have the potential to contribute to a successful digitally supported energy transition, including through AI-based and other digital solutions. To complement the work of those individual projects, the Commission has launched the BRIDGE initiative, allowing ongoing Horizon 2020 Smart Grid and Energy Storage projects to create a common view on cross cutting issues: <https://www.h2020-bridge.eu/>.

⁵⁹ See for instance the Encompass project: <http://www.encompass-project.eu/>.

⁶⁰ New AI-based solutions help prepare cities for the future of mobility. See for instance the EU funded project called Fabulos: <https://fabulos.eu/>.

⁶¹ See for instance the PRO4VIP project, which is part of the European Vision 2020 strategy to combat preventable blindness, especially due to old age. Mobility and orientation was one of the project's priority areas.

⁶² <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>.

⁶³ The European UP-Drive project for instance aims to address the outlined transport-related challenges by providing contributions enabling gradual automation of and collaboration among vehicles, facilitating a safer, more inclusive and more affordable transportation system. <https://up-drive.eu/>.

⁶⁴ See for instance the REVOLVER (Repeated Evolution of Cancer) project: <https://www.healtheuropa.eu/personalised-cancer-treatment/87958/>, or the Murab project which conducts more accurate biopsies, and which aims at diagnosing cancer and other illnesses faster: <https://ec.europa.eu/digital-single-market/en/news/murab-eu-funded-project-success-story>.

⁶⁵ See for instance the Live INCITE project: www.karolinska.se/en/live-incite. This consortium of healthcare procurers challenges the industry to develop smart AI and other ICT solutions that enable lifestyle interventions in the perioperative process. The target concerns new innovative eHealth solutions that can influence patients in a personalised way to take the necessary actions both prior and after surgery in their lifestyle to optimise the healthcare outcome.

⁶⁶ The EU-funded project CARESSES deals with robots for elderly care, focusing on their cultural sensitivity: they adapt their way of acting and speaking to match the culture and habits of the elderly person they are assisting: <http://caressesrobot.org/en/project/>. See also the AI application called Alfred, a virtual assistant helping older people stay active:

basis, thus saving lives.⁶⁷

Trustworthy AI can also assist on a broader scale. For example, it can examine and identify general trends in the healthcare and treatment sector,⁶⁸ leading to earlier detection of diseases, more efficient development of medicines, more targeted treatments⁶⁹ and ultimately more lives saved.

- *Quality education and digital transformation*

New technological, economic and environmental changes mean that society needs to become more proactive. Governments, industry leaders, educational institutions and unions face a responsibility to bring the citizens into the new digital era ensuring they have the right skills to fill the future jobs. Trustworthy AI technologies could assist in more accurately forecasting which jobs and professions will be disrupted by technology, which new roles will be created and which skills will be needed. This could help governments, unions and industry with planning the (re)skilling of workers. It could also give citizens who may fear redundancy a path of development into a new role.

In addition, AI can be a great tool to fight educational inequalities and create personalised and adaptable education programmes that could help everyone acquire new qualifications, skills and competences according to his or her own ability to learn.⁷⁰ It could increase both the learning speed and the quality of education – reaching from primary school to university.

2. Examples of critical concerns raised by AI

A critical AI concern arises one of the components of Trustworthy AI is violated. Many of the concerns listed below will already fall within the scope of existing legal requirements, which are mandatory and must therefore be complied with. Yet even in circumstances where compliance with legal requirements has been demonstrated, these may not address the full range of ethical concerns that may arise. As our understanding of the adequacy of rules and ethical principles invariably evolves and may change over time, the following non-exhaustive list of concerns may be shortened, expanded, edited or updated in the future.

- *Identifying and tracking individuals with AI*

AI enables the ever more efficient identification of individual persons by both public and private entities. Noteworthy examples of a scalable AI identification technology are face recognition and other involuntary methods of identification using biometric data (i.e. lie detection, personality assessment through micro expressions, and automatic voice detection). Identification of individuals is sometimes the desirable outcome, aligned with ethical principles (for example in detecting fraud, money laundering, or terrorist financing). However, automatic identification raises strong concerns of both a legal and ethical nature, as it may have an unexpected impact on many psychological and sociocultural levels. A proportionate use of control techniques in AI is needed to uphold the autonomy of European citizens. Clearly defining if, when and how AI can be used for automated identification of individuals and differentiating between the identification of an individual vs the tracing and tracking of an individual,

<https://ec.europa.eu/digital-single-market/en/news/alfred-virtual-assistant-helping-older-people-stay-active>. Moreover, the EMPATTICS project (EMpowering PATients for a BeTTer Information and improvement of the Communication Systems) will research and define how health care professionals and patients use ICT technologies including AI systems to plan interventions with patients and to monitor the progression of their physical and mental state: www.empattics.eu.

⁶⁷ See for instance the MyHealth Avatar (www.myhealthavatar.eu), which offers a digital representation of a patient's health status. The research project launched an app and an online platform that collects, and gives access to, your digital long-term health-status information. This takes on the form of a life-long health companion ('avatar'). MyHealthAvatar also predicts your risk for stroke, diabetes, cardiovascular disease and hypertension.

⁶⁸ See for instance the ENRICHME project (www.enrichme.eu), which tackles the progressive decline of cognitive capacity in the ageing population. An integrated platform for Ambient Assisted Living (AAL) and a mobile service robot for long-term monitoring and interaction will help the elderly to remain independent and active for longer.

⁶⁹ See for instance the use of AI by Sophia Genetics, which leverages statistical inference, pattern recognition and machine learning to maximize the value of genomics and radiomics data: <https://www.sophiagenetics.com/home.html>.

⁷⁰ See for instance the MaTHiSiS project, aimed at providing a solution for affect-based learning in a comfortable learning environment, comprising of high-end technological devices and algorithms: (<http://mathisis-project.eu/>). See also IBM's Watson Classroom or Century Tech's platform.

and between targeted surveillance and mass surveillance, will be crucial for the achievement of Trustworthy AI. The application of such technologies must be clearly warranted in existing law.⁷¹ Where the legal basis for such activity is “consent”, practical means⁷² must be developed which allow meaningful and verified consent to be given to being automatically identified by AI or equivalent technologies. This also applies to the usage of “anonymous” personal data that can be re-personalised.

- *Covert AI systems*

Human beings should always know if they are directly interacting with another human being or a machine, and it is the responsibility of AI practitioners that this is reliably achieved. AI practitioners should therefore ensure that humans are made aware of – or able to request and validate the fact that – they interact with an AI system (for instance, by issuing clear and transparent disclaimers). Note that borderline cases exist and complicate the matter (e.g. an AI-filtered voice spoken by a human). It should be borne in mind that the confusion between humans and machines could have multiple consequences such as attachment, influence, or reduction of the value of being human.⁷³ The development of human-like robots⁷⁴ should therefore undergo careful ethical assessment.

- *AI enabled citizen scoring in violation of fundamental rights*

Societies should strive to protect the freedom and autonomy of all citizens. Any form of citizen scoring can lead to the loss of this autonomy and endanger the principle of non-discrimination. Scoring should only be used if there is a clear justification, and where measures are proportionate and fair. Normative citizen scoring (general assessment of “moral personality” or “ethical integrity”) in *all* aspects and on a large scale by public authorities or private actors endangers these values, especially when used not in accordance with fundamental rights, and when used disproportionately and without a delineated and communicated legitimate purpose.

Today, citizen scoring – on a large or smaller scale – is already often used in purely descriptive and domain-specific scorings (e.g. school systems, e-learning, and driver licences). Even in those more narrow applications, a fully transparent procedure should be made available to citizens, including information on the process, purpose and methodology of the scoring. Note that transparency cannot prevent non-discrimination or ensure fairness, and is not the panacea against the problem of scoring. Ideally the possibility of opting out of the scoring mechanism when possible without detriment should be provided – otherwise mechanisms for challenging and rectifying the scores must be given. This is particularly important in situations where an asymmetry of power exists between the parties. Such opt-out options should be ensured in the technology’s design in circumstances where this is necessary to ensure compliance with fundamental rights and is necessary in a democratic society.

- *Lethal autonomous weapon systems (LAWS)*

Currently, an unknown number of countries and industries are researching and developing lethal autonomous weapon systems, ranging from missiles capable of selective targeting to learning machines with cognitive skills to decide whom, when and where to fight without human intervention. This raises fundamental ethical concerns, such as the fact that it could lead to an uncontrollable arms race on a historically unprecedented level, and create military contexts in which human control is almost entirely relinquished and the risks of malfunction are not addressed. The European Parliament has called for the urgent development of a common, legally binding position addressing ethical and legal questions of human control, oversight, accountability and implementation of international human rights law, international humanitarian law and military strategies.⁷⁵ Recalling the European Union’s aim to promote peace as enshrined in Article 3 of the Treaty of the European Union, we stand with, and look to support, the Parliament’s resolution of 12 September 2018 and all related efforts on LAWS.

⁷¹ In this regard, Article 6 of the GDPR can be recalled, which provides, among other things, that processing of data shall only be lawful if it has a valid legal basis.

⁷² As current mechanisms for giving informed consent in the internet show, consumers typically give consent without meaningful consideration. Hence, they can hardly be classified as practical.

⁷³ Madary & Metzinger (2016). Real Virtuality: A Code of Ethical Conduct. Recommendations for Good Scientific Practice and the Consumers of VR-Technology. *Frontiers in Robotics and AI*, 3(3).

⁷⁴ This also applies to AI-driven avatars.

⁷⁵ European Parliament’s Resolution 2018/2752(RSP).

- *Potential longer-term concerns*

AI development is still domain-specific and requires well-trained human scientists and engineers to precisely specify its targets. However, extrapolating into the future with a longer time horizon, certain critical long-term concerns can be hypothesized.⁷⁶ A risk-based approach suggests that these concerns should be kept into consideration in view of possible unknown unknowns and “black swans.”⁷⁷ The high-impact nature of these concerns, combined with the current uncertainty in corresponding developments, calls for regular assessments of these topics.

D. CONCLUSION

This document constitutes the AI Ethics Guidelines produced by the High-Level Expert Group on Artificial Intelligence (AI HLEG).

We recognise the positive impact that AI systems already have and will continue having, both commercially and societally. However, we are equally concerned to ensure that the risks and other adverse impacts with which these technologies are associated are properly and proportionately handled. AI is a technology that is both transformative and disruptive, and its evolution over the last several years has been facilitated by the availability of enormous amounts of digital data, major technological advances in computational power and storage capacity, as well as significant scientific and engineering innovation in AI methods and tools. AI systems will continue to impact society and citizens in ways that we cannot yet imagine.

In this context, it is important to build AI systems that are worthy of trust, since human beings will only be able to confidently and fully reap its benefits when the technology, including the processes and people behind the technology, are trustworthy. When drafting these Guidelines, Trustworthy AI has, therefore, been our foundational ambition.

Trustworthy AI has three components: (1) it should be lawful, ensuring compliance with all applicable laws and regulations, (2) it should be ethical, ensuring adherence to ethical principles and values and (3) it should be robust, both from a technical and social perspective since to ensure that, even with good intentions, AI systems do not cause any unintentional harm. Each component is necessary but not sufficient to achieve Trustworthy AI. Ideally, all three components work in harmony and overlap in their operation. Where tensions arise, we should endeavour to align them.

In Chapter I, we articulated the fundamental rights and a corresponding set of ethical principles that are crucial in an AI-context. In Chapter II, we listed seven key requirements that AI systems should meet in order to realise Trustworthy AI. We proposed technical and non-technical methods that can help with their implementation. Finally, in Chapter III we provided a Trustworthy AI assessment list that can help operationalising the seven requirements. In a final section, we provided examples of beneficial opportunities and critical concerns raised by AI systems, on which we hope to stimulate further discussion.

Europe has a unique vantage point based on its focus on placing the citizen at the heart of its endeavours. This focus is written into the very DNA of the European Union through the Treaties upon which it is built. The current document forms part of a vision that promotes Trustworthy AI which we believe should be the foundation upon which Europe can build leadership in innovative, cutting-edge AI systems. This ambitious vision will help securing human flourishing of European citizens, both individually and collectively. Our goal is to create a culture of “Trustworthy AI for Europe”, whereby the benefits of AI can be reaped by all in a manner that ensures respect for our foundational values: fundamental rights, democracy and the rule of law.

⁷⁶ While some consider that Artificial General Intelligence, Artificial Consciousness, Artificial Moral Agents, Super-intelligence or Transformative AI can be examples of such long-term concerns (currently non-existent), many others believe these to be unrealistic.

⁷⁷ A black swan event is a very rare, yet high impact, event – so rare, that it might not have been observed. Hence, probability of occurrence typically can only be estimated with high uncertainty.

GLOSSARY

This glossary pertains to the Guidelines and is meant to help in the understanding of the terms used in this document.

Artificial Intelligence or AI systems

Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans⁷⁸ that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.

As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems).

A separate document prepared by the AI HLEG and elaborating on the definition of AI used for the purpose of this document is published in parallel, titled "A definition of AI: Main capabilities and scientific disciplines".

AI Practitioners

By AI practitioners we denote all individuals or organisations that develop (including research, design or provide data for) deploy (including implement) or use AI systems, excluding those that use AI systems in the capacity of end-user or consumer.

AI system's life cycle

An AI system's life cycle encompasses its development (including research, design, data provision, and limited trials), deployment (including implementation) and use phase.

Auditability

Auditability refers to the ability of an AI system to undergo the assessment of the system's algorithms, data and design processes. This does not necessarily imply that information about business models and Intellectual Property related to the AI system must always be openly available. Ensuring traceability and logging mechanisms from the early design phase of the AI system can help enabling the system's auditability.

Bias

Bias is an inclination of prejudice towards or against a person, object, or position. Bias can arise in many ways in AI systems. For example, in data-drive AI systems, such as those produced through machine learning, bias in data collection and training can result in an AI system demonstrating bias. In logic-based AI, such as rule-based systems, bias can arise due to how a knowledge engineer might view the rules that apply in a particular setting. Bias can also arise due to online learning and adaptation through interaction. It can also arise through personalisation whereby users are presented with recommendations or information feeds that are tailored to the user's tastes. It does not necessarily relate to human bias or human-driven data collection. It can arise, for example, through the limited contexts in which a system is used, in which case there is no opportunity to generalise it to other contexts. Bias can be good or bad, intentional or unintentional. In certain cases, bias can result in discriminatory and/or unfair outcomes, indicated in this document as unfair bias.

⁷⁸

Humans design AI systems directly, but they may also use AI techniques to optimise their design.

Ethics

Ethics is an academic discipline which is a subfield of philosophy. In general terms, it deals with questions like “What is a good action?”, “What is the value of a human life?”, “What is justice?”, or “What is the good life?”. In academic ethics, there are four major fields of research: (i) Meta-ethics, mostly concerning the meaning and reference of normative sentence, and the question how their truth values can be determined (if they have any); (ii) normative ethics, the practical means of determining a moral course of action by examining the standards for right and wrong action and assigning a value to specific actions; (iii) descriptive ethics, which aims at an empirical investigation of people's moral behaviour and beliefs; and (iv) applied ethics, concerning what we are obligated (or permitted) to do in a specific (often historically new) situation or a particular domain of (often historically unprecedented) possibilities for action. Applied ethics deals with real-life situations, where decisions have to be made under time-pressure, and often limited rationality. AI Ethics is generally viewed as an example of applied ethics and focuses on the normative issues raised by the design, development, implementation and use of AI.

Within ethical discussions, the terms “moral” and “ethical” are often used. The term “moral” refers to the concrete, factual patterns of behaviour, the customs, and conventions that can be found in specific cultures, groups, or individuals at a certain time. The term “ethical” refers to an evaluative assessment of such concrete actions and behaviours from a systematic, academic perspective.

Ethical AI

In this document, ethical AI is used to indicate the development, deployment and use of AI that ensures compliance with ethical norms, including fundamental rights as special moral entitlements, ethical principles and related core values. It is the second of the three core elements necessary for achieving Trustworthy AI.

Human-Centric AI

The human-centric approach to AI strives to ensure that human values are central to the way in which AI systems are developed, deployed, used and monitored, by ensuring respect for fundamental rights, including those set out in the Treaties of the European Union and Charter of Fundamental Rights of the European Union, all of which are united by reference to a common foundation rooted in respect for human dignity, in which the human being enjoy a unique and inalienable moral status. This also entails consideration of the natural environment and of other living beings that are part of the human ecosystem, as well as a sustainable approach enabling the flourishing of future generations to come.

Red Teaming

Red teaming is the practice whereby a “red team” or independent group challenges an organisation to improve its effectiveness by assuming an adversarial role or point of view. It is particularly used to help identifying and addressing potential security vulnerabilities.

Reproducibility

Reproducibility describes whether an AI experiment exhibits the same behaviour when repeated under the same conditions.

Robust AI

Robustness of an AI system encompasses both its technical robustness (appropriate in a given context, such as the application domain or life cycle phase) and as well as its robustness from a social perspective (ensuring that the AI system duly takes into account the context and environment in which the system operates). This is crucial to ensure that, even with good intentions, no unintentional harm can occur. Robustness is the third of the three components necessary for achieving Trustworthy AI.

Stakeholders

By stakeholders we denote all those that research develop, design, deploy or use AI, as well as those that are (directly or indirectly) affected by AI – including but not limited to companies, organisations, researchers, public services, institutions, civil society organisations, governments, regulators, social partners, individuals, citizens,

workers and consumers.

Traceability

Traceability of an AI system refers to the capability to keep track of the system's data, development and deployment processes, typically by means of documented recorded identification.

Trust

We take the following definition from the literature: "Trust is viewed as: (1) a set of specific beliefs dealing with benevolence, competence, integrity, and predictability (trusting beliefs); (2) the willingness of one party to depend on another in a risky situation (trusting intention); or (3) the combination of these elements."⁷⁹ While "Trust" is usually not a property ascribed to machines, this document aims to stress the importance of being able to trust not only in the fact that AI systems are legally compliant, ethically adherent and robust, but also that such trust can be ascribed to all people and processes involved in the AI system's life cycle.

Trustworthy AI

Trustworthy AI has three components: (1) it should be lawful, ensuring compliance with all applicable laws and regulations (2) it should be ethical, demonstrating respect for, and ensure adherence to, ethical principles and values and (3) it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm. Trustworthy AI concerns not only the trustworthiness of the AI system itself but also comprises the trustworthiness of all processes and actors that are part of the system's life cycle.

Vulnerable Persons and Groups

No commonly accepted or widely agreed legal definition of vulnerable persons exists, due to their heterogeneity. What constitutes a vulnerable person or group is often context-specific. Temporary life events (such as childhood or illness), market factors (such as information asymmetry or market power), economic factors (such as poverty), factors linked to one's identity (such as gender, religion or culture) or other factors can play a role. The Charter of Fundamental Rights of the EU encompasses under Article 21 on non-discrimination the following grounds, which can be a reference point amongst others: namely sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age and sexual orientation. Other articles of law address the rights of specific groups, in addition to those listed above. Any such list is not exhaustive, and may change over time. A vulnerable group is a group of persons who share one or several characteristics of vulnerability.

⁷⁹ Siau, K., Wang, W. (2018), Building Trust in Artificial Intelligence, Machine Learning, and Robotics, *CUTTER BUSINESS TECHNOLOGY JOURNAL* (31), S. 47–53.

listed here below in alphabetical order

Pekka Ala-Pietilä, Chair of the AI HLEG	Pierre Lucas
AI Finland, Huhtamaki, Sanoma	Orgalim – Europe’s technology industries
Wilhelm Bauer	Ieva Martinkenaite
Fraunhofer	Telenor
Urs Bergmann – Co-Rapporteur	Thomas Metzinger – Co-Rapporteur
Zalando	JGU Mainz & European University Association
Mária Bielíková	Catelijne Muller
Slovak University of Technology in Bratislava	ALLAI Netherlands & EESC
Cecilia Bonefeld-Dahl – Co-Rapporteur	Markus Noga
DigitalEurope	SAP
Yann Bonnet	Barry O’Sullivan, Vice-Chair of the AI HLEG
ANSSI	University College Cork
Loubna Bouarfa	Ursula Pacht
OKRA	BEUC
Stéphan Brunessaux	Nicolas Petit – Co-Rapporteur
Airbus	University of Liège
Raja Chatila	Christoph Peylo
IEEE Initiative Ethics of Intelligent/Autonomous Systems &	Bosch
Sorbonne University	
Mark Coeckelbergh	Iris Plöger
University of Vienna	BDI
Virginia Dignum – Co-Rapporteur	Stefano Quintarelli
Umea University	Garden Ventures
Luciano Floridi	Andrea Renda
University of Oxford	College of Europe Faculty & CEPS
Jean-Francois Gagné – Co-Rapporteur	Francesca Rossi
Element AI	IBM
Chiara Giovannini	Cristina San José
ANEC	European Banking Federation
Joanna Goodey	George Sharkov
Fundamental Rights Agency	Digital SME Alliance
Sami Haddadin	Philipp Slusallek
Munich School of Robotics and MI	German Research Centre for AI (DFKI)
Gry Hasselbalch	Françoise Soulié Fogelman
The thinkdotank DataEthics & Copenhagen University	AI Consultant
Fredrik Heintz	Saskia Steinacker – Co-Rapporteur
Linköping University	Bayer
Fanny Hidvegi	Jaan Tallinn
Access Now	Ambient Sound Investment
Eric Hilgendorf	Thierry Tingaud
University of Würzburg	STMicroelectronics
Klaus Höckner	Jakob Uszkoreit
Hilfsgemeinschaft der Blinden und Sehschwachen	Google
Mari-Noëlle Jégo-Laveissière	Aimee Van Wynsberghe – Co-Rapporteur
Orange	TU Delft
Leo Kärkkäinen	Thiébaut Weber
Nokia Bell Labs	ETUC
Sabine Theresia Köszegi	Cecile Wendling
TU Wien	AXA
Robert Kroplewski	Karen Yeung – Co-Rapporteur
Solicitor & Advisor to Polish Government	The University of Birmingham
Elisabeth Ling	
RELX	

Urs Bergmann, Cecilia Bonefeld-Dahl, Virginia Dignum, Jean-François Gagné, Thomas Metzinger, Nicolas Petit, Saskia Steinacker, Aimee Van Wynsberghe and Karen Yeung acted as rapporteurs for this document.

Pekka Ala-Pietilä is Chairing the AI HLEG. Barry O’Sullivan is Vice-Chair, coordinating the AI HLEG’s second deliverable. Nozha Boujemaa, Vice-Chair until 1 February 2019 coordinating the first deliverable, also contributed to the content of this document.

Nathalie Smuha provided editorial support.

Brussels, 7.12.2018
COM(2018) 795 final

ANNEX

ANNEX

to the

**COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN
ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE
REGIONS**

Coordinated Plan on Artificial Intelligence

Coordinated Plan on the Development and Use of Artificial Intelligence Made in Europe – 2018

Artificial Intelligence (AI) can help us address some of the world's biggest challenges. It can enable doctors to improve diagnoses and develop therapies for diseases for which none exist yet; it can reduce energy consumption by optimising resources; it can contribute to a cleaner environment by lessening the need for pesticides; it can help improve weather prediction and anticipate disasters; and so on. The list is virtually endless. AI will be the main driver of economic and productivity growth and will contribute to the sustainability and viability of the industrial base in Europe¹. Like the steam engine or electricity in the past, AI is transforming the world.

The Union aims to develop trusted AI based on ethical and societal values building on its Charter of Fundamental Rights. People should not only trust AI, but also benefit from the use of AI for their personal and professional lives. Europe aims at creating an innovation friendly ecosystem for AI: an environment where economic players find the infrastructure, research facilities, testing environments, financial means, legal framework and adequate skills levels to invest in and deploy AI. **Overall, the ambition is for Europe to become the world-leading region for developing and deploying cutting-edge, ethical and secure AI, promoting a human-centric approach in the global context.**

AI has been high on the agenda of the Council of the EU since the Digital Summit organised by the Estonian Presidency in September 2017. The Communication “Artificial Intelligence for Europe” of 25 April 2018² proposes a European strategy in support of this goal. The Communication also proposes a coordinated plan on the development of AI in Europe³, to be drawn up with the Member States by the end of 2018. This was endorsed by the European Council⁴. The present document responds to this request. Only if Member States and the Commission work together will Europe be able to turn vision into reality.

The coordinated plan builds on a "declaration of cooperation" that was signed by all EU Member States and Norway in the context of the Digital Day 2018⁵, emphasising the

¹ The High Level Strategy Groupon Industrial Technologies has recommended including AI as one of the Key Enabling Technologies due to its cross cutting enabling potential crucial for European Industry. <https://publications.europa.eu/en/publication-detail/-/publication/28e1c485-476a-11e8-be1d-01aa75ed71a1/language-en>

² COM(2018) 237

³ As spelled out in the above-mentioned Communication of 25 April 2018, Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications). We are using AI on a daily basis, e.g. to translate languages, generate subtitles in videos or to block email spam. Many AI technologies require data to improve their performance. Once they perform well, they can help improve and automate decision making in the same domain. For example, an AI system will be trained and then used to spot cyberattacks on the basis of data from the concerned network or system.

⁴ <https://www.consilium.europa.eu/en/press/press-releases/2018/06/29/20180628-euco-conclusions-final/>

⁵ The Digital Day 2018, which took place on 10 April in Brussels, reached for joint commitments of Member States in relation to the digital future of Europe. The signing of the AI declaration takes place through a voluntary and participatory process.

willingness to cooperate more closely on AI. The Austrian Presidency of the EU also included AI as a priority in the context of the transformation of the industry⁶.

The main goals of the coordinated plan are to maximise the impact of investments at EU and national levels, encourage synergies and cooperation across the EU, including on ethics, foster the exchange of best practices and collectively define the way forward. By working together the Union can maximise its impact to compete globally.

The Member States' Group on Digitising European Industry and AI and the Commission discussed between June and November 2018 possible strands for cooperation. In order to accommodate the fast pace of the AI-induced change in societies and economies Member States, Norway and Switzerland agreed to put in place a rolling coordinated plan that is to be monitored and reviewed annually, so as to ensure it remains up to date. The present document is the first edition of this plan and mainly comprises activities for 2019 and 2020, with emphasis on planned EU-level activities under the current financial framework. The plan is expected to run into the next decade, possibly until 2027, in line with the next multi-annual financial framework.⁷

A need for coordinated action has been identified in the fields of investment, excellence in and diffusion of AI, data availability, societal challenges, ethics and the regulatory framework. Actions concern both private and public sectors with many synergies.

“AI made in Europe” meeting citizens’ aspirations, responding to societal needs, and boosting competitiveness

The coordinated plan will maximise the benefits of AI for all Europeans by fostering the development of trusted AI that corresponds to European ethical values, and citizens’ aspirations. Europe will progressively increase its effort in public interest areas such as healthcare, transport, security, education and energy as well as in other areas such as manufacturing and financial services (including through blockchain).

This plan brings together a set of concrete and complementary actions at EU, national and regional level⁸ in view of:

- Boosting investments and reinforcing excellence in AI technologies and applications which are trustworthy and “ethical and secure by design”. Investments shall take place in a stable regulatory context which enables experimentation and supports disruptive innovation across the EU, ensuring the widest and best use of AI by the European economy and society.
- Building on Europe’s strengths, to develop and implement in partnership with industry and Member States shared agendas for industry-academia collaborative Research and Development (R&D) and innovation.
- Adapting learning and skilling programmes and systems to prepare Europe’s society and its future generations for AI.
- Building up essential capacities in Europe underpinning AI such as data spaces and world-class reference sites for testing and experimentation.

⁶ See Presidency note 11972/18 of 14 September 2018.

⁷ All budget figures related to envisaged EU contributions from 2020 onwards are subject to the adoption of the underlying legal basis, work programmes and yearly budgets by the relevant authorities.

⁸ All of these actions must comply with the EU rules on competition law and state aid.

- Making public administrations in Europe frontrunners in the use of AI.
- Implementing, on the basis of expert work, clear ethics guidelines for the development and the use of AI in full respect of fundamental rights, with a view to set global ethical standards and be a world leader in ethical, trusted AI.
- Where needed, reviewing the existing national and European legal framework to better adapt them to specific challenges.

Key enablers

Progress in AI opens the door to new opportunities in areas such as personalised and precision healthcare, mobility (autonomous driving⁹), fintech, advanced manufacturing, space-based applications, smart power grids, sustainable circular and bio economy, improved detection and investigation of criminal activities (*e.g.* money laundering, tax fraud), media, etc.

This digital transformation requires in many cases a significant upgrading of the currently available infrastructure. The effective implementation of AI will require the completion of the Digital Single Market and its regulatory framework including the swift adoption of the Commission proposal for a European Cybersecurity Industrial, Technology and Research Competence Centre and the Network of National Coordination Centres¹⁰, reinforced connectivity through spectrum coordination, very fast 5G mobile networks and optical fibres, next generation clouds, as well as satellite technologies¹¹. High-performance computing and AI will increasingly intertwine as we transit to a future using new computing, storage and communication technologies. Furthermore, infrastructures should be both accessible and affordable to ensure an inclusive AI adoption across Europe, particularly by small and medium-sized enterprises (SMEs). Industry, and in particular small and young companies, will need to be in a position to be aware and able to integrate these technologies in new products, services and related production processes and technologies, including by upskilling and reskilling their workforce. Standardisation will also be essential for the development of AI in the Digital Single Market, helping notably to ensure interoperability.

A paradigm change is also required in cases where data needs to be processed locally (for example in connected automated driving that must be able to take swift decisions without waiting for an answer from a remote server). This trend drives demand for advanced, low-power semiconductor technologies. New paradigms beyond scaling are already emerging and new energy-efficient computing architectures (such as neuromorphic and quantum) will be needed to ensure sustainable use of energy. Ongoing partnerships between the Member States and the Union through joint undertakings such as ECSEL¹² (for electronic components and systems), EuroHPC (high-performing computing)¹³ as well as the quantum flagship under the Research and Innovation Programme Horizon 2020 (H2020) are key to processing big data and sustain further developments in AI.

Member States and the Commission will continuously support the deployment of the key enablers and will mainstream AI in the related initiatives.

⁹ The Communication on connected and automated mobility recognised for example the benefits stemming from the progress in the field of AI, which will serve to open up new areas for business development and pave the way for new mobility services making transport safer, more accessible and sustainable.

¹⁰ COM(2018) 630

¹¹ *E.g.* the EU-owned Global Satellite Navigation System Galileo.

¹² https://europa.eu/european-union/about-eu/agencies/ecsel_en

¹³ <https://ec.europa.eu/digital-single-market/en/blogposts/eurohpc-joint-undertaking-looking-ahead-2019-2020-and-beyond>

The coordinated plan links to the ongoing parallel strategies in these areas.

A. Strategic actions and coordination

The Commission put forward its approach to AI in the Communication “Artificial Intelligence for Europe” published in April 2018, resting on three pillars:

- Boosting the EU's technological and industrial capacity and AI uptake across the economy, both by the private and public sectors;
- Preparing for socio-economic changes brought about by AI;
- Ensuring an appropriate ethical and legal framework, based on the Union's values and in line with the Charter of Fundamental Rights of the European Union.

As announced in the April Communication, the Commission has tasked a High-level Expert Group on AI to draft AI ethics guidelines. The High-Level Expert Group will also propose policy recommendations on investments and the regulatory framework.

The Commission has also tasked the Expert Group on Liability and New Technologies to assist the Commission in drawing up guidance on the implementation of the Product Liability Directive and in developing EU-wide principles which can serve as guidelines for possible adaptations of applicable laws at EU and national level.

The Commission's Expert Group for the Observatory of the Online Platform Economy will additionally explore policy issues in AI-related regulatory areas, such as data access¹⁴, online advertising and the role of algorithms in the digital platform economy.

Moreover, the Commission also established a High-Level Expert Group on the impact of digital transformation on EU labour markets which will deliver a report addressing strategies to deal with employment disruption in spring 2019¹⁵.

Investment levels for AI in the Union are low and fragmented, relative to other parts of the world such as the US and China. To remedy this shortcoming, the April Communication sets an ambitious objective, aiming to increase investment and reach a total (public and private sectors combined) of at least EUR 20 billion in the period 2018-2020, and to increase investments progressively to EUR 20 billion per year in the course of the next decade.

The Commission is increasing investments in AI under the research and innovation framework programme Horizon 2020 to EUR 1.5 billion in the period 2018-2020, constituting a 70% increase compared to 2014-2017. Under the next multi-annual financial framework, the Commission has proposed to dedicate at least EUR 1 billion per year from Horizon Europe¹⁶ and the Digital Europe Programme¹⁷ to AI. Options to mobilise resources from the European Fund for Strategic Investments and the European structural and investment funds are being

¹⁴ Duch-Brown et al (2017), The economics of ownership, access and trade in digital data. Joint Research Centre Digital Economy Working Paper 2017-01. <https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/economics-ownership-access-and-trade-digital-data>

¹⁵ <https://ec.europa.eu/digital-single-market/en/news/appointment-members-high-level-expert-group-impact-digital-transformation-eu-labour-markets>

¹⁶ COM(2018) 435 and COM(2018) 436

¹⁷ COM(2018) 434

explored. For example the European Regional Development Fund is expected to invest in AI based on the next generation of Smart Specialisation Strategies.

As of today, France, Finland, Sweden, the UK and Germany have targeted AI strategies in place. Some countries like Denmark, Luxembourg, the Netherlands, Ireland and Norway include AI related actions in a broader digitalisation strategy. Moreover, Austria, Belgium, Czech Republic, Denmark, Estonia, Germany, Italy, Latvia, Poland, Portugal, Slovenia, Slovakia and Spain are in the process of developing strategies.¹⁸ In order to maximise investments, pool important resources such as data, provide a seamless regulatory environment, all Member States need to put in place national AI strategies, in line with their intentions expressed in the Digital Day Declaration of Cooperation on AI, including support measures.

Member States and the Commission will monitor progress of the implementation of the plan on annual basis.¹⁹

- ✓ **By mid-2019 all Member States are encouraged to put in place - and share with other Member States and the Commission - national AI strategies or programmes or add AI dimensions in other relevant strategies and programmes²⁰ outlining investment levels and implementation measures**, taking into account this coordinated plan. The exact form, contents and governance of the national AI strategies will be up to each Member State to decide based on national characteristics²¹.
- ✓ Discussions between Member States and the Commission will be steered by the **Member States' Group on Digitising European Industry and Artificial Intelligence** (MS Group on DEI and AI), assisted on technical matters by a Sherpa Group²². The group will meet at least biannually. It will ensure coordination across different national Ministries and other stakeholders, for example from industry, academia and civil society. Dedicated Member States' groups²³ will provide the necessary input in the different areas covered by the plan. The Member States and the Commission will also organise topical workshops.
- ✓ To assess impacts, in **2019 Member States and the Commission will identify relevant investment parameters** and comparable benchmarks for uptake so as to achieve common targets. Progress will be monitored annually.

¹⁸ Five regions have AI related priorities in their Smart Specialisation Strategies and AI may play a role in the context of the digital industrial transition dimensions of the national or regional Smart Specialisation Strategies in view of the post-2020 ERDF programmes: Lower Saxony [DE], Pohjois-Savo [FI], Łódzkie [PL], North-West [RO] and North-East [RO]. See: <http://s3platform.jrc.ec.europa.eu/map> . There are also other regions in Europe, for example in Belgium, with AI strategies.

¹⁹ AI Watch developed by the Joint Research Centre will contribute to monitoring AI-related development and will provide a number of analyses necessary to support the implementation of the European AI initiative. Among others it will develop AI indexes addressing all dimensions relevant for policy making. Such information will be made available at the AI Watch portal https://ec.europa.eu/knowledge4policy/ai-watch_en

²⁰ Including in the field of social inclusion and employment, eGovernment, eHealth, Key Enabling Technologies, skills, industrial transition / smart specialisation, etc.

²¹ Member States and regions are encouraged to analyse digital – including AI - dimensions in the process for reviewing the Smart Specialisation Strategies in view of the future European Regional Development Fund investments.

²² Representatives nominated by the Member States' Group on Digitising European Industry and Artificial Intelligence.

²³ Building on existing groups and respecting the specific governance of the different EU instruments involved.

B. Maximising investments through partnerships

To facilitate and reinforce investment in AI, and to maximise its impact in both the public and the private sectors, joint efforts between the Commission, Member States and the private sector are necessary. Only if both the Commission and Member States direct their investments in the same direction through joint programming and leverage significant private investments, will Europe as a whole have an impact and establish its strategic autonomy in AI.

- **Paving the way to a new partnership on AI:** AI is addressed today in different Public-Private Partnerships (PPP) under Horizon 2020, notably the Robotics and the Big Data PPPs with separate Research & Innovation agendas. The academic research community is also organised in networks such as EurAI, the European association for AI. The Commission, with support from the Member States, will work together with industry and academia on a common Research and Innovation Agenda in AI. Building on Europe's strengths, it will aim at the development of a dynamic EU-wide AI innovation ecosystem, fostering close cooperation between industry and academia, and reinforcing competitiveness across the whole AI value chain. For that, it will facilitate discussions with initially stakeholders from the robotics and big data PPP, but later expanding to representatives of all relevant stakeholders from industry and research institutes to develop a common strategic research and innovation agenda for AI. The Commission is planning to set up a Leaders' Group.
- **Teaming up for investment in AI:** It is key for Europe to identify and invest in the next generation of AI and to roll it out widely. One important element is making available sufficient investment for start-ups in their early stage as well as for companies in their scale-up phase. To this end, the Commission aims at making available resources for start-ups and innovators in AI and blockchain in their early stage as well as for companies in their scale-up phase, using existing instruments such as the European Fund for Strategic Investments, Horizon 2020 and the European Investment Fund. EUR 100 million should be initially mobilised in 2020. As certain early applications of blockchain utilising mining (Bitcoin) consume high amounts of energy, in the investment selection criteria for such a financial programme, the Commission will give preference to supporting newer energy efficient blockchain infrastructures and applications. Activities could focus on (i) financing a portfolio of innovative AI/blockchain companies; (ii) developing a dynamic EU-wide investors community focusing on AI; (iii) multiplying investments at the national level by involving the national promotional banks (NPBs) that are willing to participate; (iv) incentivising private sector investments and (v) making Europe become more attractive for start-ups to stay and grow. In the following years, AI and blockchain could be further supported through the InvestEU Programme.
- As a follow-up of the European Council conclusions of June 2018²⁴, the European Commission is preparing an enhanced pilot for a European Innovation Council (EIC) to support the scale-up of innovative companies (start-ups and SMEs) carrying out breakthrough market creating innovation, as well as breakthrough science and key enabling technologies that could lead to disruptive innovation.

²⁴ EUCO 9/18 – 28 June 2018.

Maximising investment:

- ✓ In 2019, **the Commission** will bring together the stakeholders initially including the robotics and big data PPPs but later expanding to other involved parties, in order to develop a **common strategic research and innovation agenda for AI**, to be supported starting in 2020. For that, it will set up a Leaders' Group representing the stakeholders at CEO level from industry and research institutes to develop the agenda and ensure commitment at the highest level in its implementation, paving the way to a new partnership in AI (first meeting early 2019).
- ✓ The Commission aims at making available resources for start-ups and innovators in AI and blockchain in their early stage as well as for companies in their scale-up phase, using existing instruments such as the European Fund for Strategic Investments, Horizon 2020 and the European Investment Fund. EUR 100 million should be initially mobilised in 2020. The Commission will also launch an investment support programme to facilitate portfolio development, co-investment with Member States and private investors, and in order to raise the awareness of start-ups and other companies, including both traditional and innovative SMEs, facilitating engagement in otherwise risky projects. This will help prepare for strengthening access to finance for AI under the InvestEU programme.
- ✓ Member States can actively support the process **through the participation of national promotional banks**, and taking part in awareness raising support programmes.
- ✓ The **European Innovation Council** will support disruptive innovation through the enhanced European Innovation Council pilot which will in particular support cutting-edge, high reward research and innovation projects that aim to demonstrate a new technological paradigm in fields such as for example **Human centric AI**, through a fund totalling EUR 100 million in 2019-2020.
- ✓ **Member States are encouraged to** explore the implementation of innovation vouchers, small grants and loans targeting SMEs in their digital transformation, including notably the integration of AI in products, process and business models.

C. From the lab to the market: (i) building up research excellence, (ii) establishing world-reference testing facilities and (iii) accelerating AI take-up through Digital Innovation Hubs.

The Commission and the Member States acknowledge the importance of continuing to strengthen their science base²⁵ and support research and innovation in order to ensure competitiveness in technology, address innovation challenges, as well as facilitate the transfer of research results to industry.

The Commission will reinforce its investment in research and innovation throughout Horizon 2020, and mainstream AI in all the topics where its benefit can be developed or exploited. For example, it will be necessary that substantial funding is allocated to AI for security purposes, namely, on the one hand to prevent the malicious use of AI technologies utilized by malevolent actors for

²⁵ As an example, the European Research Council has funded over 150 cutting edge AI projects by Europe's leading researchers in areas such as deep learning, neural networks, prediction, machine translation, natural language processing, computer vision, robotics, artificial agents and medical imaging as well as governance and standards.

criminal activities or terrorism, and on the other hand to deploy AI tools and solutions in support of law enforcement agencies to better prevent, detect and investigate criminal activities and terrorism.²⁶

*For **AI made in Europe** one key principle will be “**ethics by design**” whereby ethical and legal principles, on the basis of the General Data Protection Regulation, competition law compliance, absence of data bias are implemented since the beginning of the design process. When defining the operational requirements, it is also important to take into account the interactions between humans and AI systems. The Commission will explore how to introduce an “ethics by design” principle in relevant calls for proposals under the research programme.*

*Another key principle will be “**security by design**”, whereby cybersecurity, the protection of victims and the facilitation of law enforcement activities should be taken into account from the beginning of the design process.*

In addition, the Commission will work towards reinforcing European AI excellence centres, establishing world reference testing facilities and accelerating the uptake of AI through Digital Innovation Hubs, making sure that Europe benefits from the results of the research activities.

The Commission will maintain a geographical balance on its efforts to reinforce Digital Innovation Hubs and will encourage geographical coverage in the networks of excellence centres and testing facilities and will encourage complementarities with Cohesion Policy investments. While Europe has undeniable strengths with its many leading research centres, joining forces is essential in order to be competitive at the global level. Europe will scale up national research capacities and reach a critical mass through tighter **networks of European AI research excellence centres**²⁷. The objective is to foster cooperation among the best research teams in Europe, joining forces to tackle more efficiently major scientific and technological challenges in AI and mobilise industry to be integrated in and find synergies with the research teams.

Establishing world-reference testing facilities²⁸: An important step in bringing technology to market relates to experimenting and testing state-of-the art technology in real-world environments. To optimise investment and avoid duplication or competing efforts, a limited number of AI specialised large-scale reference sites should be developed and be opened to all actors across Europe.

Examples of such testing facilities include the cross-border testing of connected and autonomous driving, test sites for autonomous shipping and the creation of data spaces. The Commission and the Member States will identify the need for new large-scale testing facilities for the latest AI technologies in key areas such as mobility, healthcare, manufacturing, agro-food or security. These testing facilities may include regulatory sandboxes (i.e. areas where

²⁶ This would also enable businesses to enhance their security capacity. The Commission will discuss with the Secure Societies Programme Committee the inclusion of relevant actions in the Horizon 2020 Work Programme for 2020.

²⁷ An AI Excellence Centre is a research centre with a strong expertise in AI. The main purpose of such centres is to advance progress in specific fields of science and technology.

²⁸ A Reference Testing and Experimentation Facility is a technology infrastructure that has specific expertise and experience of testing mature technology in a given sector, under real or close to real conditions (smart hospital, clean rooms, smart city, experimental farm, corridor for connected and automated driving, etc.).

regulation is limited or favourable to testing new products and services) in selected areas where the law provides regulatory authorities with sufficient leeway, relaxing specific legal and regulatory requirements for the duration of the sandbox.

Accelerating AI take-up through Digital Innovation Hubs: Equally important will be fostering the uptake of AI in the wider economy, in particular by SMEs. This includes transmitting knowledge and scientific advances generated in European AI research excellence centres as well as technologies validated in the testing facilities above. **Digital Innovation Hubs (DIH)** can help ensure that every company, small or large, high-tech or not, as well as the public sector, can grasp the digital opportunities. With technical universities or research organisations at the core, DIHs act as one-stop-shops where companies and the public sector can get access to technology, testing and technical support, financing advice, market intelligence and networking opportunities. More specifically, in the area of AI, DIH can help SMEs and public administrations identify necessary datasets, develop algorithms, train AI and they can link to computing facilities building on the “AI-on on-demand” platform. They can help training professionals from the SMEs in the use of AI solutions and advise on existing financial support. They link both to the research excellence centres and to available testing facilities.

Similarly, the current eleven nodes of the Digital Knowledge and Innovation Communities of the European Institute of Technology & Innovation bring together prominent actors of the digital sectors in targeted regions.

i) Strengthening research excellence through networks of European AI research excellence centres:

- ✓ **In 2019, Member States will map national AI research excellence centres** and their competences and further support their EU-wide cooperation and networking through national programmes.
- ✓ **The Commission** plans to fund **networks of AI research excellence centres** with EUR 50 million in 2020 through Horizon 2020, supporting collaborative research addressing industrial and scientific challenges identified by such networks in joint research agendas.
- ✓ **Member States** are encouraged to mobilise their industry to be integrated in or develop synergies with the **networks of AI research excellence centres**.

(ii) Establishing world-reference testing facilities

- ✓ **2018-20, Member States and the Commission will:**
 - ✓ Building on a first set of 5G cross-border corridors for connected and autonomous driving²⁹, **support additional testing corridors** with up to EUR 30 million in 2020 through Horizon 2020.
 - ✓ Work on the development of **platforms and large-scale pilots** integrating AI elements in areas such as energy, healthcare, manufacturing, geo-information and agriculture. For 2019-20, the Commission will make available EUR 160 million from

²⁹ <https://ec.europa.eu/digital-single-market/en/news/new-5g-cross-border-corridors-connected-and-automated-mobility-baltics-will-allow-testing>

Horizon 2020.

- ✓ In 2019 and 2020, under the **ECSEL Joint Undertaking**³⁰, AI and data analytics will be integrated in lighthouse initiatives in manufacturing, mobility and personalized medicine, with a total budget of around EUR 200 million, from components up to full systems.

In total, the Commission will make available around EUR 390 million in developing platforms and large-scale pilots in the period 2019-20; it is expected that this investment will be matched with close to EUR 200 million from Member States and EUR 550 million from the private sector.

Beyond 2020,

- ✓ **Under the Digital Europe Programme, the Commission envisages making available around EUR 1.5 billion to establish world-leading testing and experimentation sites for AI-powered products and services throughout Europe.** These test sites will be identified and developed in close collaboration with Member States in 2019, who will provide a mapping of existing national test sites and will cover the whole AI supply chain from components (neuromorphic computing and quantum technologies) up to integrated applications in areas like health, mobility, energy, security, safety and industrial production.
- ✓ **Member States** will be encouraged to match the investments in the Digital Europe Programme so that an overall investment volume of EUR 3 billion is available. The use of other sources of funding, such as the European Regional Development Fund, will also be encouraged.

(iii) Accelerating AI take-up through Digital Innovation Hubs (DIH)

- ✓ **In 2019 Member States are invited to strengthen their networks of Digital Innovation Hubs** with a focus on supporting their local SME communities in the digital transformation. Member States will be asked to identify the DIHs that have AI competence.
- ✓ **In 2019 and 2020 the Commission will make available more than EUR 100 million for Digital Innovation Hubs in selected, AI-relevant areas** (Big Data, Smart Manufacturing). This includes activities for regions where currently few Digital Innovations Hubs exist, such as in the EU13 countries.
- ✓ In addition, the European Institute of Technology (EIT) activities will contribute to the adoption of AI by the public and private sectors. Between 2018 and 2020, the Union will invest in activities to support EIT Digital and its network of nodes across the Union. This investment will focus on the digital transformation of industries, cities, health, infrastructure, and finance, notably to adapt to the opportunities offered by AI.
- ✓ **Beyond 2020, funding from the Digital Europe Programme is proposed to contribute to establishing Digital Innovation Hubs in every Member State ensuring a broad**

³⁰ The ECSEL Joint Undertaking is a tripartite model of co-investment of the Commission, Member States and the Industry to support research and innovation, including large scale demonstrations and pilots in the areas of microelectronics, small system integration and embedded software with special focus on integration projects.

geographical representation (possibly with, on average, one in every NUTS2 region³¹). It is envisaged that the Union will provide up to EUR 900 million to support the development of these hubs, an amount to be matched by similar amounts from Member States. Horizon Europe is expected to enable Digital Innovation Hubs to engage further in digital transformation experiments and support up to 10 000 SMEs throughout Europe.

D. Skills and life-long learning

Talent is one essential ingredient for AI to be developed and used. AI and digitalisation are rapidly transforming society and the economy as a whole, including the work environment. In Europe, there is a significant and persistent ICT skills gap. Demand for skills in emerging areas such as AI are particularly acute and the problem is growing as the offer lags behind the market. Almost all Member States are facing shortages of ICT professionals, including in the area of AI³². The current offer of specialised higher education programmes is limited and not equally available in all Member States.³³

Poor general technical knowledge in the broader population hampers the accessibility and uptake of AI-based solutions. Access to the necessary skills should be fostered in primary and secondary schools, although training of teachers remains an important challenge. Fast track retraining programmes need to be designed in order to enable the population to gain experience in AI. Technology like Massive Open Online Courses (MOOCs) could be used to scale up learning. The topic of AI needs to become part of non-technical study-programmes through formal and informal education, so as to provide the future workforce with knowledge needed to operate and navigate in a working environment where AI will be part of the day-to-day tasks.

Beside ICT skills, other skills are equally important for a human-centred AI development. Ethics and other non-STEM skills are equally important and should be part of the talent fostering chapter of AI national and international strategies. Furthermore, re- and up-skilling should be accompanied by the modernisation of labour market and social policies in order to better cope with more frequent labour market transitions.

Difficulties in attracting and retaining talent in Europe contribute to the skills gap. Talented researchers and promising start-ups frequently receive attractive offers from abroad. In 2017, 38% of Silicon Valley's population were foreigners who had entered the US to fill a specific job in the tech industry, and 8% of them were European³⁴. Action is in particular needed to attract and retain the best talent in Europe and to create a competitive environment. Closer cooperation with industry will help to ensure the relevance of the learning content to labour market needs.

³¹ Nomenclature of territorial units for statistics, NUTS 2 are defined as basic regions for the application of regional policies.

³² There is a lack of more than 80,000 professionals for data handling and management (1 per 20 scientists) http://www.pocbigdata.eu/monitorICTonlinevacancies/general_info/

³³ "In 2018, about two third of the EU Member States have less than 10 Master's programmes strongly focusing on AI. While AI modules are becoming relatively more common across different educational domains, still only one third of the EU Member States have more than 20 Master's programmes including at least one AI module". López-Cobo et al. (2018), Academic offer and demand for advanced profiles in the EU. Artificial Intelligence, High Performance Computing and Cybersecurity. Joint Research Centre Scientific Report.

³⁴ <https://jointventure.org/images/stories/pdf/index2018.pdf>

The Commission adopted a Digital Education Action Plan³⁵ to support technology-use and digital competence development in education. As announced in the April Communication, the European Institute of Innovation and Technology (EIT) is in the process of integrating AI across curricula in the education courses it supports, at Master and PhD level, and the Digital Opportunity Traineeships pilot (2018-20)³⁶ makes available internships in advanced digital skills for Erasmus students. The Blueprint on sectoral cooperation on skills supports the development of sectoral strategies to address the skills gaps³⁷, and improve multidisciplinary approaches by integrating elements of AI into other disciplines.

Mutual recognition of certification, including for studies in new disciplines such as AI, is important. In May 2018, the Commission made a proposal for a Council Recommendation³⁸ on promoting automatic mutual recognition of higher education and upper secondary education diplomas and the outcomes of learning periods abroad. The proposed Council Recommendation invites Member States to make a political commitment to take steps to introduce automatic recognition by 2025.

Talent, skills and life-long learning

- ✓ Skills and education are domains that fall to a large extent under national, and sometimes regional, competences. However, Union level action is important in order to share experiences and seize common opportunities.
- ✓ **Member States are encouraged to :**
 - ✓ **Exchange best practices** on (i) how to reinforce excellence and to retain AI talent in Europe and (ii) on the re- and upskilling of the current workforce (in 2019).
 - ✓ Exchange best practices on how to make full use of the possibilities offered by the EU **Blue Card system** to attract and retain highly-skilled AI professionals in the EU and accelerate its implementation and on how to facilitate AI-related entrepreneurship (by the end of 2019).
 - ✓ **Include the skills dimension in the national AI strategies** (by mid-2019) and map the national education offer, the skills needs (AI also needs to be part of/integrated in other disciplines such as law, human sciences, environment, health) and the training priorities for AI, paying particular attention to inclusion and to attracting more women to AI studies (by end 2020). Strategies should address the whole cycle of formal education, vocational training, higher education and post-docs. At the same time, more focus should be put on life-long learning, in order to allow also people already in employment to acquire and improve their skills related to AI.
 - ✓ Explore how **AI** could be incorporated into the curricula of programmes for secondary and tertiary education, including **vocational training**. Commission will issue a report with the support of the Member States on this by early 2020 and support model actions in selected regions.
- ✓ **the Commission will:**
 - ✓ **Include a component of common PhD programmes and post-PhD programmes in the call for strengthening the AI research excellence centres in 2020** with a focus

³⁵ COM(2018) 22

³⁶ <https://ec.europa.eu/digital-single-market/en/digital-opportunity-traineeships-boosting-digital-skills-job>

³⁷ <http://ec.europa.eu/social/BlobServlet?docId=16962&langId=en>

³⁸ COM(2018) 270

on industrial challenges. The ambition is to establish a unique and world-recognised brand for a European programme for industrially-oriented PhDs in AI and to keep researchers in Europe after they complete their PhDs. The Marie Skłodowska-Curie Actions³⁹ will contribute to this objective.

- ✓ Explore avenues to support the inclusion of AI modules in multi-disciplinary **Master programmes (e.g. in e-health, fintech, e-government) and in adult training programmes with the focus on people with higher education and work experience.**
- ✓ Member States and Commission will work together and develop material to be used in **awareness campaigns** on the benefits of AI.
- ✓ **Beyond 2020** the Commission proposes that the Union provides a total of EUR 700 million to support advanced skills (for AI, HPC and cybersecurity) as part of the Digital Europe Programme through
 - ✓ Master degrees,
 - ✓ On-the-job trainings and traineeships for young people and professionals needing to gain experience
 - ✓ Short-term trainings for the workforce to become knowledgeable in AI.
- ✓ Mainstream ethical principles developed by the Union in the training schemes and programmes proposed above.

E. Data: a cornerstone for AI - Creating a Common European Data Space

The current expansion of AI is fuelled by the availability of large data sets combined with increases in computing power and connectivity. Making secure, robust quality data available for a broad range of users across borders is a cornerstone of European policy. Openness to international data flows will continue to be ensured in full respect of the EU rules for the protection of personal data and in accordance with applicable legal instruments, including free trade agreements that the EU concludes with its partners and Commission adequacy findings as regards the level of personal data protection in third countries. Full implementation of sectoral legislation, improving access to and re-use of information (e.g. the INSPIRE⁴⁰ Directive) will provide the domain-specific data needed to support powerful AI applications for the public sector, for analytical or policy monitoring purposes⁴¹.

Where data relating to individuals is processed, the General Data Protection Regulation (GDPR) lays down the rules applicable to the collection, use and sharing of such personal data. Additionally, the recently adopted Regulation on the free flow of non-personal data further facilitates cross-border data flows across the Union as a cornerstone of the Digital Single Market. Consideration will also be given to blockchain-based, fully GDPR- and privacy-compliant solutions for sharing and providing access to data. Proposed rules⁴² on trading practices between online intermediation services such as market places, app stores, or

³⁹ <https://ec.europa.eu/research/mariecurieactions/>

⁴⁰ Directive 2007/2/EC

⁴¹ Cetl V., Tomas R., Kotsev A., de Lima V.N., Smith R.S., Jobst M. (2019) Establishing Common Ground Through INSPIRE: The Legally-Driven European Spatial Data Infrastructure. In: Döllner J., Jobst M., Schmitz P. (eds) Service-Oriented Mapping. Lecture Notes in Geoinformation and Cartography. Springer, Cham.

⁴² <https://ec.europa.eu/digital-single-market/en/business-business-trading-practices>

accommodation booking platforms, set the conditions for predictable and transparent data use amongst hosting services and their business users. Such measures are intended to bring further fairness and trust in business relations and valuable use of data in the online platforms ecosystem.

Action is needed to facilitate sharing of data held by public and private sectors by creating a common European Data Space⁴³: a seamless digital area with the scale that will enable the development of new products and services based on data. In particular, data generated and held by the public sector is often of very high quality and constitutes a major asset for European innovators and businesses.

For increased usage, data within a space should be made interoperable as much as possible, notably by agreeing on aiming for data formats that are open, FAIR, machine readable, standardised and documented, both in the interaction between public and private sectors, within sectors and across sectors⁴⁴.

The Public Sector Information (PSI) directive⁴⁵ sets the framework for the reuse of such data by businesses. Actions should focus on making datasets more easily accessible in practice, notably to start-ups and SMEs and facilitating aggregation. Of particular importance are design and implementation of interoperable data and meta-data formats as well as the deployment of standardised Application Programming Interfaces (APIs) based on the European Interoperability Framework (EIF)⁴⁶.

These actions will complement efforts by the Member States to promote the accessibility, interoperability and re-usability of data in the sectors of high relevance for AI, such as health⁴⁷, (see points below), environment, mobility, security, migration and a sustainable and circular bioeconomy and food system.

Data

AI for areas of public interest: Member States are encouraged to cooperate with the Commission to:

- ✓ **Identify public data sets** to be made more openly reusable across the Union, especially those suitable for training AI applications. This may be supported by a mandate to establish a list of high-value datasets as provided for by the recast proposal of the Public Sector Information Directive, currently under negotiation.

⁴³ <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN>

⁴⁴ See practices European Open Science Cloud (EOSC) It will foster best practices of global data findability and accessibility (FAIR data), <https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>

⁴⁵ <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>

⁴⁶ APIs facilitate sharing and use of government data between Member States through developing common approaches that should be promoted through experiments and applied research on APIs A study (APIs4DGov) has been started in 2018 by the Joint Research Centre. It aims to be participatory and benefit from work performed in European public administrations at all levels and contributes to the implementation of the European Interoperability Framework and the Building Block approach adopted in the CEF telecom programme. <https://ec.europa.eu/digital-single-market/en/news/new-study-digital-government-apis-apis4dgo-project>

⁴⁷ In the health field, such data sharing is for instance vital for the joint innovation investments prepared under the thematic Smart Specialisation Platform "Artificial Intelligence and Human Machine Interface. With Emilia-Romagna (IT), Autonomous Province of Trento (IT), Baden-Württemberg (DE), Navarra (ES), North Brabant (NL) and Örebro Län (SE) participating. See: <http://s3platform.jrc.ec.europa.eu/artificial-intelligence>

- ✓ Invest together in the necessary tools to facilitate access to, connectivity, interoperability and aggregation of public data including the development of relevant Application Programming Interfaces (APIs) for accessing data of dynamic nature. The action will support the definition and application of data and metadata standards in close cooperation with relevant stakeholders (e.g. European standardisation bodies). The Commission is planning to make available up to EUR 100 million from Horizon 2020 (H2020) and Connecting Europe Facility (CEF).
- ✓ Support the development and operations of a data infrastructure to enable the management and sharing of data in real-time and the experimentation through a data sandbox of data-driven AI-powered services, for governments and public administrations at large, including for secure cross-border trans-European IT systems. Such services are provided through the Public Open Data service infrastructure funded under CEF, which already supports the deployment of the European Data Portal⁴⁸ for facilitating discovery and access to European public data assets including those from Member States Open Data initiatives at the national, regional and local levels.
- ✓ Ensure the further development of the **European Open Science Cloud** as a key asset for the best use of AI in science and technology and in applications ranging from advanced medicine to environment and climate change. Include training and testing of data related algorithms to maximise the benefits of Open Data.
- ✓ Support the development of **blockchain-based and other secure solutions for giving access to data and securing the integrity of data**. The Commission is planning to make available EUR 27 million under H2020 in this area.
- ✓ **Health** is particularly primed to benefit from AI. Patient information, medical records, diagnostic results, and clinical studies are just a few of the data sources available in healthcare. The Commission proposes, as a matter of priority, to concentrate on two big projects:
 - (i) Following commitments made by 19 Member States towards building a research cohort of at least 1 million sequenced genomes accessible in the EU by 2022⁴⁹, the Commission will support an initiative on linking genomics repositories. The Commission will also support building rare disease registries. Organisational, regulatory, security, ethical and technical compliance and interoperability will be duly taken into account. This will be instrumental in allowing AI technologies to be studied, developed and trialled with a view to identifying new knowledge, support clinical research and decision making.
 - (ii) In 2020, the Commission will support – in coordination with Member States – the development of common database of health images initially dedicated to the most common forms of cancer (anonymised, and based on data donorship by patients). The work must meet all necessary organisational, regulatory, security, ethical and technical requirements. It will be combined with relevant AI tools to improve diagnosis, treatment and follow-up.

Overall, the initial EU contribution to these initiatives will be around EUR 35 million from the Horizon 2020 programme. Member States are encouraged to match this amount.

⁴⁸ <https://www.europeandataportal.eu/en/homepage>

⁴⁹ <https://ec.europa.eu/digital-single-market/en/news/eu-countries-will-cooperate-linking-genomic-databases-across-borders>

These initiatives and investments will be a basis for larger scale, common health data space with possible support, starting in 2021, from the Digital Europe Programme.

- ✓ **Geo-information /Earth Observation:** The EU's Copernicus Programme is the biggest provider of earth observation and monitoring information worldwide. Copernicus has adopted a free, full and open data policy and has launched advanced Data and Information Access Services (DIAS), which co-locate a massive amount of structured data and computing capacities. Building on this, the Commission proposes to develop and deploy AI capacities using Copernicus data and infrastructures to foster geo-location based services for climate, agriculture, air quality, emissions, the marine environment, water management, security and migration monitoring and citizen science⁵⁰. It will also launch initiatives to support AI powered exploitation of Earth observation data and information in the public and private sectors.
- ✓ **Linguistic data:** The Commission language resources used for the deployment of AI enabled automated translation and natural language processing services are among the most downloaded datasets on the European Data Portal. To further improve such services, the Commission is planning to make available an additional EUR 10 million from the Connecting Europe Facility to collect further language resources for languages less represented on the web.

Industrial data platforms: The Commission has already launched Research and Development actions on platforms for secure and controlled sharing of proprietary data under Horizon 2020 including industrial data spaces and personal data spaces⁵¹. On the basis of the Commission Communication "Towards a common European data space"⁵², a set of guidelines⁵³ with the aim of providing a toolbox for sharing of data by holders, users, or both was published. Building on this, the Commission will:

- ✓ In 2019, support strategic, next-generation digital industrial platforms through large scale federating projects with an investment of EUR 50 million from Horizon 2020 programme.
- ✓ Member States are encouraged to connect existing and planned national investments in platforms with EU level activities in order to ensure scale up and interoperability.
- ✓ Beyond 2020, the Commission proposes that the Union co-invests with Member States and the private sector in the creation of a common European Data Space that makes data easily available for re-use to innovators, businesses and public sector for up to EUR 1 billion through the Digital Europe Programme, as part of its AI section.
- ✓ Particular attention will be given to developing local ecosystems at the regional and sub-regional levels bringing together local businesses and SMEs, public administrations, training centres, Digital Innovation Hubs and technology infrastructures developing and sharing algorithms trained on high quality local data to address local problems. In this way upskilling and training are linked to local data spaces to foster innovation.

⁵⁰ The Commission has already set up a Framework Partnership Agreement with the Member States to co-finance the use and mainstreaming of space data from Copernicus and Galileo in combination with data available from Member States and other sources.

⁵¹ ICT-13-2018-2019 of [H2020 work programme 2018-2020 Information and Communication Technologies](#)

⁵² <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2018:0232:FIN>

⁵³ <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1539766272141&uri=CELEX%3A52018SC0125>

Support centre for data sharing: The Commission will launch by mid-2019 a Support Centre for data sharing, to propose model contracts for the sharing of private sector data, provide practical advice, best practices and methodologies for data sharing and data analytics to all European actors in the data economy.

European High-Performance Computing initiative (EuroHPC): The Commission and Member States will work together on the timely implementation of the EuroHPC initiative in order to develop a pan-European supercomputing infrastructure, which will be critical for AI.

F. Ethics by design and regulatory framework

An appropriate and predictable, ethical and regulatory framework that relies on effective safeguards for the protection of fundamental rights and freedoms is vital for citizens to trust AI and for companies, in need of investment security, to take up new business opportunities. Spearheading the ethics agenda, while fostering innovation, has the potential to become a competitive advantage for European businesses on the global marketplace. In addition, the emerging uptake of AI in the public sector is also bound to raise similar ethical and fundamental rights concerns which need to be addressed up-stream.

As announced in the “Artificial Intelligence for Europe” strategy, the Commission tasked a High-Level Expert Group on AI to draft AI ethics guidelines⁵⁴. A first draft of these guidelines will be ready in December 2018, and a final version is expected for March 2019, after wide consultation through the European AI Alliance⁵⁵. One key principle will be “ethics by design” by which ethical principles are embedded in AI products and services right at the beginning of the design process.

It is important that legislation offers the right framework for AI-driven innovation and uptake of AI solutions, while addressing possible risks raised by the use of, and interactions with the technology, including cybersecurity concerns. This means to provide for “cybersecurity” in the sense of preventing abuse (e.g. hacking or manipulation of the AI algorithms or manipulation of the data processed by the AI algorithm), as well as for the inclusion of mechanisms to ensure the safety of consumers and effective redress to victims in case of damage and to facilitate investigations if the AI system is compromised. AI cyber security requirements would need to be specified and should benefit from the certification scheme under the proposed European Cybersecurity Certification Framework⁵⁶. Moreover, in case of businesses acting in security relevant fields (e.g. financial institutions, producers of radio-active materials, etc.) the use of certain AI products and processes serves public interest therefore their use may be made compulsory.

An adequate safety and liability framework guaranteeing a high level of safety and effective redress mechanisms for victims in case of damages is essential for building trust in AI.

⁵⁴ <https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>. This work builds on the work of the European Group on Ethics in Science and New Technologies. http://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

⁵⁵ <https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

⁵⁶ <https://ec.europa.eu/digital-single-market/en/eu-cybersecurity-certification-framework>

Moreover, with the appropriate safeguards in place, regulatory sandboxes, and other methods for policy experimentation and development, can play an important role to encourage AI based innovation for areas where the law provides regulatory authorities with a sufficient margin of manoeuvre. In 2019, a focus will be put on assessing whether the regulatory framework in Europe is fit for purpose for AI-enabled technologies in general and for connected and automated driving in particular.

Innovation Deals⁵⁷ can serve as tools within the boundaries of existing legislation for assessing regulatory barriers linked to the development and deployment of AI. Innovation Deals are voluntary cooperation agreements between the EU, innovators, and national, regional and local authorities. The objective of an Innovation Deal is to gain an in-depth understanding of how an EU rule or regulation works in practice. If the rule or regulation is found to be an obstacle to innovations, the deal will make it visible and feed into possible further action.

Other important elements for creating an integrated European home market for AI-enhanced products, services and applications are, for example, data protection and privacy⁵⁸, consumer protection and compliance with competition law by design. In addition, important considerations for the development and uptake of AI in particular in areas with a high societal and policy stake are related to fairness, transparency and accountability of algorithmic decision-making and related governance models⁵⁹ and impact of AI on human behaviour⁶⁰.

Finally, intellectual property (IP) issues should also be explored, to ensure that the related regulatory framework properly addresses a number of challenges that are specific to AI, and is thus able to promote its development effectively.

Ethics by design and the regulatory framework

- ✓ The Commission tasked a High-Level Expert Group on AI to draft **AI ethics guidelines**. A final version is expected for March 2019.
- ✓ The Commission will firmly respect and anchor the “ethics by design” principle in its calls for proposals that deal with artificial intelligence.
- ✓ **The Commission, taking into account the input from the Member States, is assessing whether and to what extent the existing legislation is fit for purpose** to allow for the new opportunities and tackle the challenges raised by AI, taking into account the policy

⁵⁷ https://ec.europa.eu/info/research-and-innovation/law-and-regulations/identifying-barriers-innovation_en

⁵⁸ Building on the existing regulatory framework such as the General Data Protection regulation which entered into application in May 2018.

⁵⁹ Regulatory approaches in the [General Data Protection Regulation](#), the [Markets in Financial Instruments Directive](#), the [proposal for a regulation on promoting fairness and transparency for business users of online intermediation services](#), the [Commission Recommendation on measures to effectively tackle illegal content online](#) proposal, all set a precedent and models for meaningful transparency and risk assessment and risk management. The Commission is further exploring (supported by the European Parliament AlgoAware pilot project) areas of concern and opportunities in algorithmic decision making in the online platforms environment, where different approaches to meaningful transparency, fairness and accountability can enhance trust. The analysis is carefully considering the balance set by the regulatory framework in place and the enforcement of recent rules, as well as technical, market and societal developments, and exploring policy and regulatory tools.

⁶⁰ The Joint Research Centre HUMAINT project aims to understand the impact of AI on human behaviour, with a focus on cognitive and socio-emotional capabilities and decision making (<https://ec.europa.eu/jrc/communities/community/humaint>).

recommendations proposed by the AI High-Level Expert Group.

- ✓ By mid-2019, the Commission will publish a report on potential gaps in and orientations for the safety and liability framework for AI.
- ✓ The Commission is ready to support stakeholders on the application of EU rules to the development and deployment of AI, for example in the fields of competition and State aid, where necessary and in the appropriate format.
- ✓ **In 2019 Member States and Commission will discuss the creation of environments⁶¹ that are conducive to innovation such as regulatory sandboxes⁶², and public testing arrangements for specific AI applications in Europe.** Following these discussions Member States will be encouraged to create such environments and public testing arrangements for AI solutions by the end of 2020. To this end, Member States will be encouraged to establish a one-stop-shop for companies that are developing AI applications to discuss the specific needs for the creation of such environments and testing arrangements.

G. AI for the Public Sector

AI applications can contribute to better public services in a variety of ways, for example by enabling smarter analytical capabilities and better understanding of real-time processes (e.g. population, economic, environmental and climate change) in economy, society and environment, including the detection of criminal activities such as tax fraud and money laundering.

AI-enabled solutions can deliver shorter and richer feedback loops for all levels of governance, providing an opportunity to speed up, improve the efficiency and effectiveness of service delivery. They have the potential to:

- increase the quality and consistency of services delivered,
- improve the design and implementation of policy measures,
- allow more efficient and targeted interventions,
- enhance the efficiency and effectiveness of public procurement, and
- strengthen security, identity management, improve health and employment services.

For beneficiaries of public support, AI-enabled decision can simplify the relationship between authorities and beneficiaries through the integration of wider public interest or regulatory considerations in daily decision-making (through targeted communication, behavioural nudges etc.)

AI can improve citizen-government interaction through conversational systems (including digital assistants and government chatbots), multilingual services and automated translation.

⁶¹ While regulatory sandboxing is a powerful and sometimes needed tool, in other circumstance innovation can be supported with softer approaches such as innovation centers and policy labs that advice and participate more or less hands on.

⁶² For selected areas where the law provides regulatory authority with a sufficient margin of manoeuvre.

Efforts are also under way to apply AI in the social and health sector, to support doctor's decision making, or to support early recognition of young people's marginalisation.⁶³

As outlined above, concrete measures are proposed to open up public sector data for the use by AI applications in areas of public interest such as for medical images or genomics.

Member States are encouraged to engage in peer-learning with other Member States, especially with respect to regulatory sandboxes and testing arrangements.

AI for the public sector

- ✓ Building on and scaling up current investments engaged under the Digital Service Infrastructure actions in the current Connecting Europe Facility Programme and ISA² Programme, the Union will progressively increase efforts for the take-up of AI in areas of public interest, such as healthcare, transport, security and education. Beyond 2020 under the proposed Digital Europe Programme, Member States and the Union will co-invest in the full deployment of EU-wide AI-enabled services in areas of public interest.
- ✓ **In 2019, Member States and the Commission are planning to engage in peer-learning** and EU-wide exchange of best practices, experiences and data⁶⁴. They will work together to produce an overview of the relevant applications already in place in the Member States, their impact and added value in support of public service delivery. The Commission is also ready to assist public buyers, for instance by setting up an assistance hub for purchasing AI and cybersecurity solutions. One concrete example is the framework of the European Public Employment Services (PES) Network, where national PES will exchange best practices on AI in service delivery, in matching and automatic processes.⁶⁵
- ✓ **Member States are encouraged to work with the Commission to identify areas for joint procurement of AI solutions**, leading to efficiency gains and better value for money. One concrete example are AI-powered self-healing systems in cybersecurity where the combined buying power of the Union and all Member States can facilitate the development and scaling up of EU-developed solutions. The aim is to issue a joint report by mid-2019 to describe areas where joint procurement is envisaged. Beyond 2020 the Commission proposes to begin work under the new Digital Europe Programme.
- ✓ **In 2019, the Commission is planning to offer eTranslation**, the AI-enabled automatic translation service developed under the Connecting Europe Facility, to public administrations in Member States. The Commission proposals for the Horizon Europe and the Digital Europe programmes foresee investments in further developing natural language processing services and tools to enhance multilingualism in the public sector.

⁶³ <https://www.sitra.fi/en/news/artificial-intelligence-based-systems-help-achieve-better-services-cost-savings-social-health-sector/>

⁶⁴ The European Commission is walking the talk and deploying actions in the AI@EC action plan to support deployment of AI solutions in Tran-European Systems (TES) supporting key EU policy areas in cooperation with Member States.

⁶⁵ A second example is the activity foreseen under AI Watch to develop a methodology to identify risks and opportunities, drivers and barriers of the use AI in public service provision AI Watch will provide an overview of the use and added value of AI tools supporting public service delivery by looking at most relevant examples in prioritized public services. Based on the results of the analysis the task will draw up recommendations on the way forward for further development of AI based systems and solutions in government. Another example is the co-innovation initiatives between the European Commission and Member States on the deployment of AI-based solutions for smart public services.

- ✓ **In 2020**, in line with the Tallinn Ministerial Declaration on eGovernment, Member States, with the support of the Commission, and in particular exploiting the role of the DIHs proposed under the next multi-annual financial framework, are encouraged to devote resources for experimentation with AI-enabled services to understand better the added value and potential impact of AI-enabled public services and policy making. AI-based solutions will also benefit the justice⁶⁶ and law enforcement sectors. Another promising public application sector is the monitoring and enforcement of single market rules for goods, services and people.
- ✓ **Member States and the Commission** are planning to continue to develop integrated earth observation and AI machine learning solutions to support evidence-based policy making, implementation and monitoring in areas such as climate change, environmental protection, agriculture, urban development, disaster response, migration, infrastructure monitoring.

H. International cooperation

With AI currently being discussed all around the world and in many international fora such as the UN, OECD, the G7 or the G20⁶⁷, international outreach is crucially important. The development of AI will benefit from international cooperation, in particular among advanced countries with research and innovation strengths and investments in AI. The joint development of international standards will facilitate AI deployment and acceptance. The Union will promote the AI ethics guidelines internationally and open up a dialogue and cooperation with all non-EU countries and stakeholders from third countries that are willing to share the same values.

To make these efforts a success, however, Member States and the Union should attempt to align bilateral outreach efforts related to AI between individual Member States and third countries and pool their efforts pushing for a responsible development of AI at the global level. The Union needs to speak with one voice to third countries and the world at-large on this topic. In synergy with activities of the Member States, the EU should also seek alliances with stakeholders - tech companies, academia and other parties - to engage in a multi-stakeholder alliance at the global level for responsible AI.

Moreover, the Union will organise an international ministerial meeting on AI in the first half of 2019 with the aim of forging a global consensus on the ethical implications of AI. Furthermore, the EU is using its Foreign Policy Instrument to engage with international partners on regulatory and ethical matters. Some Member States propose an intergovernmental process similar to the panel on Climate Change. As regards the international security dimension, the AI policy will build on the work of the High Representative in the Global Tech Panel and within the United Nations and other multilateral fora.

Finally, the Union will contribute its expertise and dedicated financial means to anchor AI more firmly in **development policy**. Artificial intelligence is destined to make impactful contributions to global challenges as well as development policy. AI-powered precision farming, for example, promises to reduce pesticides, fertiliser and water consumption, making

⁶⁶ E.g. solutions based on predictive justice implementations and LegalTech applications.

⁶⁷ ISO/IEC JTC1/SC 42

it an ideal technology to help a growing population in the developing world. AI can also be used to model weather, climate and other natural phenomena so that local populations can e.g. be warned in case of extreme weather conditions or imminent disasters and adapt in advance. AI and digital technologies can underpin affordable high-tech solutions including for people in precarious circumstances, while respecting ethical and privacy issues.

International Cooperation

- ✓ The Union will reach out to its **international partners and promote AI ethics guidelines** internationally in the course of 2019.
- ✓ Member States and the Union are encouraged to align their international outreach efforts on AI and ensure that Europe sends consistent messages to the world.
- ✓ The Union will organise **an international ministerial meeting on AI in 2019** with the aim of forging a global consensus on the ethical implications of AI.
- ✓ The Union will contribute its expertise and dedicated financial means to anchor AI more firmly in **development policy**. A particular focus will be given to Southern Mediterranean countries and Africa.

Links:

Commission Communication 'Artificial Intelligence for Europe'

<https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>

Declaration of cooperation on Artificial Intelligence

<https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence>)

AI Alliance

<https://ec.europa.eu/digital-single-market/en/european-ai-alliance>

Brussels, 25.4.2018
COM(2018) 237 final

**COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN
PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN
ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE
REGIONS**

Artificial Intelligence for Europe

{SWD(2018) 137 final}

1. INTRODUCTION – EMBRACING CHANGE

Artificial intelligence (AI) is already part of our lives – it is not science fiction. From using a virtual personal assistant to organise our working day, to travelling in a self-driving vehicle, to our phones suggesting songs or restaurants that we might like, AI is a reality.

Beyond making our lives easier, **AI is helping us to solve some of the world's biggest challenges: from treating chronic diseases or reducing fatality rates in traffic accidents¹ to fighting climate change or anticipating cybersecurity threats.**

In Denmark, AI is helping save lives by allowing emergency services to diagnose cardiac arrests or other conditions based on the sound of a caller's voice. In Austria, it is helping radiologists detect tumours more accurately by instantly comparing x-rays with a large amount of other medical data.

Many farms across Europe are already using AI to monitor the movement, temperature and feed consumption of their animals. The AI system can then automatically adapt the heating and feeding machinery to help farmers monitor their animals' welfare and to free them up for other tasks. And AI is also helping European manufacturers to become more efficient and to help factories return to Europe.²

These are some of the many examples of what we know AI can do across all sectors, from energy to education, from financial services to construction. Countless more examples that cannot be imagined today will emerge over the next decade.

Like the steam engine or electricity in the past, AI is transforming our world, our society and our industry³. Growth in computing power, availability of data and progress in algorithms have turned AI into one of the **most strategic technologies of the 21st century**. The stakes could not be higher. **The way we approach AI will define the world we live in.** Amid fierce global competition, **a solid European framework is needed.**

What is artificial intelligence?

Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals.

AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).

We are using AI on a daily basis, e.g. to translate languages, generate subtitles in videos or to block email spam.

Many AI technologies require data to improve their performance. Once they perform well, they can help improve and automate decision making in the same domain. For example, an AI system will be trained and then used to spot cyber-attacks on the basis of data from the concerned network or system.

¹ It is estimated that around 90% of road accidents are caused by human errors. See Commission's report on Saving Lives: Boosting Car Safety in the EU (COM(2016) 0787 final).

² Why AI is the future of growth, Accenture, 2016. The economic impact of the automation of knowledge work, robots and self-driving vehicles could reach between EUR 6.5 and EUR 12 trillion annually by 2025 (including improved productivity and higher quality of life in ageing populations). Source: Disruptive technologies: Advances that will transform life, business, and the global economy, McKinsey Global Institute, 2013.

³ AI is part of the Commission's strategy to digitise industry (COM(2016) 180 final) and a renewed EU Industrial Policy Strategy (COM(2017) 479 final).

The European Union (EU) should have a **coordinated approach** to make the most of the opportunities offered by AI and to address the new challenges that it brings. **The EU can lead the way in developing and using AI for good and for all**, building on its values and its strengths. It can capitalise on:

- **world-class researchers, labs and startups**. The EU is also strong in **robotics** and has **world-leading industry**, notably in the transport, healthcare and manufacturing sectors that should be at the forefront of AI adoption;
- the **Digital Single Market**. Common rules, for example on data protection and the free flow of data in the EU, cybersecurity and connectivity help companies to do business, scale up across borders and encourage investments; and
- a **wealth of industrial, research and public sector data** which can be unlocked to feed AI systems. In parallel to this Communication, the Commission is taking action to make data sharing easier and to **open up more data – the raw material for AI – for re-use**. This includes data from the public sector in particular, such as on public utilities and the environment, as well as research and health data.

European leaders have put AI at the top of their agendas. On 10 April 2018, 24 Member States⁴ and Norway committed to working together on AI. Building on this **strong political endorsement**, it is time to make significant efforts to ensure that:

- **Europe is competitive in the AI landscape**, with bold investments that match its economic weight. This is about supporting research and innovation to develop the next generation of AI technologies, and deployment to ensure that companies – in particular small and medium-sized enterprises which make up 99% of business in the EU – are able to adopt AI.
- **No one is left behind in the digital transformation**. AI is changing the nature of work: jobs will be created, others will disappear, most will be transformed. Modernisation of education, at all levels, should be a priority for governments. All Europeans should have every opportunity to acquire the skills they need. Talent should be nurtured, gender balance and diversity encouraged.
- **New technologies are based on values**. The General Data Protection Regulation will become a reality on 25 May 2018. It is a major step for building trust, essential in the long term for both people and companies. This is where the **EU's sustainable approach to technologies** creates a competitive edge, by embracing change on the basis of the Union's values⁵. As with any transformative technology, some AI applications may raise new ethical and legal questions, for example related to liability or potentially biased decision-making. The EU must therefore ensure that AI is developed and applied in an appropriate framework which promotes innovation and respects the Union's values and fundamental rights as well as ethical principles such as accountability and transparency. The EU is also well placed to lead this debate on the global stage.

This is how the EU can make a difference – and be the champion of **an approach to AI that benefits people and society as a whole**.

⁴ Austria, Belgium, Bulgaria, Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, United Kingdom.

⁵ Article 2 of the Treaty on EU: "The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities". The Member States share a "society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail."

Launching a European initiative on AI

In May 2017, the Commission published its mid-term review of the Digital Single Market strategy⁶. It highlighted the importance of building on Europe's scientific and industrial strengths, as well as on its innovative startups, to be in a leading position in the development of AI technologies, platforms, and applications.

The European Council of October 2017 stated that the EU needs a sense of urgency to address emerging trends such as AI "while at the same time ensuring a high level of data protection, digital rights and ethical standards" and invited "the Commission to put forward a **European approach to artificial intelligence**".⁷ The European Parliament made wide-ranging recommendations on civil law rules on robotics and the European Economic and Social Committee has also issued an opinion on the topic.⁸

This Communication sets out a European initiative on AI, which aims to:

- **Boost the EU's technological and industrial capacity and AI uptake across the economy**, both by the private and public sectors⁹. This includes investments in research and innovation and better access to data.
- **Prepare for socio-economic changes** brought about by AI by encouraging the modernisation of education and training systems, nurturing talent, anticipating changes in the labour market, supporting labour market transitions and adaptation of social protection systems.
- **Ensure an appropriate ethical and legal framework**, based on the Union's values and in line with the Charter of Fundamental Rights of the EU. This includes forthcoming guidance on existing product liability rules, a detailed analysis of emerging challenges, and cooperation with stakeholders, through a European AI Alliance, for the development of AI ethics guidelines.¹⁰

All this requires **joining forces**. Building on the approach set out in this Communication and the declaration¹¹ of cooperation signed by 24 Member States on 10 April 2018, the Commission will work with **Member States on a coordinated plan on AI**. The discussion will take place in the framework of the existing European platform of national initiatives to digitise industry, with the view to **agree this plan by the end of 2018**. The main aims will be to maximise the impact of investments at EU and national levels, encourage synergies and cooperation across the EU, exchange best practices and collectively define the way forward to ensure that the EU as a whole can compete globally.

In the coming weeks the Commission will issue a Communication on the future of connected and automated mobility in Europe and a Communication on the future research and innovation ambitions for Europe. AI will be a key element of these initiatives.

⁶ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2017:228:FIN>

⁷ <http://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf>

⁸ European Parliament resolution with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)); European Economic and Social Committee opinion on AI (INT/806-EESC-2016-05369-00-00-AC-TRA).

⁹ AI can significantly improve public services and contribute to the objectives set out in the Ministerial Declaration on eGovernment – the Tallinn Declaration (October 2017, <https://ec.europa.eu/digital-single-market/en/news/ministerial-declaration-egovernment-tallinn-declaration>). For example, the Commission will look into AI's potential to analyse large amounts of data and help check how single market rules are applied.

¹⁰ Building on the work of the European Group on Ethics in Science and New Technologies

¹¹ <https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence>

2. THE EU'S POSITION IN A COMPETITIVE INTERNATIONAL LANDSCAPE

Most developed economies recognise the game-changing nature of AI and have adopted different approaches which reflect their own political, economic, cultural and social systems.¹²

The government of the United States presented an AI strategy and invested around EUR 970 million in unclassified AI research in 2016. With its 'Next Generation Artificial Intelligence Development Plan', China is targeting global leadership by 2030 and is making massive investments.¹³ Other countries, such as Japan and Canada, have also adopted AI strategies.

In the United States and in China, large companies are significantly investing in AI and are exploiting large amounts of data.¹⁴

Overall, **Europe is behind in private investments** in AI which totalled around EUR 2.4-3.2 billion in 2016, compared with EUR 6.5-9.7 billion in Asia and EUR 12.1-18.6 billion in North America.¹⁵

It is therefore crucial that the EU continues its work to **create an environment that stimulates investments** and uses public funding to leverage private investments. To do so, the EU needs to **preserve and build on its assets**.

Europe is home to a **world-leading AI research community**, as well as innovative **entrepreneurs and deep-tech startups** (founded on scientific discovery or engineering).¹⁶ It has a **strong industry**, producing more than a quarter of the world's industrial and professional service robots (e.g. for precision farming, security, health, logistics)¹⁷, and is leading in manufacturing, healthcare, transport and space technologies – all of which increasingly rely on AI. Europe also plays an important role in the development and exploitation of platforms providing **services to companies and organisations (business-to-business)**, applications to progress towards the "intelligent enterprise" and e-government.

One of the main challenges for the EU to be competitive is to **ensure the take-up of AI technology across its economy**. European industry cannot miss the train. Only a fraction of European companies have already adopted digital technologies. This trend is particularly acute in small and medium-sized businesses. In 2017, 25% of EU large enterprises and 10% of small and medium-sized enterprises used big data analytics. Only one in five small and medium-sized enterprises was highly digitised, while one third of the workforce still does not possess basic digital skills.¹⁸ At the same time, the benefits of adopting AI are widely recognised. For example, the 2018 Digital Transformation Scoreboard shows that businesses

¹² See also the Commission's European Political Strategy Centre's Strategic Note: The Age of Artificial Intelligence, 2018.

¹³ Recent announcements include a EUR 1.7 billion AI technology park in Beijing.

¹⁴ With 1.4 billion mobile phone subscriptions and 800 million internet users – more than the USA and the EU combined – Chinese people generate vast amounts of personal data that are used to develop related AI products.

¹⁵ 10 imperatives for Europe in the age of AI and automation, McKinsey, 2017.

¹⁶ Europe accounts for the largest share of top 100 AI research institutions worldwide. 32 research institutions in the global top 100 for AI-related research paper citations vs 30 from the USA and 15 from China. Source: Atomico, State of European Tech, 2017. It can also be noted that the German Research Centre for Artificial Intelligence (DFKI) founded in 1988 is one of the world's largest research centres in the field of AI.

¹⁷ World Robotics 2017, International Federation of Robotics. Europe is home to three of the world's largest producers of industrial robots (KUKA, ABB and Comau).

¹⁸ <https://ec.europa.eu/digital-single-market/digital-scoreboard>. According to McKinsey (2016), European companies operating at the digital frontier only reach a digitisation level of 60% compared to their US peers.

in the agrifood and construction sectors which have adopted AI confirm positive impacts on entering new markets, improving products or services, and gaining new clients.¹⁹

EU efforts so far: laying the groundwork to make the most of AI

AI has featured in the EU research and development framework programmes since 2004 with a specific focus on robotics. Investments increased to up to EUR 700 million for 2014-2020, complemented by EUR 2.1 billion of private investments as part of a public-private partnership on robotics.²⁰ These efforts have significantly contributed to **Europe's leadership in robotics**.

Overall, around EUR 1.1 billion has been invested in AI-related research and innovation during the period 2014-2017 under the Horizon 2020 research and innovation programme, including in big data, health, rehabilitation, transport and space-oriented research.

Additionally, the Commission has launched major initiatives which are key for AI. These include the development of more efficient electronic components and systems, such as **chips specifically built to run AI operations** (neuromorphic chips)²¹; **world-class high-performance computers**²², as well as flagship projects on **quantum technologies** and on the mapping of the **human brain**.²³

Projects funded by the EU have developed for example:

- an unmanned agricultural vehicle that can mechanically remove weeds, reducing the need for pesticides;
- a highway pilot project using AI & Internet of Things to provide safe driving recommendations and reduce road fatalities;
- a robotic ortho-prosthesis to restore mobility to amputees;
- robots to take care of repetitive tasks for workers in car manufacturing plants and improve the efficiency of the manufacturing process.

3. THE WAY FORWARD: AN EU INITIATIVE ON AI

3.1. Boosting the EU's technological and industrial capacity and AI uptake across the economy

The public and private sectors must seize the opportunities that come both from developing innovative AI solutions and applying them to a range of fields.²⁴

The EU should be **ahead of technological developments in AI** and ensure they are swiftly taken up across its economy. This implies **stepping up investments** to strengthen fundamental research and make scientific breakthroughs, upgrade AI research infrastructure,

¹⁹ <https://ec.europa.eu/growth/tools-databases/dem/monitor/scoreboard>

²⁰ <https://eu-robotics.net/sparc/>.

²¹ Neuromorphic chips are modelled on biological structures such as brains. This project is part of the Electronic Components and Systems for European Leadership joint undertaking (EUR 4.8 billion of public-private investments by 2020).

²² <https://ec.europa.eu/digital-single-market/en/eurohpc-joint-undertaking>. This infrastructure will underpin the European Open Science Cloud that will offer researchers a virtual environment to store, process, share and re-use their data across disciplines and borders: <https://ec.europa.eu/research/openscience/>

²³ <https://ec.europa.eu/digital-single-market/en/fet-flagships>

²⁴ The recent report of the "High Level Group on Industrial Technologies" recognised AI as a "key enabling technology" highlighting the transformative role of AI and the necessity for the industry to use AI to maintain its leadership: http://ec.europa.eu/research/industrial_technologies/pdf/re_finding_industry_022018.pdf

develop AI applications in key sectors from health to transport, facilitate the uptake of AI and the access to data.

Joint effort by both the public (national and EU levels) **and private sectors** are needed to gradually increase overall investments by 2020 and beyond, in line with the EU's economic weight and investments on other continents.

Public and private **research and development investments in AI** in the EU last year were estimated to total EUR 4-5 billion.²⁵ **The EU as a whole (public and private sectors combined)** should aim to increase this investment to **at least EUR 20 billion by the end of 2020**. It should then **aim for more than EUR 20 billion per year over the following decade** (this objective does not pre-empt any decision to be taken with respect to the next EU multiannual financial framework).

The Commission will work with Member States on a coordinated plan to help align and step up investments, building on the declaration of cooperation signed on 10 April 2018.

Without such efforts, the EU risks losing out on the opportunities offered by AI, facing a brain-drain and being a consumer of solutions developed elsewhere. The EU should therefore strengthen its status as a research powerhouse while bringing more innovation to the market. A vast majority of European companies – whether large or small – should also adopt AI technologies.

Stepping up investments

2018-2020

To support joint efforts, **the Commission is increasing investments in AI** under the research and innovation framework programme Horizon 2020 to around **EUR 1.5 billion by the end of 2020** (this works out as an average of EUR 500 million per year and represents an increase of around 70%). Under the existing public-private partnerships (for example in robotics and big data), this investment will trigger an additional **EUR 2.5 billion** over the same period.

These investments will aim at consolidating research and innovation in AI, encouraging testing and experimentation, strengthening AI excellence research centres and starting efforts to bring AI to all potential users, with a focus on small and medium-sized enterprises.

If Member States²⁶ and the private sector (beyond established partnerships) **make similar investment efforts, the total investments in the EU** will grow to around EUR 7 billion per year, totalling **more than EUR 20 billion by the end of 2020**. This will position the EU well to further increase efforts over the next decade.

²⁵ Estimate based on data on public and business spending in research and development (R&D) in information and communication technologies (ICT) (source: Prospective Insights in ICT R&D, PREDICT, European Commission) and the share of funding in AI as part of the Commission's research and development budget in information and communication technologies since 2014 (around 13%). Building on previous trends, a similar share is calculated for government budget allocations for research and development and business expenditure on research and development which represents the major part of investments (ca. EUR 4 billion, which is consistent with recent findings by McKinsey).

²⁶ For example, France has just announced a EUR 1.5 billion investment in AI over five years.

Strengthening research and innovation from the lab to the market

The Commission will support **AI technologies both in basic and industrial research**²⁷. This includes investments in projects in key application areas such as health, connected and automated driving, agriculture, manufacturing, energy, next generation internet technologies, security and public administrations (including justice). Funding will also reinforce European strengths in embodied AI/robotics.

The Commission will also **support breakthrough market-creating innovation such as AI** through the pilot of the **European Innovation Council**.²⁸ A budget of EUR 2.7 billion is made available for 2018-2020 to support 1,000 potential breakthrough projects and 3,000 feasibility awards. This pilot scheme can be particularly helpful for AI development as AI technology is expected to be part of many projects, for applications in e.g. health, agriculture and manufacturing.

Funding in fundamental research is expected to be provided by the **European Research Council**, based on scientific excellence. **Marie Skłodowska-Curie actions** provide grants for all stages of researchers' careers and have supported research in AI in the past years.

Supporting AI research excellence centres across Europe

Building on Member States' efforts to **jointly establish AI-focused research centres**, the Commission will support and strengthen AI excellence centres across Europe. The Commission will also encourage and facilitate their collaboration and networking.

Bringing AI to all small businesses and potential users

Europe can only reap the full benefits of AI if it is available and accessible to all. The Commission will **facilitate access of all potential users, especially small and medium-sized enterprises, companies from non-tech sectors and public administrations, to the latest technologies and encourage them to test AI**. To this end, the Commission will support the development of an "AI-on-demand platform". This will provide a single access point for all users to relevant AI resources in the EU, including knowledge, data repositories, computing power (cloud, high performance computing), tools and algorithms. It will offer services and provide support to potential users of the technology, analyse the business case behind AI in their specific circumstances and help them to integrate AI solutions in their processes, products and services.

To facilitate access to the platform, the existing network of more than 400 Digital Innovation Hubs²⁹

***Digital Innovation Hubs** help companies (especially small and medium-sized enterprises) to take advantage of digital opportunities. They offer expertise on technologies, testing, skills, business models, finance, market intelligence and networking.*

For example, a small company that produces metal parts for the automotive industry could consult the regional hub (which can be a science park for example) and ask for advice on how to improve the manufacturing process with AI. Experts from the hub would then visit the factory, analyse the production process, consult with other AI experts in the network of hubs, make a proposal and then implement it. These activities would be partially financed with EU money.

²⁷ The guiding principle of all support for AI-related research will be the development of "responsible AI", putting the human at the centre, see the Commission's "Responsible Research and Innovation" workstream: <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation>

²⁸ <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/european-innovation-council-eic-pilot>

²⁹ See also the Commission's Communication of 19 April 2016 on Digitising European Industry (COM/2016/0180 final) and [list of hubs](#).

will be instrumental. Further Hubs are coming on stream and a dedicated network of **Digital Innovation Hubs focused on AI** will be created.

The Commission will also analyse systemic shifts in value chains in order to anticipate AI opportunities for small and medium-sized enterprises, pilot critical industrial AI applications in non-tech sectors, and reinforce the European advanced manufacturing support centre for small and medium-sized enterprises.

Supporting testing and experimentation

Testing of and experimenting with AI products and services is crucial to make them market-ready, ensure compliance with safety standards and rules as well as security by design and enable policymakers to gain experience with new technologies to devise suitable legal frameworks. The Commission will support the set-up of testing and experimentation infrastructures that are open to businesses of all sizes and from all regions. Building on the established network of Digital Innovation Hubs, a **first series of testing and experimentation infrastructures for AI** products and services will be set up in the areas of healthcare, transport, infrastructure inspection and maintenance, agrifood and agile production.

Attracting private investments

On top of investments under the research and innovation framework programme, a sufficient level of private investments in the AI transformation is crucial. The **European Fund for Strategic Investments** will be further mobilised to attract private investment to support the development and the uptake of AI, as part of the wider efforts to promote digitisation. The Commission will work with the European Investment Bank Group with the aim of reaching **at least EUR 500 million in total investments** in that area in the period 2018-2020. In addition, the European Commission and the European Investment Fund have just launched a EUR 2.1 billion Pan-European Venture Capital Funds-of-Funds programme – **VentureEU** – to boost investment in innovative startup and scale-up companies across Europe. The Commission also provides support as part of its **initiatives to digitise industry**³⁰.

In 2018-20³¹, the Commission will invest around **EUR 1.5 billion** in:

- **research and innovation in AI technologies** to strengthen European industrial leadership, excellence in science, and support AI applications which address societal challenges in sectors such as health, transport and agrifood. The Commission will also support breakthrough, market-creating innovation through the pilot phase of the European Innovation Council;
- strengthening **AI research excellence centres**; and
- the **uptake of AI across Europe**, via a **toolbox for potential users, with a focus on small and medium-sized enterprises, non-tech companies and public administrations**:

³⁰ The Commission has just launched the Strategic Forum for Important Projects of Common European Interest to identify and ensure appropriate large-scale finance for value chains of strategic importance for Europe including the integration of AI to strengthen EU industrial leadership. Moreover, the Commission supports and facilitates inter-regional partnerships for investments in advanced technologies and AI through its Smart Specialisation Platform on Industrial Modernisation.

³¹ Actions will stem from the work programme Horizon 2020. They will be financed within the current financial programming envelope and subject to future revision of the work programme in the frame of the comitology procedure.

this will include an **AI-on-demand platform** giving support and easy access to the latest algorithms and expertise; a **network of AI-focused Digital Innovation Hubs** facilitating **testing and experimentation**; and the set-up of **industrial data platforms** offering high quality datasets.

In addition, the Commission aims to stimulate more private investments in AI under the **European Fund for Strategic Investments (at least EUR 500 million in 2018-20)**.

Beyond 2020

Commission proposals under the **next EU multiannual financial framework (2021-2027)** will open the door to investments into:

- **upgrading** the pan-European **network of AI excellence centres**;
- **research and innovation in fields such as explainable AI³², unsupervised machine learning, energy and data efficiency³³**;
- additional Digital Innovation Hubs, world-leading **testing and experimentation facilities** in areas such as transport, healthcare, agrifood and manufacturing, supported by **regulatory sandboxes³⁴**;
- supporting the adoption of AI by organisations across all sectors, including **public interest applications**, through co-investment with Member States;
- exploring joint **innovation procurement** for the use and development of AI; and
- a **support centre for data sharing**, which will be closely linked with **the AI-on-demand platform** to facilitate development of business and public sector applications.

The Commission also intends to continue its support for technologies and infrastructure that underpin and enable AI such as high-performance computing, microelectronics, photonics, quantum technologies, the Internet of Things and cloud.

In doing so, the Commission will support more **energy-efficient technologies** and infrastructure, **making the AI value chain greener**.

³² In order to increase transparency and minimise the risk of bias or error, AI systems should be developed in a manner which allows humans to understand (the basis of) their actions.

³³ These being methods to use less data in order to train AIs.

³⁴ These are testing grounds for new business models that are not (yet) regulated.

Making more data available

AI needs vast amounts of data to be developed. Machine learning, a type of AI, works by identifying patterns in available data and then applying the knowledge to new data.³⁵ The larger a data set, the better even subtle relations in the data can be discovered. When it comes to using AI, data-rich environments also provide for more opportunities. This is because data is the way the algorithm learns about and interacts with its environment. For example, if all machines and processes in a factory continuously produce data, it is likely that further automation and optimisation can be achieved with the help of AI. In an analog setting, e.g. in a paper-based operation without digitised data about what is happening, that is not the case.

In view of this, access to data is a key ingredient for a competitive AI landscape, which the EU should facilitate.

The EU has made significant efforts over the past 15 years **to open up public sector information and publicly funded research results** for re-use, such as data generated by the EU's space programmes (Copernicus³⁶, Galileo). With its initiative to improve the accessibility and re-usability of such data, this body of data will grow further.

Public policy should also encourage **the wider availability of privately-held data**, while ensuring full respect for legislation on the protection of personal data. The Commission calls on companies to recognise the importance of non-personal data re-use, including for AI training purposes.

A new **support centre for data sharing** will provide public authorities and companies with legal and technical support when trying to access data from public sector bodies and companies.

The Commission will continue to study how more data can be made available.

Deep learning has been a game-changer for AI with a tremendous improvement in performance for specific tasks such as image or speech recognition, or machine translation.

Training a deep learning algorithm to classify objects works by exposing it to a large number of labelled examples (e.g. pictures) that are correctly categorised (e.g. pictures of planes).

Once trained, algorithms can correctly classify objects that they have never seen, in some cases with accuracies that exceed those of humans.

Significant advances in these technologies have been made through the use of large data sets and unprecedented computing power.

Alongside this Communication, the Commission has put forward a set of initiatives to grow the European data space³⁷. These are:

- an **updated Directive on public sector information**, e.g. traffic, meteorological, economic and financial data or business registers;
- **guidance on sharing private sector data in the economy** (including industrial data);
- an **updated Recommendation on access to and preservation of scientific information**; and

³⁵ Sometimes finding the pattern is itself the goal of the activity: in text and data mining, researchers use algorithms to “read” a large numbers of texts (e.g. scientific papers on chemistry) and automatically extract knowledge (e.g. finding facts that are not explicitly stated in any one of the papers but can be derived from the whole corpus). The Commission introduced an exception for text and data mining as part of the modernisation of EU copyright rules.

³⁶ Copernicus Data and Information Access Services: <http://copernicus.eu/news/upcoming-copernicus-data-and-information-access-services-dias>

³⁷ <https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy>

- a **Communication on the digital transformation of health and care**, including sharing of genomic and other health data sets.

3.2. Preparing for socioeconomic changes

Throughout history, the emergence of new technologies – from electricity to the internet – has changed the nature of work. It has brought major benefits to our society and economy, but also raised concerns. The emergence of automation, robotics and AI is transforming the labour market, and it is essential for the EU to manage this shift.

These technologies can make the life of workers easier. They can, for example, help them with repetitive, strenuous and even dangerous tasks (for example cleaning unsafe or difficult to access locations such as industrial pipes). They can also help summarise large amounts of data, provide more accurate information and suggest decisions, including using AI to assist doctors with diagnosis. They ultimately help to **enhance people's abilities**. Against the background of an ageing society, AI can provide new solutions to support more people to participate and remain in the labour market, including persons with disabilities. **New jobs and tasks will emerge as a result of AI**, some of which are difficult, or even impossible to predict. Other jobs and tasks will be replaced. While the exact quantification of AI's impact on jobs is difficult to determine at this stage, the need for action is clear.

Overall there are three main challenges for the EU – highlighting the fundamental role of education and training, including of teachers and trainers themselves, for which responsibility lies with Member States. The first challenge is to **prepare the society as a whole**. This means helping all Europeans to develop basic digital skills, as well as skills which are complementary to and cannot be replaced by any machine such as critical thinking, creativity or management. Secondly, the EU needs to focus efforts to help workers in **jobs which are likely to be the most transformed or to disappear** due to automation, robotics and AI. This is also about ensuring access for all citizens, including workers and the self-employed³⁸, to social protection³⁹, in line with the **European Pillar of Social Rights**. Finally, the EU needs to **train more specialists in AI**, building on its long tradition of academic excellence, create the right environment for them to work in the EU and attract more talent from abroad.

Leaving no one behind

In 2016, the European Commission launched a comprehensive plan to help equip people with the right skills for the evolving labour market: a **New Skills Agenda for Europe**⁴⁰. As part of this agenda, the Commission issued a Recommendation for Member States on "Upskilling Pathways: New Opportunities for Adults" to improve their basic literacy, numeracy and digital skills. A Recommendation was also adopted on key competences for lifelong learning, focusing notably on the acquisition of competences in sciences, technology, engineering and mathematics (STEM), digital competences, entrepreneurship and creativity. The Commission also presented a Digital Education Action Plan⁴¹ which aims to foster digital skills and

³⁸ <http://ec.europa.eu/social/BlobServlet?docId=19158&langId=en>

³⁹ Automation may impact the way social protection is financed, necessitating a proper reflection on the sustainability and adequacy of social security systems.

⁴⁰ <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016DC0381>

⁴¹ <https://ec.europa.eu/education/sites/education/files/digital-education-action-plan.pdf>

competences for all citizens. This plan explores the impact of AI in education and training through pilots.

While digitisation is affecting the structure of the labour market in particular through the automation of middle-skilled jobs, AI could have a more significant impact on lower skilled jobs.⁴² If not addressed early and proactively, this may exacerbate inequalities between people, regions and industries in the EU.

To manage the AI transformation, workers whose jobs are changing or may disappear due to automation must have every opportunity to acquire the skills and knowledge they need, to master new technology and be supported during labour market transitions. This anticipatory approach and focus on investing in people is a cornerstone of a human-centric, inclusive approach to AI, and will require a significant investment. National schemes will be essential for providing such up-skilling and training. They will benefit from support by the European Structural and Investment Funds (supporting skills development with EUR 27 billion over the period 2014-2020, out of which the European Social Fund invests 2.3 billion specifically in digital skills) and should also benefit from support from the private sector. The Commission will also continue to support research into human-AI interaction and cooperation.

Nurturing talent, diversity and interdisciplinarity

AI has brought about new job profiles, including in the area of developing machine-learning algorithms and other digital innovations.⁴³ Overall, the number of specialists in information and communication technologies in the EU has grown annually by 5% since 2011, creating 1.8 million jobs and rapidly increasing its share of total employment from 3% to 3.7% in just five years. There are at least 350,000 vacancies for such professionals in Europe, pointing to significant skills gaps.⁴⁴ This is why Europe should strive to **increase the number of people trained in AI** and **encourage diversity**. More women and people of diverse backgrounds, including people with disabilities, need to be involved in the development of AI, starting from inclusive AI education and training, in order to ensure that AI is non-discriminatory and inclusive. **Interdisciplinarity** should also be supported (by encouraging joint degrees, for example in law or psychology and AI). The importance of ethics in the development and use of new technologies should also be featured in programmes and courses. And it is not only about training the best talent, but also creating an **attractive environment for them to stay in the EU**.

Initiatives to encourage more young people to choose AI subjects and related fields as a career should be promoted. The Commission has recently launched the "Digital Opportunity Traineeships"⁴⁵, supporting internships aimed at acquiring advanced digital skills, and a number of actions of the Digital Skills and Jobs Coalition⁴⁶ aim at spreading coding skills and increasing the number of experts in digital.

Ensuring that workers are given the chance to adapt and to have access to new opportunities will be crucial for people to accept AI. Like any other technology, AI is not just imposed on society. It is up to governments, in dialogue with the social partners and civil society bodies,

⁴² Organisation for Economic Co-operation and Development, "Automation, skills use and training", 2018.

⁴³ <https://www.cognizant.com/whitepapers/21-jobs-of-the-future-a-guide-to-getting-and-staying-employed-over-the-next-10-years-codex3049.pdf>

⁴⁴ http://www.pocbigdata.eu/monitorICTonlinevacancies/general_info/

⁴⁵ <https://ec.europa.eu/digital-single-market/en/digital-opportunity-traineeships-boosting-digital-skills-job>

⁴⁶ <https://ec.europa.eu/digital-single-market/en/digital-skills-jobs-coalition>

to collectively steer the process to ensure that its benefits are widely shared, that all citizens are suitably equipped to take full advantage of this technology and that a broader reflection on potentially deeper societal changes is taking place.

In 2018, in order to support the efforts of Member States which are responsible for labour and education policies, the Commission will:

- set up **dedicated (re-)training schemes** in connection with the Blueprint on sectoral cooperation on skills⁴⁷ – which brings together businesses, trade unions, higher education institutions and public authorities – for professional profiles which are at risk of being automated, with financial support from the European Social Fund⁴⁸;
- gather detailed analysis and expert inputs to **anticipate the changes on the labour market and the skills mismatch** across the EU, and inform decision-making at EU, national and local levels. More specifically, the Commission will (i) publish a foresight report on the impact of AI in education; (ii) launch pilots to predict the training requirements for future competence profiles; and (iii) publish an expert **report addressing the labour market impacts of AI**, with **recommendations**;
- support Digital Opportunity **Traineeships** (2018-20) in **advanced digital skills** for students and fresh graduates;
- encourage, through the Digital Skills and Jobs Coalition, **business-education partnerships** to take steps to attract and retain more AI talent and to foster continued collaboration; and
- invite **social partners** to include AI and its impact on the economy and employment, including the importance of diversity and gender balance in AI jobs, in their joint work programmes at sectoral and cross-sectoral level where relevant.

The **European Institute of Innovation and Technology** will integrate **AI across curricula in the education courses it supports**, in order to contribute to developing a talent pool for AI in Europe.

Proposals under the next EU multiannual financial framework (2021-2027) will include strengthened support for the acquisition of advanced digital skills including AI-specific expertise.

The Commission also intends to broaden the scope of the current European Globalisation Adjustment Fund beyond redundancies caused by delocalisation, including to those resulting from digitisation and automation.

3.3. Ensuring an appropriate ethical and legal framework

An environment of trust and accountability around the development and use of AI is needed.

The **values** set out in Article 2 of the Treaty on European Union constitute the foundation of the rights enjoyed by those living in the Union. In addition, the **EU Charter of Fundamental**

⁴⁷ <http://ec.europa.eu/social/main.jsp?catId=1415&langId=en>

⁴⁸ The cooperation now focuses on the automotive, maritime technology, space, textile and tourism sectors, and will address six other sectors in the future: additive manufacturing; construction; green technologies and renewable energy; maritime shipping; paper-based value chain; steel industry.

Rights brings together all the personal, civic, political, economic and social rights enjoyed by people within the EU in a single text.

The EU has a strong and balanced regulatory framework to build on, which can set the global standard for a sustainable approach to this technology. The Union has **high standards in terms of safety and product liability**. The first EU-wide rules on **network and information systems security** and stronger rules on the **protection of personal data** will become a reality in May 2018.

The **General Data Protection Regulation** ensures a high standard of personal data protection, including the principles of data protection by design and by default. It guarantees the free flow of personal data within the Union. It contains provisions on decision-making based solely on automated processing, including profiling. In such cases, data subjects have the **right to be provided with meaningful information** about the logic involved in the decision.⁴⁹ The General Data Protection Regulation also gives individuals the right not to be subject solely to automated decision-making, except in certain situations.⁵⁰ The Commission will closely follow the Regulation's application in the context of AI and calls on the national data protection authorities and the European Data Protection Board to do the same.

The Commission has also put forward a series of proposals under the **Digital Single Market** strategy that will be a key enabler for the development of AI, such as the Regulation on the free flow of non-personal data, and that will strengthen trust in the online world, such as the ePrivacy Regulation and the Cybersecurity Act. These proposals need to be adopted as soon as possible. This is essential as **citizens and businesses alike need to be able to trust the technology they interact with**, have a predictable legal environment and rely on effective safeguards protecting fundamental rights and freedoms.

To further strengthen trust, people also need to understand how the technology works, hence the importance of research into the **explainability of AI systems**. Indeed, in order to increase transparency and minimise the risk of bias or error, AI systems should be developed in a manner which allows humans to understand (the basis of) their actions.

Like every technology or tool, AI can be used to positive but also to malicious ends. Whilst AI clearly generates new opportunities, it also poses challenges and risks, for example in the areas of safety and liability, security (criminal use or attacks), bias⁵¹ and discrimination.

Reflection will be needed on interactions between AI and intellectual property rights, from the perspective of both intellectual property offices and users, with a view to fostering innovation and legal certainty in a balanced way.⁵²

Draft AI ethics guidelines

As a first step to address ethical concerns, **draft AI ethics guidelines will be developed by the end of the year**, with due regard to the Charter of Fundamental Rights of the European Union. The Commission will bring together all relevant stakeholders in order to help develop these draft guidelines.

⁴⁹ Articles 13 (2) f), 14 (2) g) and 15 (1) h) of the General Data Protection Regulation.

⁵⁰ Article 22 of the General Data Protection Regulation.

⁵¹ Depending on the data input that is used to train AI systems, their outputs can be biased.

⁵² Using AI to create works can have implications on intellectual property, with questions arising for instance on patentability, copyright and right ownership.

The draft guidelines will address issues such as the future of work, fairness, safety, security, social inclusion and algorithmic transparency. More broadly, they will look at the impact on fundamental rights, including privacy, dignity, consumer protection and non-discrimination. They will build on the work of the European Group on Ethics in Science and New Technologies⁵³ and take inspiration from other similar efforts.⁵⁴ Companies, academic institutions, and other organisations from civil society bodies will be invited to contribute. In parallel, the Commission will continue its work towards progress on ethics at international level⁵⁵.

While self-regulation can provide a first set of benchmarks against which emerging applications and outcomes can be assessed, public authorities must ensure that the regulatory frameworks for developing and using of AI technologies are in line with these values and fundamental rights. The Commission will monitor developments and, if necessary, review existing legal frameworks to better adapt them to specific challenges, in particular to ensure the respect of the Union's basic values and fundamental rights.

Safety and liability

The emergence of AI, in particular the complex enabling ecosystem and the feature of autonomous decision-making, requires a reflection about the suitability of some established rules on safety and civil law questions on liability.

For instance, advanced robots and Internet of Things products empowered by AI may act in ways that were not envisaged at the time when the system was first put into operation. Given AI's widespread uses, both horizontal and sectoral rules may need to be reviewed⁵⁶.

The EU safety framework⁵⁷ already addresses the intended use and foreseeable (mis)use of products when placed on the market. This has led to the development of a solid body of standards in the area of AI-enabled devices that is continuously being adapted in line with technological progress..

The further development and promotion of such safety standards and support in EU and international standardisation organisations will help enable European businesses to benefit from a competitive advantage, and increase consumer trust⁵⁸.

⁵³ The European Group on Ethics in Science and New Technologies is an advisory group of the Commission.

⁵⁴ At the EU level, the EU Fundamental Rights Agency will carry out an assessment of the current challenges faced by producers and users of new technology with respect of fundamental rights compliance. The European Group on Ethics in Science and New Technologies also published a relevant statement on AI, Robotics and 'Autonomous' Systems on 9 March 2018. Examples of international efforts: Asilomar AI principles (<https://futureoflife.org/ai-principles/>), Montréal Declaration for Responsible AI draft principles (<https://www.montrealdeclaration-responsibleai.com/>), UNI Global Union Top 10 Principles for Ethical AI (<http://www.thefutureworldofwork.org/opinions/10-principles-for-ethical-ai/>).

⁵⁵ The European Commission's International Dialogue on Bioethics and Ethics in Science and New Technologies brings together the National Ethics Councils of EU Member States and of third countries, to work together on those matters of common concern.

⁵⁶ For any new regulatory proposals that shall be needed to address emerging issues resulting from AI and related technologies, the Commission applies the Innovation Principle, a set of tools and guidelines that was developed to ensure that all Commission initiatives are innovation friendly: https://ec.europa.eu/epsc/publications/strategic-notes/towards-innovation-principle-endorsed-better-regulation_en

⁵⁷ For example, the Machinery Directive, the Radio Equipment Directive, the General Product Safety Directive as well as specific safety rules for example for medical devices or toys.

⁵⁸ Standards should also cover interoperability, which is crucial for offering consumers greater choices and ensuring fair competition.

The Commission is currently assessing whether the safety and national and EU liability frameworks are fit for purpose in light of these new challenges or whether any gaps should be addressed. A high level of safety and an efficient redress mechanism for victims in case of damages helps to build user trust and social acceptance of these technologies.

Evaluations of the Product Liability Directive⁵⁹ and the Machinery Directive have already been conducted.⁶⁰ An initial assessment was also carried out on the current liability frameworks in light of AI and emerging technologies.⁶¹ An expert group will help the Commission to analyse these challenges further.⁶²

Empowering individuals and consumers to make the most of AI

The large-scale use of AI-enabled tools in business-to-consumer transactions needs to be fair, transparent and compliant with consumer legislation. Consumers should receive clear information on the use, features and properties of AI-enabled products. Individuals should be able to control the data generated by using these tools and should know whether they are communicating with a machine or another human. In particular, when interacting with an automated system, consideration should be given to when users should be informed on how to reach a human and how to ensure that a system's decisions can be checked or corrected.

The Commission will:

- set a framework for stakeholders and experts – the European AI Alliance – to develop **draft AI ethics guidelines**, with due regard to fundamental rights, **by the end of the year**, in cooperation with the European Group on Ethics in Science and New Technologies;
- **issue a guidance document on the interpretation of the Product Liability Directive** in light of technological developments **by mid-2019**. This will seek to ensure legal clarity for consumers and producers in case of defective products;
- publish, **by mid-2019**, a **report on the broader implications** for, potential **gaps in and orientations for**, the **liability and safety frameworks** for AI, Internet of Things and robotics;
- support research in the development of **explainable AI** and implement a pilot project proposed by the European Parliament on **Algorithmic Awareness Building**⁶³, to gather a solid evidence-base and support the design of policy responses to the challenges brought by automated decision-making, including biases and discrimination (2018-2019); and
- support national and EU-level **consumer organisations and data protection supervising authorities** in building an understanding of AI-powered applications with the input of the European Consumer Consultative Group and of the European Data Protection Board.

⁵⁹ The Product Liability Directive states that if a defective product causes any damage to consumers or their property, the producer has to provide compensation irrespectively of whether there is negligence or fault on their part.

⁶⁰ The evaluation of the Machinery Directive indicates that some provisions do not explicitly address certain aspects of emerging digital technologies, and the Commission will examine whether this requires legislative changes. On the evaluation of the Product Liability Directive, the Commission will issue an interpretative guidance document, clarifying important concepts in the Directive.

⁶¹ See the Staff Working Document on Liability accompanying this Communication (SWD (2018)137).

⁶² http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=615947

⁶³ <https://ec.europa.eu/digital-single-market/en/algorithmic-awareness-building>

3.4. Joining forces

Engaging Member States

Several Member States have developed or are working towards strategies to support AI. On 29 March 2018, France presented its national strategy for AI, building on the Villani report.⁶⁴ Germany, following the example of "Industrie 4.0", has set up a platform on learning systems to enable a strategic dialogue between academia, industry and the government, and it has put forward a report on the ethics of automated and connected driving.⁶⁵ Finland has put forward its 'Tekoälyaika' strategy to make it a leader in the field.⁶⁶ Every Member State is encouraged to have an AI strategy, including on investment.

Sharing best practices, identifying synergies and aligning action where relevant will maximise the impact of investments in AI and help the EU as a whole to compete globally. Cooperating on interoperability and data sets, and working together on legal solutions will prevent a fragmentation of the single market and therefore fuel the emergence of AI startups. 24 Member States and Norway have already committed to joining forces on AI and entering into a strategic dialogue with the Commission.⁶⁷ **The Commission will facilitate this dialogue and aim to agree a coordinated plan on AI with Member States by the end of the year.**

Engaging stakeholders: setting up a European AI Alliance

Given the scale of the challenge associated with AI, the full mobilisation of a diverse set of participants, including businesses, consumer organisations, trade unions, and other representatives of civil society bodies is essential. The Commission will therefore facilitate the creation and operation of **a broad multi-stakeholder platform, the European AI Alliance**, to work on all aspects of AI.⁶⁸ The Commission will also facilitate interactions of the Alliance with the European Parliament, Member States, the European Economic and Social Committee, the Committee of the Regions as well as international organisations. The Alliance will be a space for sharing best practices, encourage private investments and activities related to the development of AI.

⁶⁴ <https://www.aiforhumanity.fr>

⁶⁵ <https://www.plattform-lernende-systeme.de>

⁶⁶ <https://tekoalyaika.fi/>

⁶⁷ <https://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence>

⁶⁸ <https://ec.europa.eu/digital-single-market/en/news/call-high-level-expert-group-artificial-intelligence>

Monitoring AI development and uptake

Many of today's debates about AI are based on opinions, hearsay and assumptions – not always on facts and science. To ensure quality input and to inform policy-making, the Commission will monitor the uptake of AI applications across the economy and identify potential shifts in industrial value chains caused by AI as well as societal and legal developments and the situation on the labour market. It will also benchmark the technical capabilities of AI components and systems to give a realistic understanding of where the technology stands, and help increase public awareness.⁶⁹ The Commission will also regularly assess progress towards the objectives and the initiatives set out in this Communication.

International outreach

International discussions on AI have intensified after Japan's G7 Presidency put the topic on the table in 2016. The EU has supported these discussions both in G7 ministerial meetings and at the Organisation for Economic Co-operation and Development, which is becoming a major international venue to discuss this topic. More specifically, the Commission has encouraged the discussions on AI ethics within the G7.

With AI being easily tradeable across borders, only global solutions will be sustainable in this domain. The G7/G20, United Nations and Organisation for Economic Co-operation and Development have begun to address the role of AI, including in the military domain. The EU will continue to encourage discussions on AI and its various dimensions – including research and innovation cooperation as well as competitiveness – in such fora. It will promote the use of AI, and technologies in general, to help solve global challenges, support the implementation of the Paris Climate agreement and achieve the United Nations Sustainable Development Goals.

The EU can make a unique contribution to the worldwide debate on AI based on its values and fundamental rights.

- **By the end of the year**, the Commission will work, as part of the existing European platform of national initiatives to digitise industry, on a **coordinated plan with Member States** in order to maximise the impact of investments at EU and national levels, exchange on the best way for governments to prepare Europeans for the AI transformation and address legal and ethical considerations. In parallel, the Commission will **systematically monitor AI-related developments**, e.g. policy initiatives in the Member States, AI uptake and its impact on labour markets as well as AI capabilities, including high-level benchmarking, showcasing current capabilities and developing an AI index in order to inform the discussions.
- **By July 2018**, the **European AI Alliance** will be set up. It will involve all relevant stakeholders to gather input, exchange views, develop and implement common measures to encourage the development and use of AI.

4. CONCLUSION

The EU has a strong scientific and industrial base to build on, with leading research labs and universities, recognised leadership in robotics as well as innovative startups. It has a

⁶⁹ This work shall also be informed by the EU Agency on Fundamental Rights.

comprehensive legal framework which protects consumers while promoting innovation and it is making progress in creating a Digital Single Market. **The main ingredients are there for the EU to become a leader in the AI revolution**, in its own way and based on its values.

The approach to AI described in this document shows the way forward and highlights the need to join forces at European level, to ensure that all Europeans are part of the digital transformation, that adequate resources are devoted to AI and that the Union's values and fundamental rights are at the forefront of the AI landscape.

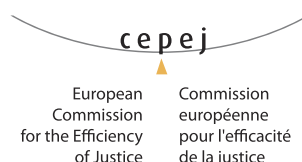
Together, we can place the **power of AI at the service of human progress**.

EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ)

European ethical Charter
on the use of Artificial Intelligence in
judicial systems and their environment



Adopted at the 31st plenary meeting
of the CEPEJ (Strasbourg, 3-4 December 2018)



EUROPEAN COMMISSION FOR THE EFFICIENCY OF JUSTICE (CEPEJ)

European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment

Adopted at the 31st plenary meeting
of the CEPEJ (Strasbourg, 3-4 December 2018)

Council of Europe

French edition:
*Commission européenne pour l'efficacité
de la justice (CEPEJ) – Charte éthique
européenne d'utilisation de l'intelligence
artificielle dans les systèmes
judiciaires et leur environnement*

*The opinions expressed in this work are
the responsibility of the authors and do
not necessarily reflect the official policy
of the Council of Europe.*

All requests concerning the
reproduction or translation of all
or part of this document should
be addressed to the Directorate of
Communication (F-67075 Strasbourg
Cedex or publishing@coe.int). All
other correspondence concerning
this document should be addressed
to European Commission for the
Efficiency of Justice (CEPEJ)
cepej@coe.int

Cover and layout:
Documents and Publications
Production Department (SPDP),
Council of Europe
Photo: Shutterstock

This publication has not been
copy-edited by the SPDP Editorial
Unit to correct typographical and
grammatical errors.

© Council of Europe, February 2019
Printed at the Council of Europe

Contents

INTRODUCTION	5
THE FIVE PRINCIPLES OF THE ETHICAL CHARTER ON THE USE OF ARTIFICIAL INTELLIGENCE IN JUDICIAL SYSTEMS AND THEIR ENVIRONMENT	7
1. Principle of respect for fundamental rights: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights	8
2. Principle of non-discrimination: Specifically prevent the development or intensification of any discrimination between individuals or groups of individuals	
3. Principle of quality and security: With regard to the processing of judicial decisions and data, use certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment	
4. Principle of transparency, impartiality and fairness: Make data processing methods accessible and understandable, authorise external audits	
5. Principle "under user control": Preclude a prescriptive approach and ensure that users are informed actors and in control of their choices	
APPENDIX I – IN-DEPTH STUDY ON THE USE OF AI IN JUDICIAL SYSTEMS, NOTABLY AI APPLICATIONS PROCESSING JUDICIAL DECISIONS AND DATA	13
1. State of the use of artificial intelligence algorithms in the judicial systems of Council of Europe member States	16
2. Overview of open data policies relating to judicial decisions in the judicial systems of Council of Europe member States	18
3. Operating characteristics of artificial intelligence (machine learning) applied to judicial decisions	29
4. Can artificial intelligence model legal reasoning in advance?	35
5. Can AIs explain judges' behaviour in retrospect?	39
6. How is AI to be applied in civil, commercial and administrative justice?	41
7. Issues specific to criminal justice: prevention of offences, risk of recidivism and assessment of the level of danger	48
8. Specific questions relating to the protection of personal data	56
9. The potential and limitations of predictive justice tools	57
10. The need for an in-depth public debate on these tools prior to the implementation of public policies for their development. The urgent need for cyberethics to provide a framework for the development of artificial intelligence algorithms while respecting fundamental rights	59

APPENDIX II – WHICH USES OF AI IN EUROPEAN JUDICIAL SYSTEMS?	63
Uses to be encouraged	64
Possible uses, requiring considerable methodological precautions	64
Uses to be considered following additional scientific studies	66
Uses to be considered with the most extreme reservations	66
APPENDIX III – GLOSSARY	69
APPENDIX IV – CHECKLIST FOR INTEGRATING THE CHARTER'S PRINCIPLES INTO YOUR PROCESSING METHOD	76
CHECKLIST FOR EVALUATING YOUR PROCESSING METHODS	77

Introduction

Acknowledging the increasing importance of artificial intelligence¹ (AI) in our modern societies, and the expected benefits when it will be fully used at the service of the efficiency and quality of justice, the CEPEJ formally adopts the 5 fundamental principles entitled “European Ethical Charter on the use of AI in the judicial systems and their environment”.

The Charter is intended for public and private stakeholders responsible for the design and deployment of artificial intelligence tools and services that involve the processing of judicial decisions and data (machine learning or any other methods deriving from data science).

It also concerns public decision-makers in charge of the legislative or regulatory framework, of the development, audit or use of such tools and services.

The use of such tools and services in judicial systems seeks to improve the efficiency and quality of justice, and should be encouraged. It must, however, be carried out with responsibility, with due regard for the fundamental rights of individuals as set forth in the European Convention on Human Rights and the Convention on the Protection of Personal Data, and in compliance with other fundamental principles set out below, which should guide the framing of public justice policies in this field.

Judicial decision processing by artificial intelligence, according to their developers, is likely, in civil, commercial and administrative matters, to help improve the predictability of the application of the law and consistency of court decisions, subject to compliance with the principles set out below. In criminal matters, their use must be considered with the greatest reservations in order to prevent discrimination based on sensitive data, in conformity with the guarantees of a fair trial.

Whether designed with the aim of assisting in the provision of legal advice, helping in drafting or in the decision-making process, or advising the user, it is essential that processing is carried out with transparency, impartiality and equity, certified by an external and independent expert assessment.

1. For the definition of artificial intelligence, see the Glossary appended.

Application of the Charter

The principles of the Charter should be subject to regular application, monitoring and evaluation by public and private actors, with a view to continuous improvement of practices.

In this respect, it is desirable that a regular review of the implementation of the principles of the Charter be made by these actors, explaining, where appropriate, the reasons for non-implementation or partial implementation, accompanied by an action plan to introduce the necessary measures.

The independent authorities mentioned in the Charter could be responsible to periodically assess the level of endorsement of the Charter's principles by all actors, and to propose improvements to adapt it to changing technologies and uses of such technologies.

The five principles of the Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment

1

Principle of respect for fundamental rights: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights.

2

Principle of non-discrimination: specifically prevent the development or intensification of any discrimination between individuals or groups of individuals.

3

Principle of quality and security: with regard to the processing of judicial decisions and data, use certified sources and intangible data with models elaborated in a multi-disciplinary manner, in a secure technological environment.

4

Principle of transparency, impartiality and fairness: make data processing methods accessible and understandable, authorise external audits.

5

Principle “under user control”: preclude a prescriptive approach and ensure that users are informed actors and in control of the choices made.

1

Principle of respect for fundamental rights: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights

■ The processing of judicial decisions and data must serve clear purposes, in full compliance with the fundamental rights guaranteed by the European Convention on Human Rights (ECHR) and the Convention on the Protection of Personal Data (Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, ETS No. 108 as amended by the CETS amending protocol No. 223).

■ When artificial intelligence tools are used to resolve a dispute or as a tool to assist in judicial decision-making or to give guidance to the public, it is essential to ensure that they do not undermine the guarantees of the right of access to the judge and the right to a fair trial (equality of arms and respect for the adversarial process).

■ They should also be used with due respect for the principles of the rule of law and judges' independence in their decision-making process.

■ Preference should therefore be given to *ethical-by-design*² or *human-rights-by-design* approaches. This means that right from the design and learning phases, rules prohibiting direct or indirect violations of the fundamental values protected by the conventions are fully integrated.

2. The ethical choice is made upstream by the program designers and is therefore not left to the user.

2

Principle of non-discrimination: specifically prevent the development or intensification of any discrimination between individuals or groups of individuals

■ Given the ability of these processing methods to reveal existing discrimination, through grouping or classifying data relating to individuals or groups of individuals, public and private stakeholders must ensure that the methods do not reproduce or aggravate such discrimination and that they do not lead to deterministic analyses or uses.

■ Particular care must be taken in both the development and deployment phases, especially when the processing is directly or indirectly based on “sensitive” data. This could include alleged racial or ethnic origin, socio-economic background, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, health-related data or data concerning sexual life or sexual orientation. When such discrimination has been identified, consideration must be given to corrective measures to limit or, if possible, neutralise these risks and as well as to awareness-raising among stakeholders.

■ However, the use of machine learning and multidisciplinary scientific analyses to combat such discrimination should be encouraged.

3

Principle of quality and security: with regard to the processing of judicial decisions and data, use certified sources and intangible data with models conceived in a multi-disciplinary manner, in a secure technological environment

■ Designers of machine learning models should be able to draw widely on the expertise of the relevant justice system professionals (judges, prosecutors, lawyers, etc.) and researchers/lecturers in the fields of law and social sciences (for example, economists, sociologists and philosophers).

■ Forming mixed project teams in short design cycles to produce functional models is one of the organisational methods making it possible to capitalise on this multidisciplinary approach.

■ Existing ethical safeguards should be constantly shared by these project teams and enhanced using feedback.

■ Data based on judicial decisions that is entered into a software which implements a machine learning algorithm should come from certified sources and should not be modified until they have actually been used by the learning mechanism. The whole process must therefore be traceable to ensure that no modification has occurred to alter the content or meaning of the decision being processed.

■ The models and algorithms created must also be able to be stored and executed in secure environments, so as to ensure system integrity and intangibility.

4

Principle of transparency, impartiality and fairness: make data processing methods accessible and understandable, authorise external audits

■ A balance must be struck³ between the intellectual property of certain processing methods and the need for transparency (access to the design process), impartiality (absence of bias)⁴, fairness and intellectual integrity (prioritising the interests of justice) when tools are used that may have legal consequences or may significantly affect people's lives. It should be made clear that these measures apply to the whole design and operating chain as the selection process and the quality and organisation of data directly influence the learning phase.

■ The first option is complete technical transparency (for example, open source code and documentation), which is sometimes restricted by the protection of trade secrets. The system could also be explained in clear and familiar language (to describe how results are produced) by communicating, for example, the nature of the services offered, the tools that have been developed, performance and the risks of error. Independent authorities or experts could be tasked with certifying and auditing processing methods or providing advice beforehand. Public authorities could grant certification, to be regularly reviewed.

3. Of interest in this connection is the suggestion made on page 38 of the Council of Europe's MSI-NET study on "Algorithms and Human Rights": "The provision of entire algorithms or the underlying software code to the public is an unlikely solution in this context, as private companies regard their algorithm as key proprietary software that is protected. However, there may be a possibility of demanding that key subsets of information about the algorithms be provided to the public, for example which variables are in use, which goals the algorithms are being optimised for, the training data and average values and standard deviations of the results produced, or the amount and type of data being processed by the algorithm." Or even the suggestions appearing on page 117 of the "AI for humanity" report drafted by Mr Cédric Villani, a member of the French National Assembly as part of a mission assigned to him by the Prime Minister of the French Republic: "The auditors may be satisfied with simply checking the fairness and equity of a programme (doing only what is required of them), by submitting a variety of false input data, for example, or by creating a large quantity of system user profiles according to precise guidelines." In addition, there are also the statements in the report by the House of Lords, "AI in the UK: ready, willing and able?", paragraphs 92, 96-99.
4. In this connection, it is interesting to note the solutions generally considered for ensuring the neutrality of algorithms in the above-mentioned report by the House of Lords (paragraphs 114, 115, 116, 119, 120): more diverse datasets, more diversity and multidisciplinary approaches, more auditing of aspects such as data processing and the manner in which the machine is constructed.

5

Principle “under user control”: preclude a prescriptive approach and ensure that users are informed actors and in control of their choices

■ User autonomy must be increased and not restricted through the use of artificial intelligence tools and services.

■ Professionals in the justice system should, at any moment, be able to review judicial decisions and the data used to produce a result and continue not to be necessarily bound by it in the light of the specific features of that particular case.

■ The user must be informed in clear and understandable language whether or not the solutions offered by the artificial intelligence tools are binding, of the different options available, and that s/he has the right to legal advice and the right to access a court. S/he must also be clearly informed of any prior processing of a case by artificial intelligence before or during a judicial process and have the right to object, so that his/her case can be heard directly by a court within the meaning of Article 6 of the ECHR.

■ Generally speaking, when any artificial intelligence-based information system is implemented there should be computer literacy programmes for users and debates involving professionals from the justice system.

Appendix I

In-depth study on the use of AI in judicial systems, notably AI applications processing judicial decisions and data

**prepared by Mr Xavier Ronsin, First President of the Court of Appeal of Rennes, scientific expert (France),
and**

Mr Vasileios Lamos, principal research fellow at the Computer Science department of University College London (UCL), scientific expert (United Kingdom),

and with the contribution of Ms Agnès Maïtrepiere, judge, member of the Consultative Committee of the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data of the Council of Europe (France)

The following experts also contributed to fintune the Study:

Mr Francesco Contini, Senior Researcher at the Research Institute on Judicial Systems – National Research Council (IRSIG-CNR), Bologna (Italy)

Mr Francesco De Santis, Professor of Human Rights Procedures, University of Naples (Italy)

Mr Jean Lassègue, philosopher and epistemologist, research fellow at the Centre National de Recherche Scientifique (CNRS) and associate researcher at the Institut des Hautes Etudes sur la Justice (IHEJ) (France)

Ms Dory Reiling, Honorary Senior Judge, Independent Expert on Information Technology and Judicial Reform (Netherlands)

Mr Aleš Završnik, Chief Researcher at the Institute of Criminology, Associate Professor at the Faculty of Law, University of Ljubljana (Slovenia) and EURIAS Researcher 2017-18 at Collegium Helveticum in Zürich (Switzerland)

INTRODUCTION

1. The wave of digital transformation in our societies still has an uneven effect on the judicial systems of Council of Europe member States. Many European countries seem to have already developed an extremely advanced approach to using practical applications (in terms of both technology and legal support), while for others, this is still just an emerging issue and the focus is solely on effective IT management.
2. Among the technologies at work in this great digital transformation, artificial intelligence (AI) appears to be both the most spectacular and the most striking. In the United States, “robot lawyers” are already at work and seem to converse in natural language with humans. Legal tech start-ups specialising in the design of new legal services offer new applications to legal professions, mainly lawyers, legal services and insurers, allowing in-depth access to judicial information and case law. These private companies even aim to predict judges’ decisions with “predictive justice” tools, although we will see that this may not be the best description for them⁵.
3. An initial examination of this phenomenon, however, prompts us to differentiate between this commercial discourse and the reality of the use and deployment of these technologies. For the time being judges in the Council of Europe member States do not seem to be making any practical and daily use of predictive software. Local tests⁶ and academic work⁷ have been carried out to explore the potential of these applications, but they have not yet been applied on a wide scale. The initiative for the development of these tools comes largely from the private sector, whose clientele so far has been made up mostly of insurance companies, lawyers and legal services wanting to reduce legal uncertainty and the unpredictability of judicial decisions. Nevertheless, public decision-makers are beginning to be increasingly solicited by a private sector wishing to see these tools – which are sometimes “beta” versions, i.e. they will evolve over time – integrated into public policies.

5. See in particular the frame opening Chapter 9.

6. For example, the Douai and Rennes Courts of Appeal in France conducted a three-month trial in 2017 with a software programme labelled “predictive” by a panel of judges.

7. Work on a sample of 584 decisions of the European Court of Human Rights: Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preoțiu-Pietro, Vasileios Lampos, “Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective”, published on 24 October 2016, [Online], <https://peerj.com/articles/cs-93/>

4. In the line of the thought process initiated in its “Guidelines on how to drive change towards Cyberjustice”,⁸ the CEPEJ proposes to provide public decision-makers and justice professionals with keys for a better understanding of the “predictive justice” phenomenon.
5. The first task will be to clarify the questions relating to the intrinsic nature of these mass case-law data processing systems, along with their technical and theoretical limitations. These aspects have not often been mentioned in the debate on this subject in the judicial sphere but they are very well known and discussed by specialists in these technologies (mathematicians, statisticians and computer scientists) and deserve some attention.
6. Secondly, this document will analyse the benefits and risks of these tools. While their supporters highlight their assets in terms of transparency, predictability and standardisation of case-law, their critics point to the limitations and the reasoning bias of the software currently on the market. The inherent risks in these technologies may even transcend the act of judging and affect essential functioning elements of the rule of law and judicial systems, to which the Council of Europe is particularly attached.
7. These include principles such as the primacy of law. The effect of these tools may be not only to provide incentives but to be almost prescriptive, creating a new form of normativity, which could supplement the law by regulating the sovereign discretion of the judge, and potentially leading, in the long term, to a standardisation of judicial decisions based no longer on case-by-case reasoning by the courts, but on a pure statistical calculation linked to the average compensation previously awarded by other courts.
8. There is also a need to consider whether these solutions are compatible with the individual rights enshrined in the European Convention on Human Rights (ECHR). These include the right to a fair trial (particularly the right to a natural judge established by law, the right to an independent and impartial tribunal and equality of arms in judicial proceedings) and, where insufficient care has been taken to protect data communicated in open data, the right to respect for private and family life.
9. While taking these issues into consideration, the document highlights the great potential of AI to help legal professionals in their work. There is no doubt that some AI applications which are still under development or testing, such as those designed to improve legal research, could be very useful in rendering processing of the judicial workload faster and more efficient. The

8. See in particular paragraph 51 of the document CEPEJ(2016)13, Guidelines on how to drive change towards Cyberjustice.

document highlights these positive examples and generally advocates the use of AI by legal professionals according to their needs, provided that due regard is shown for the individual rights guaranteed by the ECHR and Council of Europe standards, particularly in criminal matters. Far from being a simple instrument for improving the efficiency of judicial systems, AI should strengthen the guarantees of the rule of law, together with the quality of public justice.

10. Lastly, the document suggests means of monitoring this phenomenon in the form of an ethical charter, emphasising the need for a cautious approach to the integration of these tools into public policies. It is essential that any public debate involves all the stakeholders, whether legal professionals, legal tech companies or scientists, to enable them to convey the full scope and possible impact of the introduction of artificial intelligence applications in judicial systems and devise the ethical framework in which they must operate. Subsequently, this debate could go beyond a pure “business” framework, involving citizens themselves and hence contributing to some extent to general computer literacy, as has been achieved in Canada.⁹

1. State of the use of artificial intelligence algorithms in the judicial systems of Council of Europe member States

In 2018, the use of artificial intelligence algorithms in European judicial systems remains primarily a private-sector commercial initiative aimed at insurance companies, legal departments, lawyers and individuals.

11. The use of AI in the judicial field appears to be quite popular in the United States, which has invested in these tools in a fairly uncomplicated way, both in civil and criminal matters.¹⁰

12. Pinpointing instances of AI algorithm initiatives in the judicial systems of Council of Europe member States is a more difficult task, as most of the initiatives come from the private sector and are not often integrated into public policies.

13. The question of the use of AI in judicial systems was dealt with in a specific online survey, launched in April 2018 for representatives of the

9. Declaration of Montreal, available at <https://www.declarationmontreal-iaresponsable.com/demarche>, accessed 16 July 2018.

10. See COMPAS algorithms or tools such as RAVEL LAW or ROSS chatbot

CEPEJ member States and civil society. The response level was relatively low and did not allow clear trends to be identified. Some private operators did not seem very receptive to this survey and the members of the CEPEJ, who belong for the most part to ministries of justice or higher councils of justice, were able to quote only the tools currently used by the public sphere.

14. As a result, the inventory below is only partial and is based solely on research conducted by experts and the secretariat using publicly available literature.¹¹

15. Classifications can be made according to the service offered. The involvement of AI can vary greatly according to the applications. For illustrative purposes, the main categories are as follows:

- ▶ Advanced case-law search engines
- ▶ Online dispute resolution
- ▶ Assistance in drafting deeds
- ▶ Analysis (predictive, scales)
- ▶ Categorisation of contracts according to different criteria and detection of divergent or incompatible contractual clauses
- ▶ “Chatbots” to inform litigants or support them in their legal proceedings

16. **Latvia** stated that it was exploring the possibilities of machine learning for the administration of justice. The main purpose would be to process court statistics to draw up provisional estimates of human and financial resources to be allocated.

17. Other activities carried out by legal tech companies have not been included in this classification because they involve little or no artificial intelligence processing: some sites offer access to legal information, “cloud” solutions, electronic signatures, etc.

18. A non-exhaustive list of legal services making use of artificial intelligence in their operations is set out below:

11. See summary bibliography in Appendix IV – substantial contributions from Benoît Charpentier as well as Giuseppe Contissa and Giovanni Sartori (https://media.wix.com/ugd/c21db1_14b-04c49ba7f46bf9a5d88581cbda172.pdf) and Emmanuel Barthe (<http://www.precisement.org/blog/Intelligence-artificielle-en-droit-derriere-la-hype-la-realite.html#nb14>) (French only)

Software	State	Type
Doctrine.fr	France	Search engine
Prédicte	France	Analysis (except criminal cases)
Case Law Analytics	France	Analysis (except criminal cases)
JurisData Analytics (LexisNexis)	France	Search engine, Analysis (except criminal cases)
Luminance	United Kingdom	Analysis
Watson/Ross (IBM)	USA	Analysis
HART	United Kingdom	Analysis (criminal, risk of reoffending)
Lex Machina (LexisNexis)	USA	Analysis

2. Overview of open data policies relating to judicial decisions in the judicial systems of Council of Europe member States

The availability of data is an essential condition for the development of AI, enabling it to perform certain tasks previously carried out by humans in a non-automated manner. The more data available, the more AI is able to refine models improving their predictive ability. An open data approach to judicial decisions is therefore a prerequisite for the work of legal tech companies specialising in search engines or trend analysis ("predictive justice").

Processing of these data raises a number of issues, such as changes in the formation of case-law and protection of personal data (including the names of professionals).

19. Computer-raised data are said to be the "oil" of the 21st century as their use and cross-referencing are producing a whole new wealth. Even though some stakeholders and authors dispute this argument, the global successes of the digital industry over recent decades have confirmed the enormous growth potential of this field of activity.

20. The quantification of human activities, now on a global scale, could not fail to touch on the data produced by the public sector. This is what has prompted the movement to open up public data, based on much older imperatives which are the founding principles of our constitutional states.

21. The major change in recent years has been brought about by the emergence of downloadable public data (open data), notably in the context of the “Partnership for Open Government” (OGP). The OGP is a non-governmental organisation bringing together nearly 70 member States (including many of the Council of Europe member States) with representatives of civil society and digital giants. The aim of this openness is to improve the transparency of public activities, encourage citizens in the development and assessment of public policies and guarantee the integrity of public service and those who perform it by processing considerable amounts of information, organised into databases (big data).

2.1. Definition of open data on judicial decisions

22. First of all, let us redefine the notion of open data before dealing with the question of the impact of allowing open data on judicial activity. Firstly, there is often confusion between access to information and access to data (more precisely, access to information in the form of database)¹².

23. A certain amount of public information, requiring wide publicity, is already disseminated using information technology. In France, the government site Légifrance.fr is the main online source of certified public information, comprising not only legislative and regulatory texts but also case-law and information on appointments to public posts. This unitary information, although available on the Internet, differs completely from direct access to data organised and included in a database that can be downloaded and processed by a computer.

24. Open data therefore only involves the dissemination of “raw” data in structured computer databases. These data, aggregated in whole or in part with other structured sources, constitute what we call big data. The Council of Europe Convention 108 Consultative Committee defines big data as “the growing technological ability to collect, process and extract new and predictive knowledge from great volume, velocity, and variety of data. In terms of data protection, the main issues do not only concern the volume, velocity, and variety of processed data, but also the analysis of the data using software to extract new and predictive knowledge for decision-making

12. The data are meaningless letters and numbers. Information is data included in a context. It is the context that gives meaning to the data. We can guess that 2005 is a year, but without context, we do not know. But in “in 2005, we completed 3 projects” the context gives meaning to the number. Therefore, ‘open data’ is not data in the sense of the definition, but information. Similarly, large data are also large amounts of information, not data.

purposes regarding individuals or groups. For the purposes of these Guidelines, the definition of Big Data therefore encompasses both Big Data and Big Data analytics”¹³

25. **As this definition shows, open data should not be confused with their means of processing.** Some of the discourse on this issue actually relates to processing carried out by various advanced methods which are generally defined as data science. Predictive justice using artificial intelligence, advanced search engines applying extremely precise criteria and legal robots are all algorithmic applications which are fed with data but have nothing to do with the policy of open data itself.

26. However, this policy must be examined in the light of the possibilities it offers for further processing, whatever its nature. If certain data are filtered upstream, taking account for example, of the need for confidentiality and respect for privacy, subsequent risks of misuse appear to be reduced.

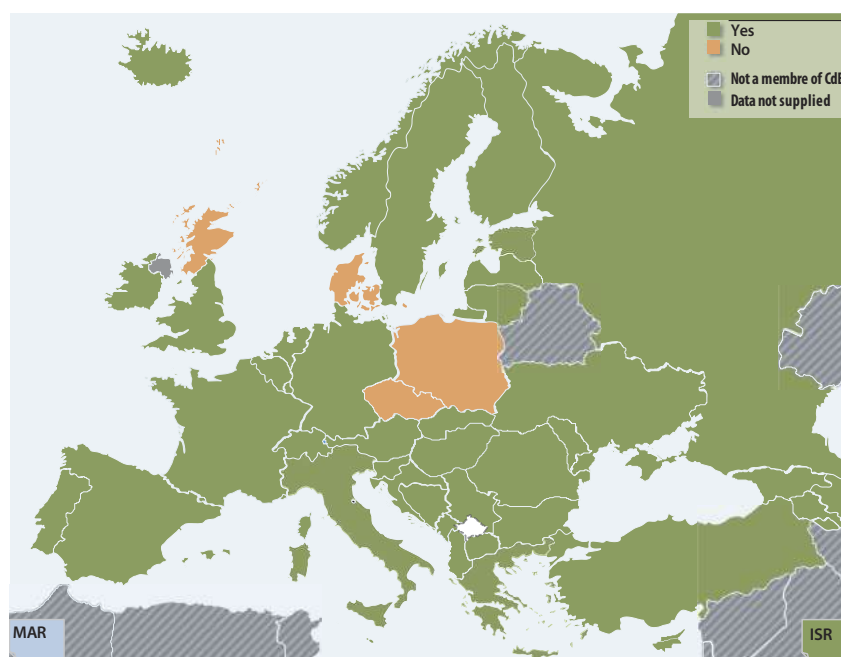
2.2. State of development of open data on judicial decisions in Council of Europe member States and consequences for the development of case law

27. What is the situation of the Council of Europe member States as regards open data on judicial decisions? The 2016-2018 CEPEJ evaluation cycle focused for the first time on the question of court decisions being provided in open data, for which some AI processing is used. The issue of data anonymisation or pseudonomysation within the European data protection framework¹⁴ provided by the General Data Protection Regulation (GDPR, EU Regulation 2016/679) and Council of Europe Convention No. 108 was the subject of a specific question designed to identify the measures implemented by member States and observers in this particularly sensitive area.

13. T-PD(2017)1, Guidelines on the protection of individuals with regard to the processing of personal data in a world of big data.

14. General Data Protection Regulations (DGPS, EU Regulation 2016/679, and Council of Europe Convention No 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data).

Figure 1. Table showing CoE member States having introduced open data on judicial decisions in 2016 (Q62-4 of the evaluation scheme)



28. Of all the States and observers surveyed, only 5 declared that they had not implemented an open data policy for judicial decisions in 2016. While this response rate should be put into perspective, since some answers confused public access to decisions with open data (**Armenia, Belgium, Bosnia and Herzegovina, Luxembourg, Russian Federation, Spain, Turkey**), it reveals, on the one hand, a desire for transparency on the part of European judicial institutions and, on the other hand, a desire on the part of many countries to make court decisions public and thus make it possible to deal with them later using AI tools. This also requires efforts by the institutions concerned, since a number of technical measures must be put in place to this end. In **France** specifically, some administrative case law is already available for download on the site *data.gouv.fr* (see below).

29. With regard to the protection of personal data, 23 countries declared that they are pseudonymising¹⁵ at least some types of disputes (e.g. personal status, family status) by erasing data making the parties or witnesses identifiable (names, addresses, telephone numbers, identity numbers, bank account numbers, tax numbers, health status, etc.). This work appears to be the responsibility of judicial personnel (e.g. **Israel, Republic of Moldova**) or public officials (e.g. **Bosnia and Herzegovina, Spain**). **Bosnia and Herzegovina** and **Hungary** on the other hand stated to publish the names of professionals.

30. However, there is a real difficulty in measuring the impact of open data on the efficiency and quality of justice. As indicated above, the initiative to re-use these data is essentially private, targeting a professional clientele (lawyers, legal departments), and an exclusively intergovernmental activity is probably not the best means of fully identifying such positive results.

31. The situation in France is representative of the questions raised by this approach and reveals a number of the issues at stake. First of all, it is important to underline that France enacted legislation in 2016 imposing a compulsory framework for the open data dissemination of decisions on its courts.

32. Articles 20 and 21 of the Law for a Digital Republic¹⁶ broke with the previous logic¹⁷ of selecting which decisions from judicial and administrative courts and tribunals were to be disseminated if they were “of particular interest”. Under the new French law, however, the opposite principle that everything is publishable has been set, except in specific cases identified by law (for judicial decisions) and with due regard for the privacy of the persons concerned. Provision is made, however, for judicial and administrative decisions to be published only after an analysis has been made of the risk of re-identification of the persons involved.

15. As defined by the T-PD in its “Guidelines on the protection of individuals with regard to the processing of personal data in a world of big data” T-PD(2017)1, pseudonymisation refers to the processing of personal data “in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.”

16. This law was adopted in order to bring French law into line with Directive 2013/37/EU of the European Parliament and of the European Council of 26 June 2013, which in turn amended the Council Directive of 17 November 2003 on the re-use of public sector information (the “PSI Directive”).

17. Article R433-3 of the Code of judicial organisation

33. There are numerous advantages to wide dissemination, promoted in particular by the Court of Cassation at two conferences in October 2016 and February 2018. They include greater awareness of judicial activity and case law trends, the increased quality of a justice system that knows it is being observed and the creation of a completely new factual reference base.

34. This theoretical desire to “regulate” case-law through digital leverage raises however a number of general questions that are relevant to all countries considering a similar approach. Firstly, it should be placed in the context of some of the principles set out by the European Court of Human Rights in cases of differences in domestic case-law. The Court clearly emphasises the need to balance legal certainty, which makes decisions more predictable, against vitality in judicial interpretation.¹⁸

35. This desire is also tempered by several aspects, of technical nature in first place:

- a. **Collection of all judicial decisions eligible for publication is not necessarily well co-ordinated between all levels of courts:** some of the business applications in European courts have not been designed for this, particularly as regards first instance decisions, and some countries will have to set up new processes for collating judgments if they want collection to be exhaustive;
- b. **Despite some promising developments, a fully effective automated post-identification mechanism that can prevent any risk of identification or re-identification has not yet been devised;**

And equally substantive aspects:

- c. **Thought should be given to the transformation of the very logic of the production of case-law.** What is the value of the “standard” resulting from the number of decisions given on a specific matter? Does this “standard” add to the law? If so, is this a new source of law?
- d. **Should judges be asked to provide additional explanations for decisions to deviate from this norm?**¹⁹ This question is not insignificant and does not prompt the same reply in all judicial systems. In French civil law,²⁰ “judges settle disputes in accordance with the rules of law applicable thereto”. In the French context, if they were asked to justify their decisions by providing all the reasons for which they deviated

18. *Greek Catholic parish Lupeni and Others v. Romania* [GC]. No. 76943/11, 29/11/2016, § 116.

19. Eloi Buat-Menard and Paolo Giambiasi, «La mémoire numérique des décisions judiciaires», *Dalloz Reports*, 2017, p. 1483. (French only)

20. Article 12 of the French Code of Civil Procedure

from the supposed majority case-law trend on how to resolve the dispute (while complying with the relevant rules of law), would this not be tantamount to removing them from office? Judges would not only settle disputes in accordance with the rules of law but also in accordance with case law trends derived from statistics compiled by a digital tool (which could also be biased or developed without external control by a private operator, see sections 6.1 and 9 below). The question that arises goes far beyond a specific national feature, relating more broadly to the place of AI tools in judges' decision-making processes. Note that this analysis is limited to a processing of case-law by algorithms and in no way refers to the role of case-law in the sources of law or the authority of case-law precedents, which, moreover, are well established principles at European level.²¹

- e. **Moreover, would it not be the case that if norms were established according to the majority trend, judicial decisions would be rendered uniform, and no longer be ordered according to the hierarchy of the courts from which they emanate, disregarding the significance of the decisions of supreme courts, which are the guarantors of the uniform interpretation of law in many European States?** What would the relationship be between norms and case-law? Since they derive from the majority, would they become a criterion for these courts when determining their own case-law, which they would in turn have to justify when they deviate from the majority view?
 - f. **Finally, isn't there a risk of court decisions being written down according to a reproductive logic?** While court decisions are likely to evolve according to the evolution of a normative framework (national, European or international), the case law available for reference (for example, from supreme courts and European courts) or the socio-economic context, would not the norm resulting from the majority become a standard to which judges would be encouraged to refer without question, with an induced effect of the excessive standardisation of judicial decisions?
36. **Doubts might also be raised about the consequences for users, who are supposed to benefit directly from the transparency of activities:** will they really benefit from the publication of all judicial decisions on the

21. The ECJ has stated, on the conditions governing the liability of a Member State for the content of a decision of a supreme national court, that "an infringement of Community law will be sufficiently serious" and must give rise to compensation for the damage "where the decision concerned was made in manifest breach of the case-law of the Court in the matter" (ECJ, Case C-224/01, Koebler, §56).

Internet or are they not more likely to find themselves overwhelmed by the mass of decisions, without necessarily deriving qualitative benefits from them, for lack of legal knowledge and the critical apparatus to analyse them?

37. Paradoxically enough, one may wonder of open data be in position of delivering meaningful information since it will certainly make it possible to download a considerable set of raw data with a simple click on a link, but the meaning will remain totally obscure for the majority of citizens.

38. The first requirement for open data to be transparent and informative is for third parties to be able to use them for analysis and development. An economic model in which public case law data, which are the subject of work by judicial authorities in order to make them technically “readable” by the AI and compliant with the legal requirements for the protection of personal data, would be processed free of charge by the private sector and then sold by to the courts, professionals and citizens, therefore deserves to be questioned and even criticised.

2.3. Protection of personal data in open data policies for judicial decisions

2.3.1. The names of parties and witnesses

39. In order to strike a fair balance in the digital age between the need to make judicial decisions public and respect for the fundamental rights of parties or witnesses, their names and addresses must not appear in published decisions, particularly in view of the risk of misappropriation and re-use of such personal information and the particular sensitivity of the data likely to be contained in the decisions.²² Automated processes can be used to systematically to conceal such information.

40. Other identifying information may also be obscured (for example, telephone numbers, e-mail addresses, dates of birth, children’s given names, rare given names, nicknames and place names). In terms of personal data protection principles, this concealment amounts to a simple pseudonymisation of the data, not complete anonymisation. The volume and variety of information contained in court decisions, combined with the growing ease

22. ECHR, *Z. v. Finland* No. 22009/93, §§95 et seq. and the European Commission’s Green Paper on public sector information in the information society (COM(1998)585) (“If special precautions are not taken, case-law databases, which are legal documentation instruments, can become information files on individuals if these databases are consulted to obtain a list of court judgments on a specific individual rather than to find out about case-law”).

of cross-referencing with other databases, makes it impossible, in practice, to guarantee that the person concerned cannot be re-identified. In the absence of such a guarantee, these data cannot be qualified as anonymous and are therefore subject to personal data protection rules.

41. Some especially sensitive items of personal data warrant particular attention, as provided for in Article 6 of Convention 108. This applies to data revealing ethnic or racial origin, political opinions, trade union membership, religious or other beliefs, physical or mental health or sex life, which are considered intimate details.

42. Court decisions may contain other, very varied, types of personal data that fall into this category of sensitive data. Courts dealing with criminal matters are particularly likely to process sensitive data such as those on criminal proceedings and convictions. All this sensitive data therefore deserves special vigilance. Their mass dissemination would present serious risks of discrimination, profiling²³ and violation of human dignity.

2.3.2. The names of professionals, including judges

43. Obviously, knowing how a judgment will be arrived at is an essential element for lawyers in predicting the outcome of a case, and they believe that knowing one's judge is sometimes almost as important as knowing the law. They have long tried to make comparisons between panels of judges, more or less empirically, so as to give better advice to clients dealing with a particular judge or panel of judges.

44. This method was sufficient when a lawyer was only speaking before a limited number of courts, but the gradual loosening of local restrictions on the bar in many countries and the freedom to move and work within the European Union make it reasonable for any national or even European lawyer to want to know the case-law of each national or European jurisdiction in which he is likely to plead in full detail.

45. We cannot therefore exclude the possibility that, in the future, highly useful, and hence very expensive, machine learning applications will be much more effective than the experience and "good sense" of litigation lawyers working through cases in the traditional way. The use of such applications could further accentuate the distortion of competition and inequality

23. Profiling is defined in section 4 of the GDPR. It is processing an individual's personal data in order to analyse and predict his/her behavior or his/her situation, such as determining his/her performance at work, financial situation, health, preferences, lifestyle habits, etc.

of arms between law firms that have or have not used such “predictive” case-law analysis software.

46. There is a real risk that, for the sake of such a competitive advantage, the principle of a fair trial established by law will be undermined. The possibility of judge profiling through cross-referencing of public and private data could allow private companies and their lawyers to engage in even more forum shopping practices. This tactic has already been observed for a long time in the United States and in France for press offences and violations of privacy in the press, where plaintiffs have already been known to choose the court which appears to award the highest amounts of damages and interest.

47. In addition, much of the discourse on this subject confuses open data with the need to publish a certain amount of public information. As a result it is sometimes argued that the names of professionals should appear in open data for the sake of publicity and transparency.

48. However, the provision of computerised case-law data is a totally separate issue from the principles of publication of original or certified copies of decisions. The objective of open data is to allow automated processing of case-law under a low-cost operating licence. As stated previously, this provision is made in the form of a comprehensive computer database, which is opaque and not directly understandable to citizens.

49. This provision obviously does not meet the need to publicise the names of professionals who have contributed to a specific decision. It should be noted that, in law, mentioning the names of judges in judicial decisions is a common obligation for member States, in connection with the principle of public trial set out in Article 6(1) of the European Convention on Human Rights, in order to ensure the objective impartiality of judges (who must be identifiable, and lawfully appointed and assigned to the duties they perform) and compliance with procedural rules (e.g. publicity and collegiality).²⁴

50. The answer to the question of the legitimacy or not of publishing the names of professionals²⁵ in open data therefore has nothing to do with the obligation to publish the names of professionals in decisions. Rather, it

24. ECHR, *Vernes v. France*, No. 30183/06 as regards the identification of the judges who rendered the decision, *Pretto v. Italy*, No. 7984/77 for the publication of the judgment, *Kontalexis v. Greece*, No. 59000/08, § 38, *DMD GROUP, a.s. v. Slovakia*, No. 19334/03, § 66, *Miracle Europe KFT v. Hungary*, No. 57774/13, § 58 for the right to a court established by law in conjunction with the right to an impartial judge.

25. These questions may also apply to *Rechtspfleger* who make judicial decisions and to clerks, assistants to the judge mentioned in the composition of the formation of the court (albeit to a lesser extent).

seems that the challenge lies in reconciling often conflicting requirements: making public activities transparent by allowing citizens to know and evaluate their judges, on one hand, while protecting the privacy of professionals (whose functions should not limit their fundamental guarantees in this field), on the other hand. There are rigorous challenges where it comes to guaranteeing the impartiality of judges and even judicial institutions as a whole, which open data policies are actually designed to meet²⁶. What practical measures can be taken to protect them from potential attempts at destabilisation which cross-reference judges' personal data included in databases with other sources (social networks, commercial sites) to try to identify hypothetical political, religious and other biases?

51. These questions do not arise in the same form everywhere in Europe and depend on the specific features of the judicial system concerned (and on the nature of the judiciary's career management body), the collegial nature or not of the judgment and the level of court concerned. In Switzerland, for example, where judges are elected, publication is a guarantee of transparency and social responsibility of judges vis-à-vis citizens and political groups. This information is already available in online search engines (which are not strictly speaking open data).²⁷

52. Nor do these questions arise in the same form depending on the level of jurisdiction. The value of characterising the case-law of lower court judges may not be the same as for judges of supreme courts or international courts. For example, the European Court of Human Rights authorises searches for judgments by the names of the judges members of the decision panel, but does not allow the calculation of statistics relating to a particular judge.²⁸ On the other hand, in countries where the judicial bodies are unfamiliar with the practice of dissenting opinions (existing within this international court), it may seem unfair to assign a judge personal responsibility for a decision which he voted against during deliberation in a collegial court.

53. These debates were well defined by a study mission conducted in France by Professor Loïc Cadiet. The conclusions of the mission remain limited since they do not recommend prohibiting publication but reserving it for certain types of litigation and ruling it out for others (for example, for

26. See *ECHR Previti v. Italy*, No. 45291/06, §§ 249 et seq., which recalls the principles of objective impartiality of the judge.

27. See the example of the Swiss Federal Court, whose case-law can be downloaded: <https://www.bger.ch/fr/index/jurisdiction/jurisdiction-inherit-template/jurisdiction-recht.htm>; or, for the cantons: <http://ge.ch/justice/dans-la-jurisprudence> (Canton of Geneva for example).

28. The decisions of the European Court of Human Rights are collegial. Publications include any dissenting opinions.

specialised criminal matters). The possibility of publishing only the names of supreme court judges was proposed, although it was conceded that this might result in a “one-way trip”.

54. As it stands, a simple precautionary principle could be applied in order to assess the interest, by type of litigation and degree of jurisdiction, in publishing the names of professionals in a downloadable database. Likewise, we cannot rule out the possibility of judicial institutions themselves or authorised third parties exploiting this information outside the open-data context to find out about case-law.

55. Publication of this sort would in any event still need to be examined in light of applicable international European rules on the protection of personal data, such as those of Convention No. 108 and the European Union’s GDPR and Directive 680/2016.

3. Operating characteristics of artificial intelligence (machine learning) applied to judicial decisions

Natural language processing and machine learning are the two techniques at the heart of processing of judicial decisions using artificial intelligence.

In most occasions, the objective of these systems is not to reproduce legal reasoning but to identify the correlations between the different parameters of a decision (for example, in a divorce claim the length of marriage, the income of spouses, the existence of adultery, the amount of the benefit pronounced, etc.) and, through the use of machine learning, to infer one or more models. Such models would then be used to “predict” or “foresee” a future judicial decision.

The purpose of this chapter is to clarify the intrinsic nature of software described as “predictive” – sometimes “beta” versions, i.e. in a development phase – both in terms of its potential and its limitations. It is intended to provide a simple explanation of machine-learning algorithms, which are at the heart of the automated analysis of case-law.

3.1. The theoretical functionalities of “predictive justice” software

56. By way of an introduction, we should briefly review the features promised by “predictive” software. They propose to establish the probabilities of the success (or failure) of a case before a court. These probabilities are established through the statistical modelling of previous decisions using methods from two broad computer science domains: natural language processing

and machine learning. These modelling approaches are often referred to as AI; in reality, these are “weak” AI (see the glossary, page

57. It should be made clear straight away that the term AI is debated by experts as it leads to many ambiguities. The term AI has now entered our everyday language to describe a diverse range of sciences and technologies that allow computers to beat the best champions in the game of Go,²⁹ to drive a car, to converse with humans, etc. Researchers prefer to identify the different applications through the exact technologies underlying them, including machine learning, and sometimes refer to all of these highly specialised AI resources as “weak” (or “moderate”) AIs. These are distinguished from an ultimate – still totally theoretical – goal of creating a “strong” AI, i.e. a self-learning machine capable of automatically comprehending the world in general, in all its complexity.

58. In relation specifically to justice, predictive justice systems are designed for use by legal departments, insurers (both for their internal needs and for their policyholders) as well as lawyers for them to anticipate the outcome of litigation. Theoretically, they could also assist judges in their decision-making.

59. They provide a graphic representation of the probability of success for each outcome of a dispute based on criteria entered by the user (specific to each type of dispute). These systems claim to be capable to calculate the likely amount of compensation distributed by the courts.

3.2. The practical functioning of artificial intelligence: statistical machines constructing models based on the past

60. A distinction must be made from the outset between what is a “prediction” and what is a “forecast”. Prediction is the act of announcing what will happen (*prae*, before – *dictare*, say) in advance of future events (by supernatural inspiration, by clairvoyance or premonition). Forecasting, on the other hand, is the result of observing (*aiming*, seeing) a set of data in order to envisage a future situation. This abuse of language and the spread thereof seems to be explained by a transfer of the term from the “hard” sciences, where it refers to a variety of data science techniques derived from mathematics, statistics and game theory that analyse present and past facts to make hypotheses about the content of future events.

29. <https://www.nature.com/articles/nature16961>

61. It should also be noted that the logic of the operation of predictive justice software is essentially based on either generative (commonly referred to as Bayesian) or discriminative methods which eventually try to estimate the current or future range of values of a variable (e.g. the outcome of a trial) from the analysis of past examples.

62. Generally speaking, it is also important to keep in mind the anthropomorphic notion that computing machines are intelligent and that their designers have managed to slip a mind inside their mechanisms.³⁰ Unfortunately, this idea still permeates many analyses of predictive justice that lend these devices immediate or future capabilities for the near replication of human intelligence. This context, fuelled every day by a further series of revolutionary AI advances,³¹ therefore leads us all to approach these predictive tools with a certain dose, conscious or not, of mysticism, sometimes stating that what is not entirely possible today will inevitably be possible tomorrow.

63. The “strong” AIs of science-fiction literature do not exist. This type of AI, which would be equipped not only with intelligence but also with conscience, remains purely fictional. The machine learning systems currently being developed are described as “weak” AIs and **are capable of extracting complex patterns and learning from large volumes of data efficiently and often with high levels of predictive accuracy.**

64. To overcome any instinctive or fantasised considerations, these information processing and analysis technologies must be understood on the basis of the following three concepts.

- **AI is not a single, homogeneous object:** it is actually an assemblage of sciences and techniques (mathematics, statistics and computer science) capable of processing data to design very complex computer processing tasks.
- **AI engines do not produce intelligence per se but proceed using an inductive approach:** the idea is to associate in a nearly automated way a set of observations (inputs) with a set of possible results (outputs) using various preconfigured properties. Specifically for predictive justice, the engine builds links between the different lexical groups composing judicial decisions. These groups are correlated between those identified

30. Dominique Cardon, *A quoi rêvent les algorithmes, nos vies à l'heure des big data*, La République des idées, Editions du Seuil, 2015, p. 58.

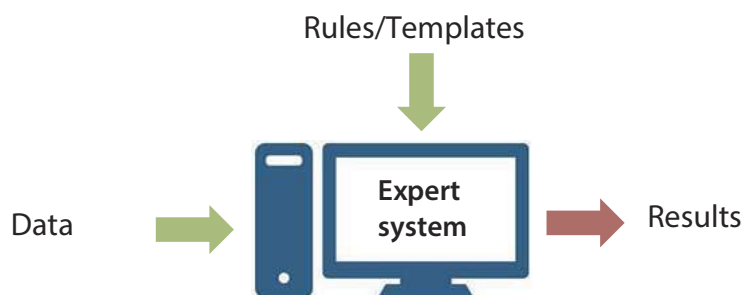
31. “AlphaZero : l’IA de DeepMind apprend et maîtrise le jeu d’échecs en moins de 4 heures”, *Génération NT*, article published on 8 December 2017, [Online], <https://www.generation-nt.com/alphazero-ia-deepmind-maitrise-echechs-4-heures-actualite-1948919.html> (page accessed on 14 December 2017).

at the input stage (facts and reasoning) and those at the output stage (the operative part of the decision) then classified.

- The reliability of the model (or function) built strongly depends on the quality of the data used and the choice of machine learning technique.

65. It is worth returning briefly to the ancestors of these systems – expert systems – which for their part relied on processing rules written by a computer scientist. Expert systems (ES) developed rapidly in the late 1980s and 1990s, especially in medicine and finance.³² These systems were able to answer specialised questions and reason using known facts, executing pre-defined encoding rules in an engine. Despite the success of ESs, such as Deep Blue against Garry Kasparov in a series of chess games in 1997, these systems ended up failing, notably because they were unable to interpret “the infinite variety of situations and contexts”³³ and became ineffective beyond 200 to 300 encoding rules, both in terms of execution performance and maintenance (the reasoning followed by the system became almost impossible to apprehend for its designers).

Fig. 2: The old expert systems were programmed with rules reproducing the logic of legal reasoning

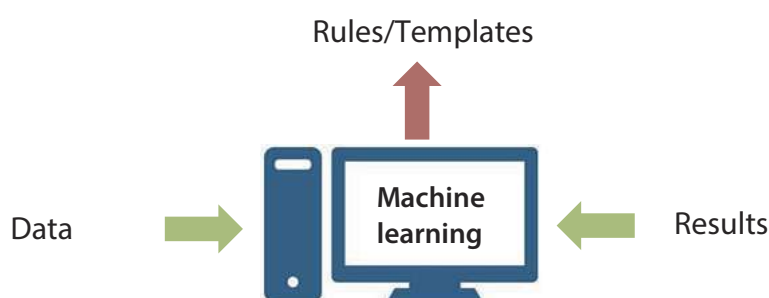


32. An example is High Frequency Trading (HFT), which is a type of financial transaction carried out at high speed by software based on algorithms. In 2013 and in Europe, 35 % of transactions were already carried out with HFT. In the USA, HFT represents 70 % of trading volumes on the equity market. In view of the automation of transactions, the fall is even faster in the event of crashes, as was the case during the 2007 financial crisis.

33. Hubert Dreyfus, *What Computers Still Can't Do. A Critique of Artificial Reason*, Cambridge, The MIT Press, 1992 cited by Dominique Cardon in *A quoi rêvent les algorithmes, nos vies à l'heure des Big Data*, p.59.

66. Today, the idea is no longer to write reasoning rules mirroring human reasoning, as with older expert systems,³⁴ but to let machine learning systems themselves identify existing statistical models in the data and match them to specific results.

Fig. 3: Machine learning alone produces models by automatically searching for correlation results



67. In doing so, these systems do not reproduce or model reasoning (such as legal reasoning). For example, today's online translators do not carry out abstract reasoning. They infer a probable estimate of the best match between groups of lexical structures and translations already done. Users' actions obviously contribute to the improvement of the match search, but the machine learning algorithm does not actually perform a translation by understanding the meaning of the sentences processed.³⁵

68. This example shows how the approach of lawyers and researchers can be different: a lawyer will seek to understand the effectiveness of pre-existing rules, which the researcher can only provide through the language of his science, and this is sometimes limited to interpreting thousands of parameters from a machine learning system.

69. Machine learning is a field of computer science in which computer programmes learn from experience. Algorithms make a machine perform a

34. Written processing rules could be based on specific code in different programming languages such as LISP or editors to model rules. See for example Radboud Winkels, "CLIME : Un projet de développement de serveurs juridiques intelligents", in Danièle Bourcier, Patricia Hasset and Christophe Roquilly (eds.), *Droit et intelligence artificielle*, 2000, Romillat, p. 59.

35. Li Gong, "La traduction automatique statistique, comment ça marche ?", *Interstices.info*, published on 29 October 2013, [Online], https://interstices.info/jcms/nn_72253/la-traduction-automatique-statistique-comment-ca-marche (page accessed on 14 December 2017).

training process, like a child learning in its environment. In summary, these learning techniques may or may not be supervised by a human. The most prominent category of machine learning is currently that of reinforcement learning: the machine alone reaps virtual “rewards” if action produces the expected result in a given environment. Machine learning methods include neural networks (or their more complex version known as deep learning), which have been quite widely publicised in view of their autonomy and their quite striking applications, such as the one succeeding in obtaining high scores on old Atari 2600 video games³⁶ (based solely on pixel positioning and scores). Other practical applications for these technologies are already affecting our daily lives and are beginning to appear in the professional world of justice.³⁷

70. However, these algorithms remain highly specialised in one particular task and present discernment problems when faced with chaotic situations or with insufficient data to allow prediction (such as the actual understanding of natural language). In the social sciences, to which law and justice belong, failure would even appear inevitable in the absence of a convincing model of cognition. For Pierre Lévy, artificial intelligence is in fact content to provide a heterogeneous toolbox (logical rules, formal syntaxes, statistical methods, neural or socio-biological simulations, etc.) that does not offer a general solution to the problem of mathematical modelling of human cognition.³⁸ Thus, real predictive learning should in reality be based on a good systemic representation of the world, which AI researcher Yann LeCun believes is a fundamental scientific and mathematical issue, not a question of technology.³⁹

71. Moreover, the uniqueness of current big data processing systems is that they do not attempt to reproduce our cognition model but produce context statistics on an unprecedented size of data, without any real guarantee of excluding false correlations.⁴⁰

36. <https://www.nature.com/articles/nature14236>

37. Artificial intelligence by IBM Watson providing a service to the medical field and providing a search tool called “Ross”, presented as a virtual lawyer – Roos, “Do more than humanly possible” [Online], <http://rossintelligence.com> (page accessed on 14 December 2017).

38. Pierre Lévy, “Intelligence artificielle et sciences humaines”, *Pierre Levy's blog*. 8 October 2014. [Online], <https://pierrelevyblog.com/2014/10/08/intelligence-artificielle-et-sciences-humaines/> (page accessed on 30 December 2017).

39. Yann LeCun, “Qu’est-ce que l’intelligence artificielle”, research published on the Collège de France website, [Online], https://www.college-de-france.fr/media/yann-lecun/UPL4485925235409209505_Intelligence_Artificielle__Y._LeCun.pdf (page accessed on 14 December 2017).

40. Dominique Cardon, *op. cit.*, p.60.

72. While it is commonly accepted that statistical and probabilistic models are improved by increasing the data feeding them, some mathematicians have warned of the risks of increasing false correlations (i.e. links between factors having absolutely no causal link) in big data. The mathematicians Cristian Sorin Calude and Giuseppe Longo point to the risk of a deluge of false correlations in big data: the larger a database used for correlations, the greater the chances of finding recurrent patterns and the greater the chances of making errors.⁴¹ What may appear as regularities for an AI (recurrent links between different data, concepts, contexts or lexical groups) may actually be random. Even if the argument of the two mathematicians should not be generalised too hastily, they note that in certain vast sets of numbers, points or objects, regular random patterns appear and it seems impossible to distinguish them algorithmically from patterns revealing causalities.⁴²

73. Lastly, algorithms, whatever their current level of sophistication, still automatically boil down to the interactions established in the learning phase and hence to their past. The content and quality of the data flows used in the composition of the calculations is therefore fundamental in understanding the results obtained and in identifying possible analytical biases. There are many challenges here again since, in an exhaustive approach, the analysis of the greatest possible amount of data relating to an activity will produce results whose meaning has to be clarified with regard to all the factors that have had an influence. In a more restrictive approach in which incoming data is sampled, risks will also arise from the trade-off biases required to select one data over another.

4. Can artificial intelligence model legal reasoning in advance?

Artificial intelligence circumvents the difficulties encountered with older expert systems: they do not attempt to manually mirror legal reasoning, whose reproduction is not in itself an objective for them. Machine learning leads to categorisations between the different parameters identified by the designers or those discovered by the machine.

74. Continental legal systems are far removed from the ideal of rationality that embodied, for example, the 1804 Civil Code in France. There are a multitude of sources that do not fit together perfectly and which relate to

41. Cristian Sorin Calude, Giuseppe Longo, "Le déluge des corrélations fallacieuses dans le big data", in Bernard Stiegler (dir.) (2017), *La toile que nous voulons*, FYP editions, p.156.

42. Theory of "Ramsey", Ronald L. Graham, Joel H. Spencer – "Ramsey Theory", *Scientific American*, vol.263, No.1, July 1990, p112-117 cited by Cristian Sorin Calude, Giuseppe Longo, "Le déluge des corrélations fallacieuses dans le big data", *ibid.*

a set of rules whose meaning remains undetermined, which the legal theorist Herbert L. A. Hart called the “open texture of law”.⁴³ Common law systems, although considered more economically efficient because they adapt dynamically to new legal needs,⁴⁴ are also evolving and offer no more legal certainty. Legal rules therefore do not evolve in a linear fashion, distinguishing them from empirical laws (those of the “exact sciences”), where each new rule generally complements the previous ones and does not invalidate a complete set of reasoning.

75. It should be stressed that, faced with this general complexity of legal systems, the old IT expert systems quickly reached their limits once 200 to 300 logical rules were nested. The division of the law into rules of production was not sufficient to provide a valid representation of the body of knowledge and methods that guide a lawyer.

76. Consequently, it has proved just as impossible to model the reasoning of judges on a computer as it has to model a positive legal system. As the theory of law has highlighted, judicial reasoning is above all a matter of assessment and interpretation, of the proven and relevant facts of a case, of the applicable rules of law (textual or precedents) – the meaning of which remains, as has been said, indeterminate –,⁴⁵ and of the subjective interpretation by judges of the concept of equity, which should undergo new changes in Europe with the requirement of a proportionality review encouraged by the European Court of Human Rights.

77. Xavier Linant de Bellefonds has stressed that the complexity of the law lies in its teleological and contentious nature: two coherent arguments can lead to different judgments according to two different priorities.⁴⁶

78. This is because the famous legal syllogism is more a way of presenting legal reasoning than its formal translation. It does not reflect the full reasoning of the judge, which is in fact made up of a multitude of decision-making factors, cannot be formalised a priori, and is sometimes based on his discretion: what are the relevant facts? Are these facts proven? Which rule applies to them? What is the meaning of this rule with regard to the case to be decided? Which source should prevail between a range of conflicting

43. Herbert L. A. Hart (1976), *Le concept de droit*, Saint-Louis university departments, Brussels.

44. <https://www.contrepoints.org/2014/08/15/177160-common-law-contre-droit-civil-experience-francaise-de-lancien-regime>

45. On these questions, Michel Troper (2001), *La théorie du droit, le droit, l'Etat*, PUF, Paris, spec. p. 69-84.

46. Xavier Linant de Bellefonds (1994), “L'utilisation des systèmes experts en droit comparé”, *Revue internationale de droit comparé*, Vol. 46, No. 2, p. 703-718.

sources? The overall consistency of judicial decisions is never achieved and is more a matter of an a posteriori account which judges use in their reasoning, having more of a function of convincing themselves of the validity of a specific solution than of describing strictly and objectively all the stages that resulted in the decision made.

79. However, this work of interpretation is precisely what machine learning techniques do not do – and do not attempt to do – today, since they carry out, as we have seen, automated processing where the correlation of vast amounts of information is supposed to replace the understanding of the real causalities of a decision. They make not attempt to formalise legal reasoning but hope that the models captured by them can anticipate the likely decisions of a judge in similar situations.

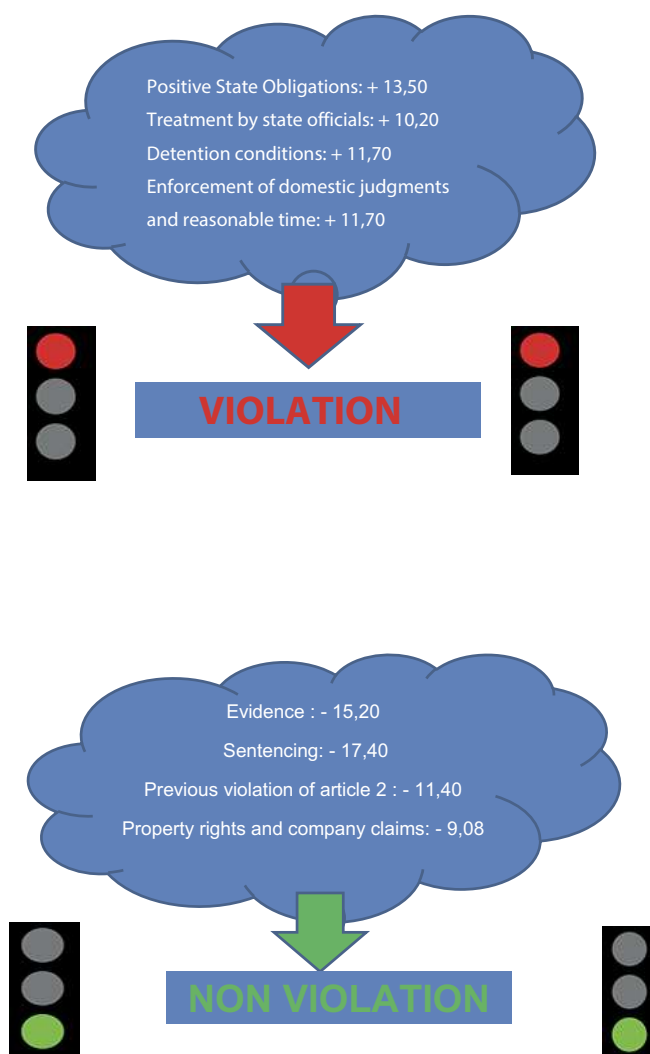
80. The results achieved by AIs are in reality unrelated to the question of the legal conformity of a particular solution and cannot discriminate between legal and illegal arguments.

81. A review of the work of the University College of London (UCL) on the case-law of the European Court of Human Rights confirms this diagnosis. The UCL study assumed that a simple automatic learning model could predict the outcome of a case with 79 % accuracy for that particular court. The machine learning model proved to be more accurate with regard to the descriptive part of the facts of the decisions studied than with regard to the reasoning relating to the application of the Convention to the case in question.⁴⁷

82. The examination of the terms correlated with the finding of a violation (with a positive weighting) and those correlated with a non-violation of the European Convention on Human Rights (with a negative weighting) sheds unambiguous light on the mechanics at work, which is in no way comparable with legal reasoning.

47. Work on a sample of 584 decisions of the European Court of Human Rights: Nikolaos Aletras, Dimitrios Tsarapatsanis, Daniel Preotiuc-Pietro, Vasileios Lamps, "Predicting judicial decisions of the European Court of Human Rights: a Natural Language Processing perspective", published on 24 October 2016, [Online], <https://peerj.com/articles/cs-93/> (page accessed on 14 December 2017).

Fig. 4: Illustration of UCL's work – Theoretical weight allocated to words or terms according to their link with findings of violation or non-violation of the European Convention on Human Rights



83. The choice of the relevant facts and their interpretation constitute one of the elements of the judge's decision. In other words, the UCL study was in reality only able to produce a probability with lexical material largely derived from the reasoning and motivation of the judge and not with that assembled by the applicant on the basis of frequencies alone. Their AI established thus a high probability of correspondence between groups of words and a decision that had already been formalised and could only give rise to a limited number of possible outcomes. Under no circumstances can it alone reproduce the reasoning of European judges nor, above all, predict an outcome, on the basis, for example, of a future applicant's raw account before the Strasbourg court, whose application will be subject to a very strict admissibility examination (nearly 70,356 applications were declared inadmissible or struck from the list in 2017)⁴⁸ based largely on the application of standards of assessment (importance and seriousness of the complaint, etc.) leaving considerable leeway in decision-making.

5. Can AIs explain judges' behaviour in retrospect?

An a posteriori explanation of a judge's behaviour, in particular the revelation of bias, would require all the potentially causative factors to be identified through an interpretative framework and a contextualised analysis. The fact that, statistically, childcare is more often entrusted to mothers than fathers does not demonstrate a bias on the part of judges but reveals the need to mobilise different disciplines from the social sciences to shed light on this phenomenon.

84. From a scientific point of view, explaining a phenomenon or, as far as we are concerned, a piece of human behaviour, amounts to determining the causal mechanisms that led to this behaviour using a certain amount of contextual data.

85. This requires, in a very schematic way, the preliminary constitution of an interpretative framework, itself derived from the repeated observation of this type of event or behaviour in the presence of certain factors or elements. The interpretative framework is made up of the hypotheses or points of view adopted by the different social science disciplines. This is an additional analytical step that can be fed into algorithms, but which they cannot perform alone.

86. Some legal tech companies went further and thought they could identify possible personal biases of judges and feed suspicions of bias. The open data of the names of certain presiding judges of administrative courts and

48. Source: 2017 Statistical analysis of CEDH, January 2018.

administrative courts of appeal in France have made it possible to develop an indicator of the rejection rate of appeals against obligations to leave French territory taken by the administrative authorities. Some commentators have argued fervently that the alleged impartiality of judges was therefore cast into doubt by artificial intelligence.

87. But can such interpretations really be achieved on the basis of an algorithmic processing of court decisions? For there to be personal bias in judges' decision-making processes (differing from their personal and public statements in the case concerned) their behaviour, or in this case their decision, needs to be determined by their personality traits, opinions or religion. However, as has been said, such a causal explanation cannot simply be deduced from the probabilistic result provided by algorithms. On the contrary, it requires additional analytical work in order to isolate, among the many correlated factors (including the identity of members of the panel of judges), those that are truly causative. For example, the fact that a family court statistically decides more often that children should live with their mother does not necessarily reflect the judge's bias in favour of women, but rather the existence of psychosocial, economic and even cultural factors specific to the jurisdiction, such as the working time of each of the parents, their income, the local availability of collective childcare, whether or not the child is in school, whether or not one of the parents is in a new relationship or even simply the lack of interest by either parent to take care of a young child.

88. Similarly, decisions on expulsion from a country given by an administrative court located near a large detention centre cannot be compared fairly with those of a court that deals with such disputes only occasionally.

89. In addition, regardless of the court's location, the issue of the case-law of a single duty judge who only occasionally deals with a certain type of litigation, but who uses (or ignores) the case law of his colleagues, is particularly interesting and may legitimately raise the question of equality of citizens in judicial proceedings. However, the focus must remain on the remodelling or preservation of the collegial nature of the judicial system rather than classification or stigmatisation via machine learning tools.

90. What can be deduced from the personality of the president of a panel of judges in a collegial court, when his name is the only nominative information visible in open administrative court decisions?

91. Furthermore, how can we account for two distinct philosophical and cultural approaches to judicial decisions, whereby, in some European countries, including France, there is a culture of precedent and a detailed knowledge by judges of the factual databases of all 1st and 2nd instance decisions

(Ariane database) in the field of administrative justice, while other countries or systems favour the intellectual independence of each court, along with a desire to deal with each situation on a case-by-case basis?

92. Providing an accurate explanation of a court decision therefore requires a much more detailed analysis of the contingent data in each case and the applicable rules of law, rather than sustaining a vain hope that the mass of links will make sense.

6. How is AI to be applied in civil, commercial and administrative justice?

The state of development of machine learning techniques does not allow today to reach reliable results regarding the “prediction” of judicial decisions. On the other hand, their application in the field of civil, commercial and administrative justice is to be considered for the creation of scales or the pre-litigation resolution of disputes online, when a later appeal to the judge remains possible.

93. As we can see, the first question raised by such use of artificial intelligence is not so much whether it is beneficial or harmful, desirable or otherwise, but whether the algorithms proposed can achieve the type of result sought. The conclusions from the experiments conducted in the Douai and Rennes appeal courts in France clearly demonstrate that in the presence of a thought-provoking discourse promoting a product of AI can hide unacceptable design flaws and totally erroneous analysis results.

94. Regardless of the quality of the software tested, anticipating judges’ decisions in civil, commercial and administrative matters would appear to be a potentially desirable benefit, albeit sometimes for very different reasons, both for those responsible for judicial public policy and for private legal professionals.

95. Whatever the legal tradition of the country, legal uncertainty, i.e. the risk of having one’s legal claim validated or rejected, prompts the desire to be able to quantify these factors using these new technological applications.

96. Lawyers (or a company’s legal department) see the possibility of using this technology to provide their clients with better informed advice by empirically and systematically assessing the chances of a procedure’s success, as well as encouraging the conclusion of transactions that, if necessary, make it possible to avoid a long and costly trial. Some insurers already offer the use of predictive systems to their clients to evaluate the merits of their business.⁴⁹

49. The Predictive software is offered to Allianz policyholders, for example.

97. At the same time, public decision-makers see this as an opportunity to better regulate the flow of new proceedings through the courts and provide themselves with a lever to reduce judicial operating costs. This is thought to encourage litigants to use alternative dispute resolution methods (conciliation, mediation or arbitration).⁵⁰

98. The approach that already exists in many judicial systems of harmonising decisions in many matters by using scales (divorce, dismissal, compensation for personal injury) could be revitalised through a probabilistic or actuarial approach.⁵¹ Alternative online dispute resolution services have even been created to help assess the amount of compensation for small disputes, *inter alia*. However, these interesting approaches are not unbiased and must not deprive citizens of access to a judge or call into question the adversarial principle.

Experiments conducted in France

At the initiative of the Ministry of Justice, the two courts of appeal in Rennes and Douai agreed to test predictive justice software on various litigation appeals in spring 2017, which in reality was an analysis of civil, social and commercial decisions of all French courts of appeal.

Although these internal and exhaustive case law data had already been available to them free of charge for many years (JURICA database), the Ministry made them specially available to the publishing company when they offered to assess the value of a quantified (innovative) analysis of the sums allocated by the two courts, in addition to a geographical classification of the discrepancies noted for similar applications and trials.

The stated objective of the software was therefore to create a decision-making tool in order to reduce, if necessary, excessive variability in court decisions, in the name of the principle of equality of citizens before the law. The result of the experiment, contradictorily debated between the two courts of appeal, the Ministry of Justice and the legal tech company who designed the product unfortunately stated the absence of added value of the tested version of the software for the work of reflection and decision-making of the magistrates.

More significantly, software reasoning biases were revealed that led to aberrant or inappropriate results due to confusion between mere lexical occurrences of judicial reasoning and the causalities that had been decisive in the judges' reasoning.

50. In this respect see information report No. 495 (2016-2017) prepared on behalf of the Senate Law Commission, and tabled on 4 April 2017, by Senator Philippe Bas

51. With respect to the actuarial approach, Case Law Analytics' offer announces more of a risk assessment than a prediction of a solution to a dispute.

6.1. A new computing tool to calculate scales

99. The procedures for calculating scales in various civil matters (for example, compensation for bodily injury, compensatory allowance and severance pay) appear to be considerably enhanced when combined with other processing techniques, subject to numerous design measures and uses (performative effect).⁵²

100. It is important to highlight what Jean-Paul Jean, chair of the CEPEJ evaluation working group, described as the qualitative challenge during a conference on open data held in 2016 in France: the procedure carried out by machine learning or any other method of processing should use certified originals, whose integrity has been verified, and which have been enriched to distinguish the important from the insignificant.⁵³

101. The other risk already encountered by the designers of these tools is that of “data-snooping”, namely selecting upstream-only data that is significant to predetermined analysis grids, for example by excluding from samples decisions that lend themselves poorly to correlations of linguistic sequences through machine learning or any other method (for example, decisions without presentation of the litigation or little reasoning).

102. But if we calculate a scale, are we not already doing a little prediction? The borderline may seem relatively sketchy unless we clearly distinguish the aim of the process: the objective here is not to deliver prescriptive information but to provide information about a state of affairs.

103. Subject to these methodological and operational precautions, the scales are available for handling certain disputes and are a powerful tool for harmonising case-law. Scales were previously calculated on more or less limited samples of decisions, but AI tools make it possible to survey more decisions and are liable to produce more accurate scales, applying a standard and hence giving more weight to the outcomes.

52. The performative or self-realisation effect is the risk that a system will produce the same output progressively by influencing the producers of input information; this effect is often mentioned with regard to judicial scales which, when informed by decisions based on these scales, tend to be representative only of themselves.

53. J-P. Jean, “Penser les finalités de la nécessaire ouverture des bases de données de jurisprudence”, Conference of 14 October 2016 held at the Court of Cassation, <https://www.courdecassation.fr/IMG///Open%20data,%20par%20Jean-Paul%20Jean.pdf>, page accessed on 17 March 2018.

6.2. Online dispute resolution

104. All European courts face to different degrees repetitive low-value civil litigation. The idea of facilitating the procedure through information technology and/or outsourcing them from the courts is quite widely shared. Great Britain, the Netherlands and Latvia are examples of countries that have already implemented or are about to implement these types of more or less automated solution⁵⁴s. For cross-border claims, the European Union has set up, by means of Regulation n°524/2013, a common framework available on the Internet (European small claims).

105. However, the scope of these online dispute resolution (ODR) services seems to have gradually extended. They have gone from restricted online services to alternative dispute resolution measures before the complaint is brought before the court, and are now being introduced

increasingly into the court process itself, to the point of offering electronic court services.⁵⁵ They do not only concern low-value disputes, but also tax disputes or disputes relating to social security services, or divorce proceedings.

106. For those who advocate such solutions, which are of interest to a number of legal professions and the private sector, access to justice could be significantly improved by a broad solution combining ODR and AI (or at least expert systems, see section 3 above for the distinction). The idea is to take complainants through an automated diagnosis of the dispute by putting a number of questions, which are then processed by the machine, resulting in proposals for a solution. The work of the Cyberjustice de Montréal laboratory, which assembles the various pre-litigation and litigation phases into a computer-based process for low-intensity disputes (for example, small claims courts in Quebec), is a good illustration of hybridisation.⁵⁶ According to the designers, there are clear benefits in terms of efficiency and quality.

54. See to this end the online dispute resolution available in the UK – <https://www.judiciary.gov.uk/wp-content/uploads/2015/02/Online-Dispute-Resolution-Final-Web-Version1.pdf>. See also the system PAs in the Netherlands, which issues automated decisions based on previously granted permits, and which has given rise to litigation at the national level and before the ECJ: there are two cases (c-293/17 en c-294/17) brought to the Council of State of the Netherlands (farmers/ nature conservation against the Netherlands) to determine whether or not a system (Programme Regulating Nitrogen) is allowed to decide whether or not farmers a.o. are violating the Habitat directive. Recently the Court of Justice of the European Union in Luxembourg has answered on the requests for a preliminary ruling concerning these joined cases (ECLI:EU:C:2018:882)

55. Darin Thompson, "Creating new pathways to justice using simple artificial intelligence and online dispute resolution", Osgoode Hall Law School of York University.

56. <http://www.cyberjustice.ca/projets/odr-plateforme-daide-au-reglement-en-ligne-de-litiges/>

107. But on what basis would any compensation proposed by such a system be calculated? What method? Does the algorithm process information fairly? Is the proposal intended to be discussed on an adversarial basis with the help of a trained and certified third party? Is access to a judge always possible? Some authors even see the widespread use of these dispute resolution methods as a new manifestation of digital “solutionism”, i.e. the systematic use of technologies to try to solve problems that do not necessarily fall within their scope.⁵⁷ It should also be noted that in Europe, a more protective regulatory framework that is binding on the member States has recently been put in place: Article 22 of the Data Safety Monitoring Plan explicitly provides for persons to be able to refuse to be the subject of a decision based exclusively on automated processing, with certain exceptions.⁵⁸

57. Evgeny Morozov, “Pour tout résoudre cliquez ici”, FYP editions, cited by David Larrousserie, “Contre le ‘solutionnisme’ numérique”, *Le Monde*, 6 October 2014, https://www.lemonde.fr/sciences/article/2014/10/06/contre-le-solutionnisme-numerique_4501225_1650684.html

58. Article 22(1) of EU Regulation 2016/679: “The data subject shall have the right not to be subject to a decision based solely on automated processing”; exceptions are provided for (such as the data subject’s consent) but “suitable measures to safeguard the data subject’s rights and freedoms and legitimate interests” must be implemented by the controller, including “the right of the data subject to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision”.

See, in the same sense, the Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, as amended by the Protocol adopted in May 2018 when the latter enters into force. Article 9(1)(a) provides the principle that “Everyone has the right not to be subject to a decision affecting him significantly, which shall be taken solely on the basis of automatic processing of data, without his point of view being taken into account. Notwithstanding this principle of prohibition, Article 9(2) states that “paragraph 1(a) shall not apply if the decision is authorised by a law to which the controller is subject and which also provides for appropriate measures to safeguard the rights, freedoms and legitimate interests of the data subject”.

The explanatory report states (§75): “It is essential that any person likely to be subject to a purely automated decision should have the right to challenge that decision by effectively putting forward his point of view and his arguments. In particular, the data subject must have the possibility to prove the possible inaccuracy of personal data before their use, the inadequacy of the profile to be applied to his particular situation or other factors which will have an impact on the outcome of the automated decision. This is particularly the case when the application of an algorithmic reasoning, by leading to the limitation of a right, the refusal of a social benefit or the evaluation of their borrowing capacity on the sole basis of the software, has the effect of stigmatizing individuals. However, the data subject may not exercise this right if the automated decision is provided for by the law to which the controller is subject, which provides for appropriate measures to safeguard the rights and freedoms and legitimate interests of the data subject”.

See also Article 9(1)(c) of the modernised Convention 108, which provides for the right of the data subject “to obtain, at his request, knowledge of the reasoning underlying the processing of data, when the results of such processing are applied to him”. The explanatory

108. The potential benefits of an ODR system, its degree of integration into a complete judicial process (from pre-litigation to actual litigation) and the almost decisive role of AI in the execution of the process must therefore be properly assessed on a case-by-case basis.

109. ODR already offers upstream knowledge of judicial processes. Its role is clearly to contribute to the implementation of conciliation, mediation and arbitration services outside the courtroom. These services can also be used during contentious proceedings under the supervision of judges before they decide on the outcome of disputes based on the merits (for some disputes this phase is considered compulsory).

110. On the other hand, the actual contribution of AI should be assessed. Is it only a question of using machine learning to establish indicative scales or prescribe a solution? Is it really AI that is being used or an expert system or just the logical rule chain? In any event, it should be possible to combine these systems with the requirements of transparency, neutrality and loyalty.⁵⁹

111. Finally, it is also necessary to examine the way in which complainants are encouraged to use the system: is there potential confusion in the very name of what is on offer? If one speaks of a court, it must be the form of organisation defined by the European Convention on Human Rights and not simply a private justice institution with the mere appearance of state justice⁶⁰. Is recourse to a judge clearly possible? In the Netherlands, private health insurance contracts seem to provide automatically for recourse to an ODR before any legal action is brought..

112. The CEPEJ Mediation Working Group (CEPEJ-GT-MED), launched in 2018, has offered its first thoughts on the contribution of information technology towards alternative dispute resolution methods. The CDCJ

(cont'd) dreport of the modernised Convention (§77) states: "Data subjects have the right to obtain knowledge of the reasoning underlying the data processing, including the consequences of this reasoning and the conclusions which may have been drawn from it, in particular when using algorithms for automated decision-making, in particular in the context of profiling. For example, in the case of a credit rating system, borrowers have the right to know the logic behind the processing of their data that leads to the decision to grant or refuse credit, rather than simply being informed.

of the decision itself. Understanding these elements contributes to the effective exercise of other essential safeguards such as the right of objection and the right of appeal to the competent authority". This "must obtain useful information concerning the underlying logic" is also found in the GDR (Article 13(1)(f); Article 14(2)(g); Article 15(1)(h)).

59. Charlotte Pavillon, "Concerns over a digital judge", nrc.nl, <https://www.nrc.nl/nieuws/2018/01/19/zorgen-om-populaire-digitale-rechter-a1588963>

60. Scarlett-May Ferrié, Algorithmes tested against fair trial, document consulted on <http://lexis360.fr>, downloaded on 09/07/2018, § 27-38

is currently doing in-depth work on ODRs to identify the potential of these tools but also any problem points amounting to possible violations of Articles 6, 8 and 13 of the European Convention on Human Rights.

6.3. The main guarantees to be reaffirmed in civil, commercial and administrative proceedings

Right of access to a court

113. The provision of online dispute resolution tools should not affect the right of access to a court within the meaning of Article 6, even if this right is not absolute and lends itself to implicit limitations.⁶¹ In civil matters, for example, every litigant has the right to submit to a court any dispute relating to his “civil rights and obligations” heard by a court.⁶² In 2015, the Parliamentary Assembly of the Council of Europe adopted a resolution on “Access to justice and the Internet: potential and challenges” in which it called to ensure that “parties engaging in ODR procedures retain the right to access a judicial appeal procedure satisfying the requirements of a fair trial pursuant to Article 6 of the Convention”.⁶³

Adversarial principle

114. It seems imperative to make a certain amount of quantitative information (for example, the number of decisions processed to obtain the scale) and qualitative information (origin of decisions, representativeness of selected samples, distribution of decisions between different criteria such as the economic and social context) accessible to citizens and, above all, to the parties to a trial in order to understand how scales have been constructed, to measure their possible limits and to be able to debate them before a judge.

Equality of arms

115. The use of technological means should not cause imbalances between parties, since the use of digital means could indeed facilitate proceedings

61. Art.6 §1 « 1. everyone is entitled to a fair and public hearing .. by an independent and impartial tribunal... who shall decide (...) the merits of any criminal charge against it.; for the limitations see *Deweert c. Belgique*, § 49 ; *Kart c. Turquie* [GC], § 67.

62. CEDH, *Golder c. Royaume-Uni*, §§ 28-36

63. Resolution 2054 (2015) of the Parliamentary Assembly of the Council of Europe (PACE), 10 November 2015, <http://assembly.coe.int/nw/xml/XRef/Xref-XML2HTML-EN.asp?fileid=22245&lang=en>

for certain operators (institutions, companies with means, computer literate persons) and, on the contrary, pose difficulties for certain population types that are more uncertain or less familiar with computers. It is important that no individuals are left alone in front of their screens, and that they are informed that they can seek legal advice and are assisted where necessary.

Impartiality and independence of judges

116. It has been posited that the norm derived from the majority trend referred to above in section 2.2. may have indirect effects on the independence and impartiality of the judiciary, particularly in systems where the independence of the judiciary is not fully achieved. In these systems, we cannot rule out the risk that such norms will place indirect pressure on judges when decisions are taken and prompt their approval, or that the executive will monitor those who depart from the norm.

Right to counsel

At the beginning of this chapter, we mentioned the advantages derived from the application of predictive justice tools for lawyers and, in particular, the possibility of providing their clients with better informed advice by empirically and systematically assessing the chances of a procedure's success. However, let us imagine a case where the chances of success for the litigant are extremely poor: could this affect the lawyer's decision to assist his client? Professional practice should aim to minimise the risk that persons requiring legal advice may ultimately be deprived of it.

7. Issues specific to criminal justice: prevention of offences, risk of recidivism and assessment of the level of danger

Even they are not specifically designed to be discriminatory, the use of statistics and AI in criminal proceedings has shown a risk of prompting the resurgence of deterministic doctrines to the detriment of doctrines of individualisation of the sanction, which have been widely acquired since 1945 in most European judicial systems.

117. The use of AI science and technology in criminal matters poses specific challenges as its application may reflect some current public debates about the alleged predictability of offending behaviour. However, this debate seemed to have been thoroughly settled for some thirty years in a number of European countries. In Italy for example, Article 220, paragraph 2, of the Code of Criminal Procedure expressly rules out the use of an expert opinion

to establish habitual or professional criminal features, the tendency to commit a crime, the character and personality of the accused and, in general, the psychological qualities of the accused, regardless of the pathological causes. In France, for example, the doctrine of “new social defence” developed by Marc Ancel was the foundation of criminal law: instead of a merely punitive and deterministic approach, a system of social rehabilitation was introduced to prevent the commission of an offence by avoiding the conditions for crime. This approach is shared by a number of European criminal policy instruments that focus on the objectives of re-educating and reintegrating offenders.⁶⁴

118. Criminal justice tools should therefore be designed in accordance with these fundamental principles of rehabilitation,⁶⁵ including the role of the judge in the individualisation of the sentence, based on objective elements of personalities (training, employment, regular medicals and social care) without any other form of analysis than that carried out by specifically trained professionals, such as probation officers. Big data analytics techniques could be used by these professionals to centralise and collect information on the person accused of a crime or misdemeanour, which could then be stored by various institutions and agencies and would then need to be examined by a judge, sometimes within a very short time frame (for example in the context of accelerated trial procedures).

7.1. Tools used by investigative authorities before the criminal trial

119. Instruments described as “predictive policing” (before the judicial process or before a court referral) are already growing rapidly and are beginning to be known by the general public (for example, think of the no fly list, which is actually a big data analytics application that collects and analyses data on potential terrorists in order to prevent the commission of acts, or algorithms used to detect fraud or money laundering).

120. In general, a large number of computer tools are commonly used to prevent the commission of criminal acts (by identifying possible places where this could happen or their authors) or prosecute them more effectively.⁶⁶ The

64. See European Court of Human Rights, Grand Chamber, *Vinter and Others vs. United Kingdom*, paras. 114-118

65. On the other hand, the use of AI for treatment and rehabilitation purposes (e.g. to collect data on treatment administered or methods of reintegration in prison) should be encouraged.

66. See Ales Zavrsnik, *Big Data, crime and social control*, page 194 et seq., which lists in detail a series of instruments used by police services in Europe and the United States.

first category includes “predictive policing” instruments that are used to prevent certain types of offences with elements of regularity in their occurrence such as burglary, street violence, theft from/of vehicles. The designation of these tools derives from their ability to determine precisely where and when these offences could be committed and to reproduce this information on a geographical map in the form of hot spots that are monitored in real time by police patrols. This process is called predictive criminal mapping. Most of the software used in this area is based on historical crime location evidence, such as police reports, but even more powerful new technologies combining various data and from different sources are also being tested.⁶⁷ These instruments, which have very persuasive rates of effectiveness, are also claimed to have deterrent effects on the commission of offences in areas surrounding hot spots, leading to a positive opinion of public policies.⁶⁸

121. However, the predictive capabilities of these tools, which show their limitations with regard to crimes of a less regular nature or targeting different locations, such as terrorism, must be put into perspective. In addition, one of their weaknesses is the effect of “vicious circles” and “self-fulfilling prophecies”: neighbourhoods considered at risk attract more police attention and police detect more crime, which leads to excessive police surveillance of the communities living in them.⁶⁹ Lastly, questions about a possible “tyranny of the algorithm” that could minimise or even progressively replace human judgment are not totally absent within the police services themselves, even

67. For example, as part of the project “E-Security – ICT for knowledge-based and predictive urban security” (<http://www.esecurity.trento.it/>), which was conducted in the Italian city of Trento between November 2012 and May 2015, a database gathering information on crimes reported to the police, the results of surveys conducted by the city hall on victimisation and real and perceived security by citizens, information on physical and social urban disorder from the police, as well as other variables relating to “SmartCity” (e.g. information relating to the socio-demographic context, urban setting, night lighting, the presence of surveillance cameras and public transport). It was created to better equip crime prevention and urban safety improvement work. The project managers testified to the reliability of the techniques used, which were said to make it possible to predict criminal acts with a success rate of approximately 60-65 % and which were said to help to better fight crime when limited resources were available. In addition, tests conducted in the United Kingdom as part of a pilot project to predict possible burglary, theft, and assault locations show that the software projections used, called PREDPOL, were accurate in 78 % of cases, compared to 51 % using traditional techniques.

68. Indication of a geographical concentration of the crime would help police forces to better consider the environmental factors making the crime more likely in the identified area (lighting, presence of shops, etc.) and to plan adequate responses in consultation with other partners.

69. “Predicting crime, LAPD style”, *The Guardian*, 25 June 2014.

if, for the time being, technology is presented as remaining at the service of human beings in order to better equip them for decision-making.⁷⁰

122. In addition, big data analytics are increasingly being applied in the prosecution of crime. Tools such as Connect, which is used by the UK police to analyse billions of data generated in financial transactions to find correlations or patterns of operations, or the International Child Sexual Exploitation Database (ICSE DB), managed by Interpol, which helps identify victims and/or perpetrators through the analysis, for example, of furniture and other objects in abusive images, or the analysis of background noise in videos, have proven particularly effective in fighting crime. With Connect, for example, searches that previously required months of investigation can now be performed in minutes, with a very high level of complexity and volume of data.

123. The doctrine nevertheless questions the managerial logic of the crime response provided by these predictive tools, in which a thorough analysis of the reasons for the crime becomes less important than doing something here and now. This is occurring at a time when available budgets are shrinking and the police must provide the same level of public protection, but with limited personnel, equipment and resources.⁷¹

7.2. Tools during the criminal trial

124. The use of predictive tools⁷² by judges in criminal trials is very rare in Europe.

125. HART (Harm Assessment Risk Tool) was developed in partnership with Cambridge University and is now being tested in the UK. This technology based on machine learning was trained using Durham Police archives dating from 2008 to 2012. By learning from decisions made by police officers during this period, and whether or not certain suspects reoffended, the machine is expected to be able to assess the risk – low, medium or high – of suspects reoffending, based on around thirty factors, some of which are not related to the crime committed (for example, postal code and gender).

70. "How technology is allowing police to predict where and when crime will happen", The Independent, 7 October 2017.

71. Ales Zavrsnik, *Big Data, crime and social control*, page 196.

72. In the literature, these tools are often referred to as "algorithmic justice" or "automated justice", or "simulated justice".

126. In tests initially conducted in 2013, during which suspect behaviour was observed over a two-year period after commission of the crime, HART predictions were found to be 98 % effective at predicting low risk and 88 % effective at high risk of recidivism. In this experimental phase, HART will have a purely advisory value for the judge. In addition, audits of HART's functioning and the reliability of its conclusions will be regularly conducted by the police.

127. Even if it is the only predictive tool identified in Europe to date, it provides the opportunity to consider the challenges that public decision-makers could face in the near future if this type of application is tested on a larger scale, particularly in the light of findings in America.

128. In the United States,⁷³ the NGO ProPublica revealed the discriminatory effects of the algorithm used in COMPAS software (Correctional Offender Management Profiling for Alternative Sanctions), which aims to evaluate the risk of recidivism when the judge must determine the sentence for an individual.

129. This algorithm, which was developed by a private company and which must be used by judges in certain American federal states, includes 137 questions answered by either the defendant or information pulled from criminal records. The questions are quite varied and include the presence of a telephone at home, difficulty paying bills, family history, criminal history of the accused, etc.⁷⁴ The algorithm rates the person on a scale from 1 (low risk) to 10 (high risk). It is an aid to judicial decision-making, its conclusions being only one of the variables considered by the judge when deciding on the sentence.

130. African-American populations were assigned a recidivism high-risk rate twice that of other populations within two years of sentencing – without this effect being naturally sought by the designers⁷⁵. In contrast, the algorithm considered that other populations appeared much less likely to repeat an

73. A 2015 study had identified about sixty predictive tools in the United States.

74. There are other algorithms that have been developed using critical observations expressed by the doctrine (see next chapter) that are based on smaller variables, more directly related to the crime committed and less related to race, gender or socio-economic status. An example is the Public Safety Assessment Tool used in 30 American jurisdictions.

75. This purely discriminatory effect can in fact be explained by the relatively permissive "calibration" of the algorithm model, which creates many "false positives".

offence⁷⁶. It must of course be noted that this type of misleading interpretation in reality reveals only the social and economic fragility of certain groups of populations which are obviously not criminogenic by nature. Researchers at Dartmouth College have also shown that this type of algorithm does not produce added value since people without any criminal history can reproduce exactly the same assessment simply by answering the questionnaire.

131. In addition, the lack of transparency in the algorithm operation processes designed by private companies (which claim intellectual property) was another cause for concern. If we take into account the fact that they take their source data from the state authorities themselves, their lack of accountability to citizens poses a major democratic problem. Accounts have shown that the public is informed about big data operations accidentally, sporadically and when there are leaks or errors: an example of this is when ProPublica revealed the flaws in the COMPAS algorithm following the owner company's refusal to share it. The NGO had to appeal to public authorities to access the data and hire its own scientist to examine the algorithm.

7.3. The challenges of “prediction” in criminal matters

132. We have seen in the previous sections that the degree of development of predictive tools in Europe in the criminal field is very varied. Although instruments described as “predictive policing” are growing rapidly and are even beginning to come to the attention of the general public, the situation is not the same when it comes to the application of this type of tool by

76. Black populations were more frequently classified as high risk, while they did not reoffend within two years of conviction; white populations, on the other hand, who were more frequently classified as low risk, committed offences within the next two years. In short, the algorithm overestimated the risk of recidivism for blacks and underestimated it for whites (the “false positives” were mostly blacks, while the “false negatives” were mostly whites). In response to ProPublica's allegations, NorthPointe (now Equivant since this controversy) replied that white and black populations were equally represented when considering the “true positives”, i.e. those who had actually reoffended. **The question of how to reconcile both the accuracy of the algorithm in detecting recidivism and the need to avoid discriminatory effects towards black populations has been the source of intense debate in the literature;** see in particular Chouldechova A (2016), “A fair prediction with a disparate impact: a study on bias in recidivism prediction instruments”, available at <http://arxiv.org/abs/1610.07524>; and also “Bias in criminal risks scores is mathematically inevitable, Researchers say”, available at <https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say>. **This debate also reflects the question of the legitimacy of a private company, without any institutional control, to arbitrate between two opposing requirements: that of defending society, on the one hand, and that of respecting the rights of individuals, on the other.**

judges in criminal trials. As regards the instruments available to prosecution services, thoughts about their advantages and disadvantages have already been expressed. Let us now study the tools specific to criminal trials.

133. First of all, it is important to rule out arguments based solely on the efficiency or inefficiency of these tools. The examples given above show that there can be tremendous opportunities but also real risks in the application of new technologies that are used without the necessary precautions. Public decision-makers and judicial stakeholders must be particularly vigilant and play an active role in the development of these technologies; continuous monitoring is necessary to determine their real effectiveness and efficiency, and to avoid unforeseen consequences. This is even more important in criminal proceedings because of their direct impact on the individuals' personal freedoms.⁷⁷

134. This implies that both the benefits and the drawbacks of the application of such tools in the judicial field should be carefully measured.

135. Supporters often argue that they are neutral and that they rely on factual and objective methods that help to make justice more accurate and transparent. Another great asset, it is claimed, is their efficiency, which sometimes exceeds human capacities and can only be extremely valuable in a general context of reduced public funds or even a shortage of resources.

136. The inclusion of algorithmic variables such as criminal history and family background means that the past behaviour of a certain group may decide the fate of an individual, who is, of course, a unique human being with a specific social background, education, skills, degree of guilt and distinctive motivations for committing a crime⁷⁸. They also argue that human decisions can be based on values and considerations (e.g. societal) that would not be retained by the machine. For example, a judge could decide to order the bail of a female offender who has a risk of recidivism, on the basis of a hierarchy of values, for example by setting greater store by her role as a mother and protector of her children, whereas the algorithm would be able to determine the risk of reoffending more accurately but would not be able to operate such a hierarchy of priorities.

77. An extract from the Wisconsin Supreme Court decision in *Wisconsin v. Loomis* can also provide inspiration at European level: "It is important to consider that tools such as COMPAS continue to change and evolve. The concerns we address today may very well be alleviated in the future. It is incumbent upon the criminal justice system to recognize that in the coming months and years, additional research data will become available. Different and better tools may be developed. As data changes, our use of evidence-based tools will have to change as well. The justice system must keep up with the research and continually assess the use of these tools."

78. Aleš Zavrsnik, "Big Data, crime and social control", page 196.

137. In criminal matters, there are also potential risks of discrimination when one considers that these tools, which are constructed and interpreted by humans, can reproduce unjustified and already existing inequalities in the criminal justice system concerned; instead of correcting certain problematic policies, technology may end up legitimising them. As already mentioned, the NGO ProPublica⁷⁹ clearly revealed the discriminatory effects of the algorithm used in COMPAS,⁸⁰ which predicted that black populations were twice as likely to reoffend as white populations within two years of sentencing while considering that white populations were much less likely to repeat the offence. They could, however, help reveal the errors in decision-making so that they may be corrected.⁸¹ Moreover, the lack of transparency in algorithm construction processes by proprietary companies and their accountability to the public is a cause for concern, all the more so if they are part of steps taken by state authorities to make data available to the public.

138. In the light of the foregoing, when algorithms are used in the context of a criminal trial it seems essential to fully guarantee respect for the principle of equality of arms and presumption of innocence established by Article 6 of the ECHR. The party concerned should have access to and be able to challenge the scientific validity of an algorithm, the weighting given to its various elements and any erroneous conclusions it comes to whenever a judge suggests that he/she might use it before making his/her decision. Moreover, this right of access is also covered by the fundamental principle of personal data protection. All people have the right not to be subject to decisions affecting them significantly made solely on the basis of automated data processing, without their point of view having been taken into account beforehand.

139. In this respect, there is a difference between Europe and the United States with regard to the right of access to algorithms: while in the United States the judicial authorities are still reluctant to recognise this right fully and weigh private interests (particularly the protection of intellectual property) against the rights of defence, in Europe the framework is more protective because of the GDPR, which establishes a right to information on the underlying logic of decisions made using algorithms.⁸²

79. www.propublica.org/article/technical-response-to-northpointe.

80. Other algorithms focus on other elements more directly related to the offence committed.

81. Mojca M. Plesnicar and Katja Sugman Stubbs, "Subjectivity, algorithms and the courtroom".

82. Article 15, 1. (h) of EU Regulation 2016/679: "The data subject shall have the right to obtain from the controller"... "the following information:"..." the existence of automated decision-making, including profiling, as referred to in Article 22, paragraphs 1 and 4, and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject".

140. The considerations expressed earlier regarding the potentially negative effects of these tools on the impartiality of the judge are also valid in criminal matters: a judge who decides against the prediction of an algorithm is likely to take risks as he assumes greater responsibility. It does not seem unrealistic to imagine that judges would be reluctant to take on this additional burden, particularly in systems where their terms of office are not permanent but subject to popular vote,⁸³ or in which their personal liability (disciplinary, civil or even criminal) is likely to be incurred, especially if their statutory guarantees in disciplinary matters are insufficient.

8. Specific questions relating to the protection of personal data

The use of algorithms raises the question of the protection of personal data when being processed. The precautionary principle should be applied to risk assessment policies.

141. For the full potential of algorithms to be exploited while complying with data protection principles, the precautionary principle should be applied and preventive policies should be put in place to counter the potential risks associated with the use of the data processed by these algorithms and the impact of their use on individuals and society in general.

142. The principle of lawfulness in the processing of personal data and the obligation to prevent or minimise the impact of data processing on the rights and fundamental freedoms of data subjects should induce prior risk assessment. This should make it possible to implement the appropriate measures, particularly during the design stage (and hence by design) and by default, in order to mitigate the risks identified.

143. Since personal data must be processed for specified and legitimate purposes, they must not be used in a way that is incompatible with those purposes and must not be further processed in a way that the data subject may consider unexpected, inappropriate or questionable (principle of loyalty). The issue of re-using personal data, making them widely accessible, therefore needs to be handled with the utmost caution.

144. The design of the data processing methods used by algorithms should minimise the presence of redundant or marginal data and avoid any potential

83. Mojca M. Plesnicar and Katja Sugman Stubbs, "Subjectivity, algorithms and the courtroom".

hidden bias and any risk of discrimination or negative impact on the fundamental rights and freedoms of the data subjects.

145. When artificial intelligence is used, the rights of data subjects are of especial importance, and the control that each of us must have over our personal information implies that it must be possible to exercise the following rights: the right of data subjects not to be subject to automated decisions significantly affecting them without their point of view being taken into account, the right to obtain information about the reasoning underlying the data processing carried out by algorithms, the right to oppose such processing and the right to a legal remedy.

9. The potential and limitations of predictive justice tools

The term predictive justice should be dismissed because it is ambiguous and misleading. These tools are based on methods of analysis of case-law, using statistical methods that do not in any way reproduce legal reasoning but may attempt to describe it. Analytical biases, if they cannot be totally eliminated, must be identified. The design process and the use of the tool must be embedded in a clear ethical framework.

146. In section 3, we already highlighted the ambiguity and fallacy of the concept of predictive justice and how it operates a slow shift in the collective mind, leading us to believe that machines, devoid of any emotion, will one day be better able to make the act of judging more reliable. More than ever, its promises need to be examined in an objective and scientific manner, based on solid foundations of fundamental research, in order to identify possible limitations. In this connection, it should be noted that the risks of distorted interpretations of the meaning of court decisions are extremely high when based on statistical modelling alone. This observation is borne out further by the lack of precise understanding of the links between the data and the obvious presence of false correlations that cannot be discerned in large masses of data.

147. Moreover, the neutrality of algorithms is a myth, as their creators consciously or unintentionally transfer their own value systems into them. The philosopher, Eric Sadin, noted that, behind their efficient and impersonal facade, algorithmic systems imperceptibly reflect the intentions of their designers or sponsors, inducing a functioning and asymmetric power over the lives of other people. Similarly, researcher Aurélien Grosdidier considers that an algorithm, in itself, is capable of nothing other than allowing us – at best – to grasp part of the designer’s intention and extends the questioning

to the entire information processing chain (designer's intention, production of computer code, execution of computer code and context of execution then maintenance). This observation is also shared by the criminologist Aleš Zavrsnik, who underlines how the stages of construction and interpretation of algorithms are carried out by man, for man, and cannot escape errors, prejudices, values, human interests and a human representation of the world, however they are conceived.

148. Despite these significant limitations, should we overlook the contribution of a technology with unequalled power? Mathematicians C. S. Calude and G. Longo themselves stress in their study on big data that the restrictive or negative scope of their results, as often happens, does not destroy the science of data but paves the way for greater thought, including the challenge of a new, more extensive scientific method capable of incorporating both new algorithmic instruments and classical tools by accompanying processing with a rigorous evaluation of evidence. As highlighted in the introduction, the use of AI is likely to offer extremely significant support for professionals, including judges and lawyers, but also the general public, especially if one day they make it possible to construct unparalleled research and documentary analytical tools in legislative, regulatory, jurisprudential and doctrinal matters and create dynamic links between all these sources. But this type of application goes beyond the scope of this article since it is not designed to predict the outcome of a dispute but to analyse case-law in a given time and space.

149. As discussed in section 6, subject to the representativeness of selected and processed samples, AI has helped to devise much more precise scales of the average or median amounts of money allocated, *mutatis mutandis*, in various fields (financial support, compensatory benefits, compensation for bodily injury, severance pay, etc.). These scales, which are based more on a consensus than on an average analysis of what already exists, already provide significant support for decision-making and guidance without being able to replace the law itself. As mentioned earlier, the risk is that, in the absence of a statistical representation of reality or of being able to predict anything, the results of predictive justice software will be set as standards without any validation by the legal system and in conflict with it.

150. Lastly, let us consider the idea of being able to step back at will from predictive systems. Rather than locking users into a probability (or set of probabilities), the idea would be to allow them to navigate through the correlations that led the system to propose its assessment and to be able to distance themselves by selecting other more relevant concepts or groups of words or to exclude false correlations. To use the UCL example, this would

consist of proposing a graphic representation of the different terms retained by the system (with their respective weightings) to find that there has been a violation (or non-violation) and to authorise other paths to be taken by proposing the selection of other terms or lexical groups.

151. However bold and seductive this proposal may be, it presupposes that the professionals themselves (judges, lawyers, universities) take it over collectively to test its feasibility and that they do not allow private operators alone, save for a few unchecked scientists, to design software and abstruse or locked modes of reasoning or calculation.

152. The ambitious (and unfulfilled) promises of some legal tech companies must not hide the immense potential of technologies and the need for applications adapted and built in directly with scientific and academic research environments, as well as with all legal professionals, such as magistrates, clerks, lawyers, notaries, bailiffs and field experts. A number of measures would appear to be able to take full advantage of these new tools through applications adapted and devised in direct association with research workers and all legal professionals, including judges, prosecutors, clerks, lawyers, notaries, bailiffs and experts in the field.

153. In this dynamic context, it seems essential, firstly, not to make hasty decisions and to take time to debate in advance the risks and practical applications of these instruments for judicial systems, and to test them in the first stage. A judicial system in keeping with its time would be one that is capable of establishing, administering and guaranteeing genuine **cyberethics** for both the public and private sectors, and insisting on total transparency and fairness in the functioning of algorithms, which may contribute one day to judicial decision-making.

10. The need for an in-depth public debate on these tools prior to the implementation of public policies for their development. The urgent need for cyberethics to provide a framework for the development of artificial intelligence algorithms while respecting fundamental rights

The challenge of integrating these tools into a judicial decision-making process justifies simplifying the concepts for the target group concerned. An ethical framework must be set up to promote the rapid development of a form of AI that includes mechanisms preventing bias and discrimination in its very design processes.

10.1. The importance of debating, testing and continually reviewing the application of these tools prior to the implementation of public policies

154. The issues connected with the implementation of predictive justice tools are so numerous and multifaceted that they call for a balanced approach by public decision-makers.

155. Firstly, it is essential to hold a public debate around these questions, bringing together both the designers of the tools and legal professionals. Judicial councils, professional associations of judges and bar associations can undoubtedly contribute to this and help to identify opportunities and more controversial aspects. In addition, judicial training and law schools can play a key role in raising awareness among justice professionals on these issues, so that they can better understand and practically contribute to current developments.

156. It is also essential to carry out research on the proposed applications and test them, both to understand their potential and their weaknesses and to be able to develop them further and adapt them to our needs. A right to examine the components and characteristics of the instruments proposed by the private sector (or those developed by independent and specialised public institutes, a solution which should be encouraged) seems equally important so that the justice service can effectively carry out its mission. A rigorous evaluation of test results should be conducted before wider deployment and integration into public policy. It also seems strongly advisable to regularly assess the impact of these tools on the work of justice professionals.

10.2. The establishment of an ethical framework

157. First of all, merely adopting a legislative or regulatory framework for AI seems vain in a digital context, which is inherently transnational in its scope. On the other hand, scrupulous attention to the nature and quality of open data is likely to minimise the risks of inadequate cross-referencing and to reinforce the relevance of the results of automated processing. With regard to the names of professionals, a simple precaution would be to prohibit their public dissemination in structured raw databases given the risks of misuse. It would not be a question of limiting access to the information already processed (for example, the composition of a panel of judges) but of filtering the raw data made freely available. In short, a distinction needs to be made between access to information and access to databases, which can be manipulated at will.

158. Researchers Buttarelli and Marr stressed how big data need to be closely controlled and protected. Other researchers (Pasquale and Morozov) have stressed the need to establish transparent procedures for the deployment of big data and more generally AI in the judicial field, because the solutions proposed can never represent life in its complexity.

159. The development of cyberethics rules to guide the activity of stakeholders in the sector and to promote the above-mentioned principles of transparency, fairness and neutrality of the tool is essential. Regular monitoring by independent experts should ensure that the drivers of artificial intelligence used to assist judges in their decisions are not biased. It is not inappropriate to anticipate the implementation, discreet or not, of paid referencing systems (based on Google's Search Engine Advertising model) allowing certain operators to give less weight to decisions that are unfavourable to them. These rules will play a key role in increasing citizens' confidence in their judicial systems.

160. In this connection, the quality of the best systems could be recognised by the award of a label or certification. In particular, they must guarantee total transparency and perfect fairness in the way information is processed, both for professionals and for citizens, to prevent the repetition of errors such as the aforementioned COMPAS algorithm. Justice professionals must be closely involved to be able to properly assess the risks and the impact of these applications on judicial systems.

161. Nowadays, all the experts involved in the development of AI, including researchers, engineers and computer developers, have quite exceptional and unprecedented responsibilities. Their work could be accompanied by an even greater strengthening of the humanities. The example of some innovative schools of computer developers shows that behind the will to "hack the system" hides, according to some observers, in reality a pragmatism devoid of any contextualisation of the responsibility that now hangs over technicians with quasi-demiurge powers. The Hippocratic Oath certainly has its limits in medicine but ritualises responsibility and provides an ethical framework.

162. Finally, cyberethics must be accompanied by large-scale training of stakeholders, from algorithm designers and legal tech companies to their users. New transdisciplinary humanities should be made available to all so that AI becomes a vector of positive development for humankind.

Appendix II

Which uses of AI in European judicial systems?

This Annex to the Charter reviews different uses of AI in European systems and encourages to a different degree their application in the light of the principles and values set out in the Ethical Charter.

The use of machine learning to constitute search engines for case-law enhancement is an opportunity to be taken up for all legal professionals. Additional applications (drawing up of scales, support for alternative dispute settlement measures, etc.) should be considered, but due care must be taken (in particular, the quality of the data source and not mass processing of the entire dispute in question). Other applications (“predictive justice”) should be assigned to the field of research and further development (in consultation with legal professionals in order to ensure that they fully tie in with actual needs) before contemplating use on a significant scale in the public sphere.

In criminal matters, this is a very sensitive issue but it should not be ignored. In the light of the many existing questions as to their compatibility with a certain number of fundamental rights, the use of algorithms to calculate the potential risks of recidivism of an individual brought to justice should be considered with the most extreme reservations. On the other hand, the processing of global quantitative data for crime prevention is an avenue to be further explored with these new techniques, taking into consideration known biases (performative effects, data quality, etc.). Similarly, the use of algorithms to form a better link between the type of community service available and an individual's personality may be a factor in the effectiveness of a measure of this kind.

Uses to be encouraged

- ▶ **Case-law enhancement:** machine Learning techniques have been increasingly deployed in the field of natural language processing in the past years (this includes initial efforts in natural language understanding) and are a considerable asset for finding search options to complement current keyword or full-text search. These tools could link various sources (e.g. constitutions and conventions, laws, case law and legal theory). Data visualisation techniques could illustrate search results.
- ▶ **Access to law:** without replacing human intervention, chatbots could be set up to facilitate access to the various existing sources of information using natural language. Document templates (court applications, lease agreements, etc.) could also be generated online.
- ▶ **Creation of new strategic tools:** the use of data science and artificial intelligence techniques on court activity data can help improve the efficiency of justice by making it possible, for example, to carry out quantitative and qualitative evaluations and to make projections (e.g. future human and budgetary resources). Key performance indicators could be drawn up on this basis. It is recommended that legal professionals, especially judges, be involved in the implementation of these tools, in terms of taking ownership of these tools and of analysing the results in conjunction with factors relating to the specific features of the court in question or the quality of justice (for example, the need to preserve access to justice).

Possible uses, requiring considerable methodological precautions

- ▶ **Help in the drawing up of scales in certain civil disputes:** an analysis of all judicial decisions is not statistically meaningful if all the causative factors (explicit and implicit in the decisions) are not identified. Knowing that the average compensation awarded in a certain geographical area is higher than in another can be explained not due to the behaviour of judges, but in the light of the characteristics of the area in question. Machine learning can therefore be useful in identifying decisions (see *Case-law enhancement* above), but the automated processing of data alone cannot produce meaningful information. An essential prerequisite is the compilation of a relevant sample of decisions to be processed (for example, via surveys).

- ▶ **Support for alternative dispute settlement measures in civil matters:** in some European countries, “predictive justice” tools are used by insurance companies to evaluate the chances of success of a dispute and to steer the litigant towards another method of dispute resolution when it is felt that there is little chance of success. Furthermore, some systems abroad offer compensation amounts without any real transparency as to the rules of calculation. However, these systems cannot be considered as impartial and reliable (see the section on machine learning techniques). Decisions are made about a citizen using truncated bases. In other cases, a litigant may be advised, by means of a virtual agent (chatbot), to opt for an alternative dispute settlement measure after a preliminary examination of the criteria entered by the litigant himself or herself, when visiting a court’s website or searching legal information online. The virtual agent may, where appropriate, also recommend that the litigant seek advice from a mediation service or a lawyer. In all these cases, the presence of a trained third party (mediator using not only techniques but maybe scales as calculated above, or a lawyer) would appear to be the most appropriate solution at this stage.
- ▶ **Online dispute resolution:** when litigants go onto an online dispute resolution platform, they should be informed in a clear and comprehensible manner whether the processing of their dispute is done in an entirely automated way or with the involvement of a mediator or arbitrator. In addition, the information given to litigants must be honest and must avoid giving them the impression that a court is involved (in this connection, the term “online court” is often used for this type of platform, whereas technically their purpose is to provide alternative dispute resolution services). These are two essential factors in enabling litigants to make an informed choice, possibly disagreeing with the advice and deciding to go to a real court within the meaning of Article 6 of the ECHR. Furthermore, in view of the requirements of Articles 6 and 13 of the ECHR, forms of review of the online dispute resolution procedure and its outcome by state courts should always be considered, especially where the litigant has consented to fully automated online dispute resolution.
- ▶ **The use of algorithms in criminal investigation in order to identify where criminal offences are being committed:** this type of application could concern not only the police but also prosecutors in the crime prevention bodies of which they are a part. Systems have been used in the United States to guide police patrols in real time towards possible locations where offences are being committed. However, this type of

quantitative approach can generate a strong “performative effect” (in a given location, there is a greater chance of discovering an offence and this then reinforces the system). Criminal analysis through approaches combining geographic information systems (GISs) and large amounts of data on procedures could be better shared with prosecutors and could certainly benefit from a significant machine learning contribution. Anti-money laundering units already use “predictive” systems to identify suspicious financial flows, but in the case of quantitative (financial) information, machines are more able to produce reliable results. Researchers should also have better access to these data to produce relevant studies for policy-makers.

Uses to be considered following additional scientific studies

- ▶ **Judge profiling:** quantifying a judge’s activity will reveal less about any possible biases than about any external factors influencing his or her decisions. The judge himself or herself is not the reason why judicial activity in an impoverished area does not produce the same results as in another territory, whatever his personality may be. When the decision is rendered in a collegial manner and without the possibility for a judge to express a divergent opinion, it is pointless to profile each of the judges of the chamber. On the other hand, offering judges a more detailed quantitative and qualitative assessment of their activities, thanks to new tools, but with a purely informative aim of assisting in decision-making and for their exclusive use, could be encouraged.
- ▶ **Anticipating court decisions:** statistical processing of lexical groups alone reveals the frequency of the use of certain groups of words but does not identify the real reasons for a decision and does not carry out a legal analysis (see the study carried out on ECHR decisions by the University College of London which produced better results on the facts than on the analysis of the law). Hybrid systems, founded on the construction of mathematical models that are supposed to represent the diverse range of judges’ reasoning, are not any more efficient because they are still limited by bias in the data sample that they have processed and have to start again from square one if a law is amended or if there is a reversal in case law.

Uses to be considered with the most extreme reservations

- ▶ **Use of algorithms in criminal matters in order to profile individuals:** experiments in other countries (COMPAS in the United States

and HART in the United Kingdom) have been criticised by NGOs (see work by ProPublica in the United States and Big Brother Watch in the UK). Because of the limitations of the methodology used, this pure statistical approach has led to wrong result: the finding that some African-American individuals are more often involved in criminal acts has led to a higher risk factor for the entire African-American population. Thus, even for minor offences, these systems have negatively weighted African-American defendants, with the result of unfairly increasing the quantum of their sentences. This approach, which has discriminatory and deterministic effects, must be replaced by one that is more respectful of European standards concerning criminal sanctions and which must offer the individual the possibility of rehabilitation and reintegration. If algorithmic systems manage to help improve the collation of information for probation services, for example, and make it possible for the relevant information to be collected more quickly for subsequent human processing, then progress would definitely be made (particularly in expedited proceedings). Any other use is prone to biases which will come into conflict with certain national and supra-national fundamental principles.

- **Quantity-based norm:** it is not only a question of producing scales, which could be legitimate, but of providing each judge with the content of the decisions produced by all the other judges and claiming to lock his future choice into the mass of these “precedents”. This approach should be rejected because this large number cannot add to or act in place of the law. For the reasons provided above (*Help in the drawing up of scales*), a quantity-based approach is not the way to go. The CEPEJ study also highlighted the dangers of the crystallisation of case law and the potentially negative effects on the impartiality and independence of judges.

Appendix III

Glossary

This glossary provides a definition of the terms used by the Ethical Charter and study document. Preference has been given to providing a narrow definition for all the vocabulary used. All the documents must be read and understood in the light of these definitions.

A

ALGORITHM Finite sequence of formal rules (logical operations and instructions) making it possible to obtain a result from the initial input of information. This sequence may be part of an automated execution process and draw on models designed through machine learning.

ANONYMISATION Method to process personal data in order to completely and irreversibly prevent the identification of a natural or legal person. Anonymisation therefore implies that there is no longer any possible link between the information concerned and the person to whom it relates. Identification then becomes completely impossible.⁸⁴ As the principles relating to data protection apply to all information relating to an identified or identifiable individual, they do not apply to anonymised data.

ARTIFICIAL INTELLIGENCE (AI) A set of scientific methods, theories and techniques whose aim is to reproduce, by a machine, the cognitive abilities of human beings. Current developments seek to have machines perform complex tasks previously carried out by humans.

84. Working Party on Article 29 Opinion 05/2014 on Anonymisation Techniques. See also No. 26 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016.

However, the term artificial intelligence is criticised by experts who distinguish between “strong” AIs (yet able to contextualise specialised and varied problems in a completely autonomous manner) and “weak” or “moderate” AIs (high performance in their field of training). Some experts argue that “strong” AIs would require significant advances in basic research, and not just simple improvements in the performance of existing systems, to be able to model the world as a whole.

The tools identified in this document are developed using machine learning methods, i. e. “weak” AIs.

B

BIG DATA (*metadata, large data sets*) The term big data refers to large sets of data from mixed sources (e.g. open data, proprietary data and commercially purchased data). For data derived from judicial activity, big data could be the combination of statistical data, records of business software connections (application logs), court decisions’ databases, etc.

C

CHATBOT (*conversational agent*) Conversational agent which converses with its user (for example, empathy robots used to help those who are ill, or automated conversation services in customer relations).⁸⁵

D

DATA Representation of information for automatic processing. When it is said that algorithms can be “applied” to the most diverse realities in the legal world or elsewhere, one presupposes the “digitalizability” of any reality in the form of “data”. But it is clear from physics that nothing tells us that physical processes can be adequately translated in terms of “data” (and integrated into the input/output cycle of algorithms). If this is already the case in physics, there is no reason why this should not also be the case in social relations. We must therefore be cautious with the idea of “data”, which always assumes that the reality we are trying to describe has a format such that it is naturally algorithmically processable.

85. CNIL Report December 2017: How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence.

DATABASE A database is a “container” that stores data such as numbers, dates or words, which can be reprocessed using a computer to produce information, for example, gathering and sorting numbers and names to form a directory.

DATA MINING Data mining makes it possible to analyse a large volume of data and highlight models, correlations and trends.

DATA SCIENCE A large field that groups together mathematics, statistics, probabilities, data processing and data visualisation in order to gain understanding from a mixed set of data (images, sound, text, genome data, links between social networks, physical measurements, etc.).

Methods and tools derived from artificial intelligence fall within this category.

DEEP LEARNING See *Machine Learning* and *Neurons*

E

EXPERT SYSTEM This is one of the ways to achieve artificial intelligence. An expert system is a tool capable of reproducing the cognitive mechanisms of an expert in a particular field. More precisely, it is software capable of answering questions, by reasoning based on known facts and rules. It consists of 3 parts:

- ▶ a fact base;
- ▶ a rule base;
- ▶ an inference engine.

The inference engine is able to use facts and rules to produce new facts, until it reaches the answer to the expert question asked.

Most existing expert systems are based on formal logic mechanisms (Aristotelian logic) and use deductive reasoning.

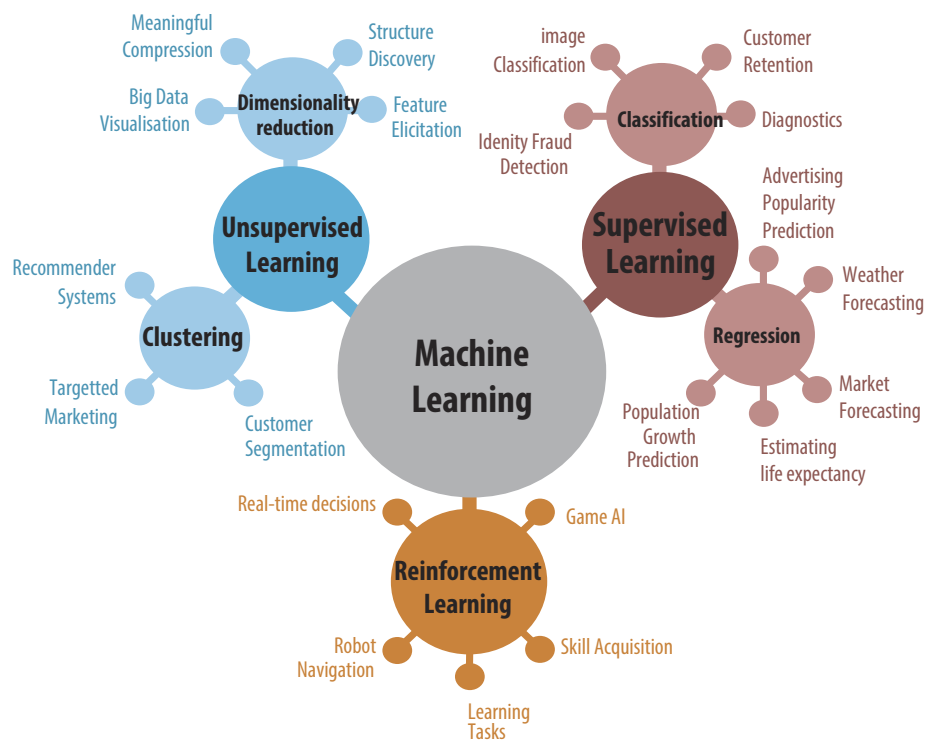
L

LEGAL TECH Companies using information technology in the field of law in order to offer innovative legal services. These companies are start-ups specialised in law. Other terms derived from business sectors have also appeared such as *Fintechs* for start-ups offering financial services and *Medtechs* in the medical field.

M

MACHINE LEARNING Machine learning makes it possible to construct a mathematical model from data, incorporating a large number of variables that are not known in advance. The parameters are configured gradually during the learning phase, which uses training data sets to find and classify links. The different methods of machine learning are chosen by the designers depending on the nature of the tasks to be completed (grouping). These methods are usually classified into three categories: (human) supervised learning, unsupervised learning and reinforcement learning. These three categories group together different methods including neural networks, deep learning, etc.

The graph below illustrates the different categories of machine learning:



METADATA Data that make it possible to define, contextualise or describe other data. In most of its computer uses, the meta prefix means “reference definition or description”.

Metadata synthesize basic information about data, they facilitate the search and manipulation of particular data instances. The author, creation date, modification date and file size are examples. Metadata and its corollary, data filtering, help to locate a specific document.

N

NEURONS/NEURAL NETWORK Neural networks are computing systems vaguely inspired by the biological neural networks that constitute animal brains.[1] Such systems “learn” to perform tasks by considering examples, generally without being programmed with any task-specific rules. For example, in image recognition, they might learn to identify images that contain cats by analyzing example images that have been manually labeled as “cat” or “no cat” and using the results to identify cats in other images. They do this without any prior knowledge about cats, e.g., that they have fur, tails, whiskers and cat-like faces. Instead, they automatically generate identifying characteristics from the learning material that they process.

An ANN is based on a collection of connected units or nodes called artificial neurons which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it.

The original goal of the ANN approach was to solve problems in the same way that a human brain would. However, over time, attention moved to performing specific tasks, leading to deviations from biology. Artificial neural networks have been used on a variety of tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis.

O

OPEN DATA The term refers to making structured databases available for public download. These data can be inexpensively re-used subject to the terms of a specific licence, which can, in particular, stipulate or prohibit certain purposes of re-use.

Open data should not be confused with unitary public information available on websites, where the entire database cannot be downloaded (for example, a database of court decisions). Open data do not replace the mandatory publication of specific administrative or judicial decisions or measures already laid down by certain laws or regulations.

Lastly, there is sometimes confusion between data (strictly speaking open data) and their processing methods (machine learning, data science) for different purposes (search engines, assistance in drafting documents, analysis of trends of decisions, predicting court decisions, etc.).

OPEN SOURCE SOFTWARE Software for which the source code is available to everyone. The software can therefore be freely used, modified and redistributed.

P

PERSONAL DATA Any information concerning an identified or identifiable natural person (the “person concerned”), directly or indirectly.

These include sensitive data relating to genetic data, biometric data uniquely identifying an individual, data relating to offences, criminal proceedings and convictions and related security measures, and any data for information they reveal on racial or ethnic origin, political opinions, trade union membership, religious or other beliefs, health or sex life.

PREDICTIVE JUSTICE Predictive justice is the analysis of large amounts of judicial decisions by artificial intelligence technologies in order to make predictions for the outcome of certain types of specialised disputes (for example, redundancy payments or alimentary pensions).

The term “predictive” used by legal tech companies comes from the branches of science (principally statistics) that make it possible to predict future results through inductive analysis. Judicial decisions are processed with a view to detecting correlations between input data (criteria set out in legislation, the facts of the case and the reasoning) and output data (formal judgment such as the compensation amount).

Correlations deemed to be relevant make it possible to create models which, when used with new input data (new facts or precisions described as a parameter, such as the duration of the contractual relationship), produce according to their developers a prediction of the decision (for example, the compensation range).

Some authors have criticised both the form and substance of this approach. They argue that, in general, the mathematical modelling of certain social phenomena is not a task comparable to other more easily quantifiable activities (isolating the really causative factors of a court decision is infinitely more complex than playing the game of Go or recognising an image for example):

here, there is a much higher risk of false correlations. In addition, in legal theory, two contradictory decisions can prove to be valid if the legal reasoning is sound. Consequently, making predictions would be a purely informative exercise without any prescriptive claim.

PROFILING An automated data processing technique that consists of applying a “profile” to a natural person, in particular in order to make decisions about him or her or to analyze or predict personal preferences, behaviours and attitudes.

PROCESSING OF PERSONAL DATA According to Article 2 of the revised Convention 108, “data processing” means any operation or set of operations performed on personal data, such as the collection, storage, preservation, alteration, retrieval, disclosure, making available, erasure, or destruction of, or the carrying out of logical and/or arithmetical operations on such data.

PSEUDONYMISATION Pursuant to Article 4 of the GDPR, this is the processing of personal data in such a manner that they can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.⁸⁶

86. Article 4 of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016.

Appendix IV

Checklist for integrating the Charter's principles into your processing method

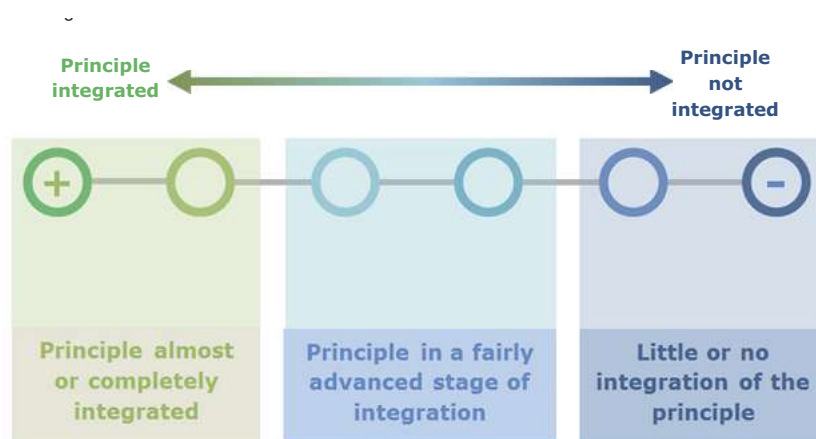
In order to assess the compatibility of your processing method with the Charter, there is a self-evaluation scale available for each of the principles listed.

For each principle, tick the box that corresponds to your processing methods.

The box furthest to the left indicates complete integration, the box furthest to the right indicates no integration.

At the bottom of the evaluation sheet (in the line marked "Total"), add up the number of boxes ticked. The column with the highest score indicates your processing method's level of compatibility with the Charter.

This evaluation is of course purely informative and by no means equates to any certification.



Checklist for evaluating your processing methods

Principle of fundamental rights:

Ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights, including the right to protection of personal data

Checklist for evaluating your processing methods

Principle of non-discrimination:

Specifically prevent the development or intensification of any discrimination between individuals or groups of individuals

Checklist for evaluating your processing methods

Principle of quality and security:

With regard to the processing of judicial decisions and data, use certified sources and intangible data with models elaborated in a secure technological environment

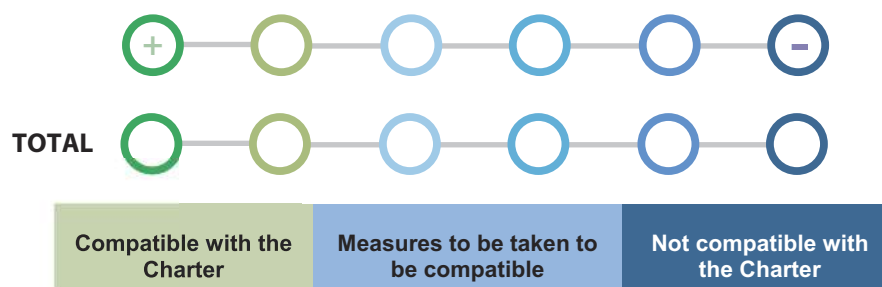
Checklist for evaluating your processing methods

Principle of transparency, impartiality and fairness:

Make data processing methods accessible and understandable, authorise external audits

Checklist for evaluating your processing methods

Principle "under user control": Preclude a prescriptive approach and ensure that users are informed actors and in control of their choices



The Charter provides a framework of principles that can guide policy makers, legislators and justice professionals when they grapple with the rapid development of Artificial Intelligence in national judicial processes.

The CEPEJ's view is that the application of Artificial Intelligence in the field of justice can contribute to improve the efficiency and quality. It must be implemented in a responsible manner which complies with the fundamental rights guaranteed in particular in the European Convention on Human Rights (ECHR) and the Council of Europe Convention on the Protection of Personal Data. It is essential to ensure that Artificial Intelligence remains a tool in the service of the general interest and that its use respects individual rights.

The Charter defines five core principles to be respected in the field of Artificial Intelligence and justice: respect of fundamental rights; non-discrimination; quality and security; transparency, impartiality and fairness; "under user control". The Charter is accompanied by a n in-depth study on the use of Artificial Intelligence in judicial systems.



www.coe.int/cepej

PREMS 005419

ENG

www.coe.int

The Council of Europe is the continent's leading human rights organisation. It comprises 47 member states, including all members of the European Union. All Council of Europe member states have signed up to the European Convention on Human Rights, a treaty designed to protect human rights, democracy and the rule of law. The European Court of Human Rights oversees the implementation of the Convention in the member states.

