si2021.eu

Virtual Conference

REGULATION OF ARTIFICIAL INTELLIGENCE – ETHICAL AND FUNDAMENTAL RIGHTS ASPECTS

EUROPEAN UNION AND INTERNATIONAL PERSPECTIVE

20 July 2021

Virtual Conference

REGULATION OF ARTIFICIAL INTELLIGENCE – ETHICAL AND FUNDAMENTAL RIGHTS ASPECTS

EUROPEAN UNION AND INTERNATIONAL PERSPECTIVE

20 July 2021

FINAL AGENDA

9.00-10.00 OPENING REMARKS

Marjan Dikaučič, Minister of Justice of the Republic of Slovenia

Francisca van Dunem, Minister of Justice of the Portuguese Republic

Christine Lambrecht, Federal Minister of Justice and Consumer Protection of Germany (video message)

Didier Reynders, European Commissioner for Justice

Adrián Vázquez Llázara, Member of the European Parliament and Chair of the Committee on Legal Affairs (JURI Committe)

Juan Fernando López Aguilar, Member of the European Parliament and Chair of the Committee on Civil Liberties, Justice and Home Affairs (LIBE Committee)

Marija Pejčinović Burić, Secretary General of the Council of Europe (video message)

10.00-12.45 1st PANEL: THE EU PERSPECTIVE

Moderator: **Dr. Maja Bogataj Jančič**, LL.M., LL.M., Founder and Head of the Intellectual Property Institute, Slovenia, Co-Chair of the GPAI Data Governance Working Group

10.00-10.50 What is Artificial Intelligence and why does regulation matter?

Panelists:
Michael O'Flaherty, Director of the Agency for Fundamental Rights

Dr Joanna Bryson, Professor of Ethics and Technology, Centre for Digital Governance at Hertie School

Miha Lobnik, Advocate of the Principle of Equality in the Republic of Slovenia and member of the Equinet Executive Board

10.50-11.00 Short break

11.00-12.45 Artificial Intelligence Act Proposal – presentation and feedback

Panelists:
Kilian Gross, Head of Unit on Artificial Intelligence Policy Development and Coordination, DG CONNECT, European Commission

Dr Joanna Bryson, Professor of Ethics and Technology, Centre for Digital Governance at Hertie School

Matthias Spielkamp, Co-founder and Executive Director, AlgorithmWatch

Catelijne Muller, LL.M., Co-founder and President of ALLAI

Discussion

12.45 -14.00 Lunch break

14.00-16.00 2nd PANEL: THE INTERNATIONAL PERSPECTIVE

Chair of the Panel: **Gregor Strojin**, LL. M., Chair of the CAHAI – Council of Europe Ad hoc Committee on Artificial Intelligence, Senior Advisor to the President of the Supreme Court of the Republic of Slovenia

Panelists:
Louisa Klingvall, Team Leader in the Fundamental rights unit, DG Justice and Consumers, European Commission

Dr David Leslie, Ethics Theme Lead at the Alan Turing Institute, CAHAI Bureau Member

Karine Perset, Head the AI Unit of the OECD Division for Digital Economy Policy (AI Policy Observatory, AI Network of Experts)

Dr Marielza Oliveira, Director for Partnerships and Operational Programme Monitoring, UNESCO

Prof Dr John Shawe Taylor, Director of IRCAI - International Research Centre on AI under the auspices of UNESCO

Discussion

16.00 Closing remarks by Trio Presidency

Prof Dr Christian Kastrop, State Secretary at the Federal Ministry of Justice and Consumer Protection of Germany

Anabela Pedroso, State Secretary at the Ministry of Justice of the Portuguese Republic

Zlatko Ratej, State Secretary at the Ministry of Justice of the Republic of Slovenia

Host: **Iztok Štefanič**, Ministry of Justice of the Republic of Slovenia

Contact person: **Maja Velič**, Ministry of Justice of the Republic of Slovenia (maja.velic@gov.si)

The conference will be held in English.

Dr. Joanna Bryson

Regulation of AI — Obstacle or Enabler?

Joanna J. Bryson

Hertie School
Centre for
Digital Governance

@j2bryson

What is AI?

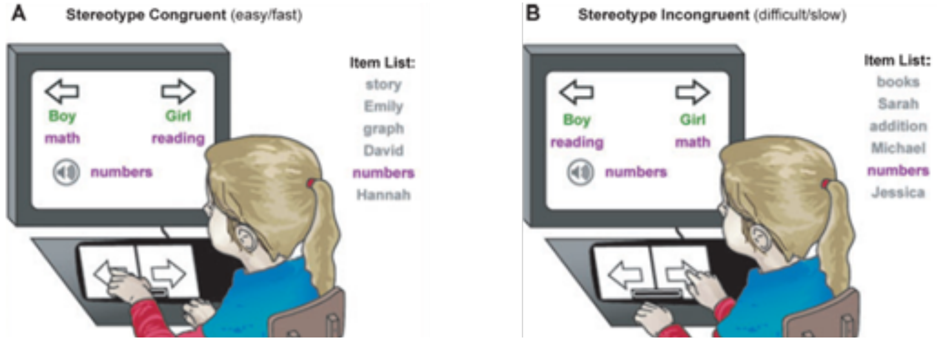
Intelligence: the capacity to do the right thing at the right time – to transform perception into action.

Artificial Intelligence: artefacts (deliberately built) facilitating our intentions through computation.

Explicit, deliberate: the parts of human intelligence humans discuss.

AI Trained on Human Language Replicates Implicit Biases

Science
AAAS



Gender bias [stereotype]

Female names: Amy, Joan, Lisa, Sarah...	Male names: John, Paul, Mike, Kevin...
Family words: home, parents, children, family...	Career words: corporation, salary, office, business, ...

Original finding [N=28k participants]: d = 1.17, $p < 10^{-2}$
Our finding [N=8x2 words]: d = 0.82, $p < 10^{-2}$

Caliskan, Bryson & Narayanan (*Science*, April 2017)

Our implicit behaviour is not our ideal.
Ideals are for explicit communication, planning.

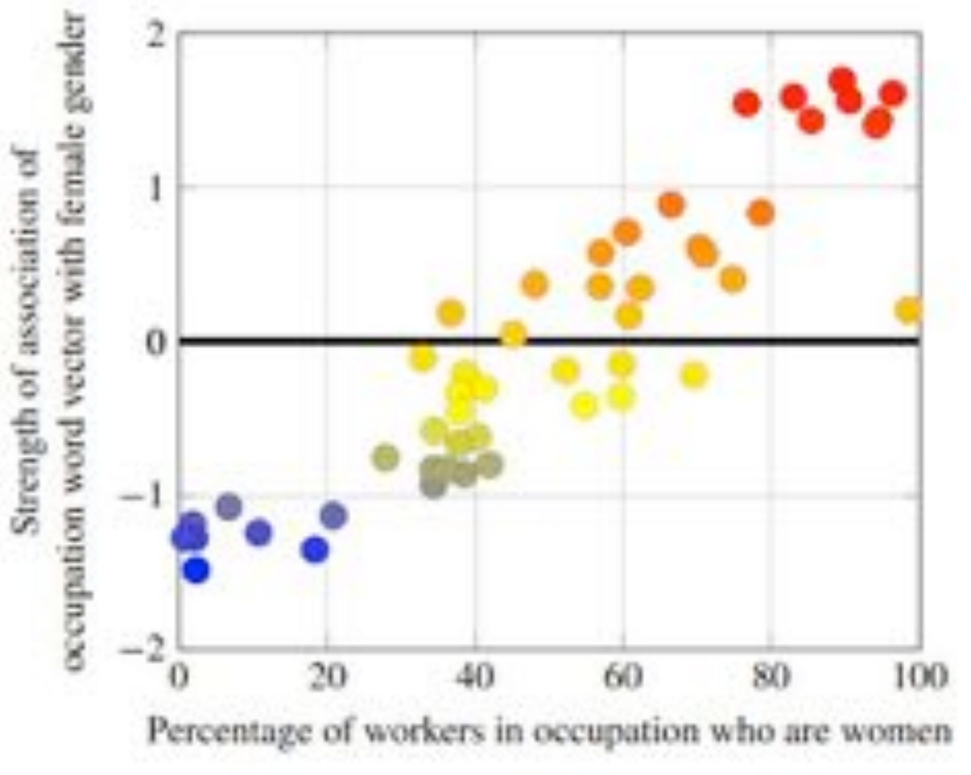
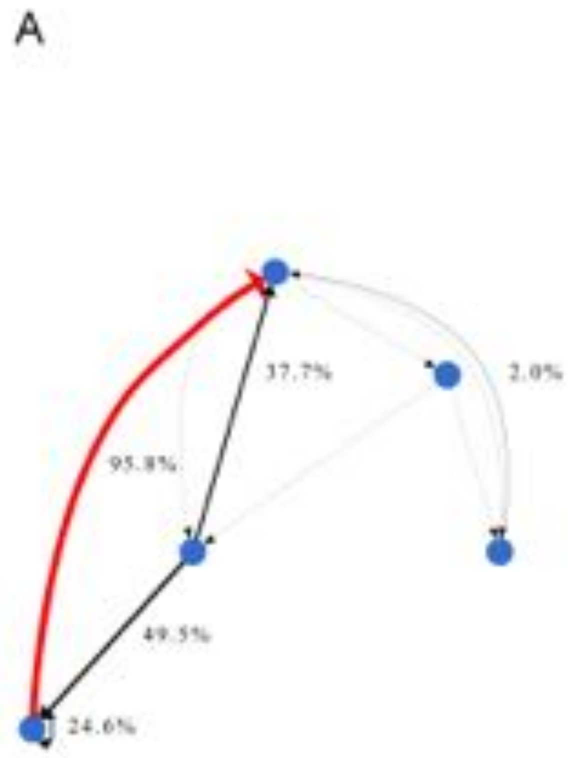


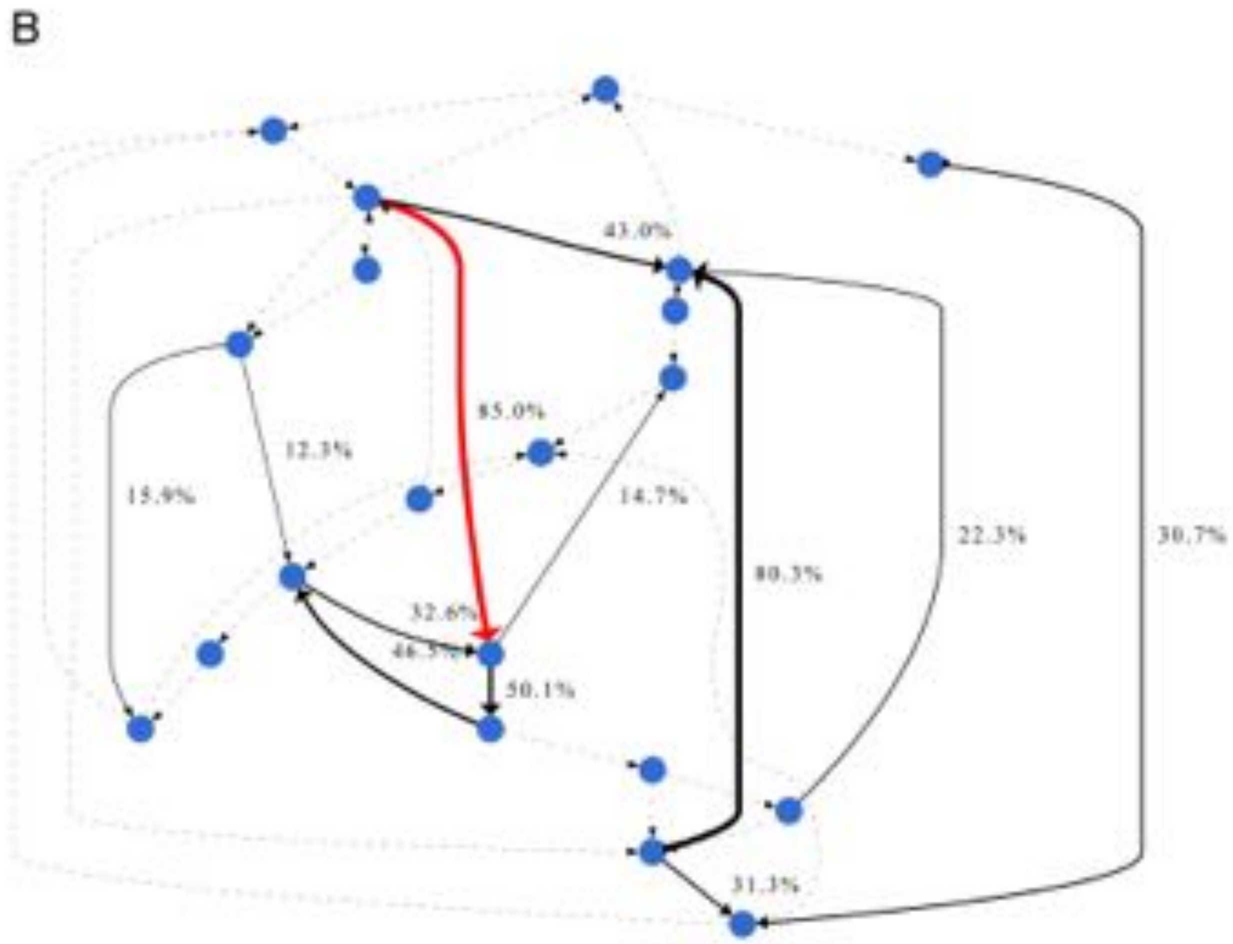
Figure 1. Occupation-gender association
Pearson's correlation coefficient $\rho = 0.90$ with p-value $< 10^{-18}$.

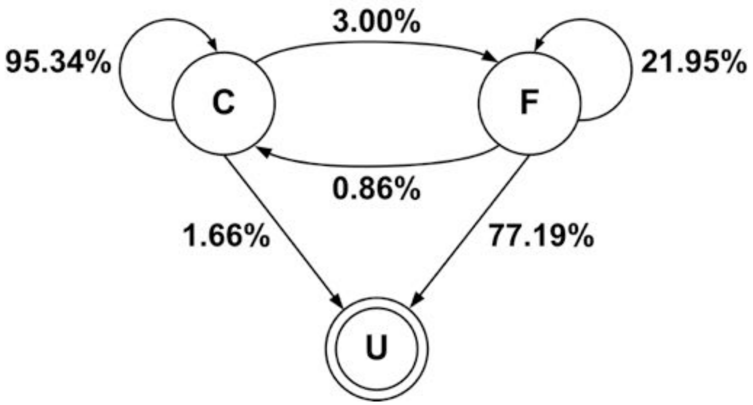
2015 US labor statistics
ρ = 0.90

What is regulation?

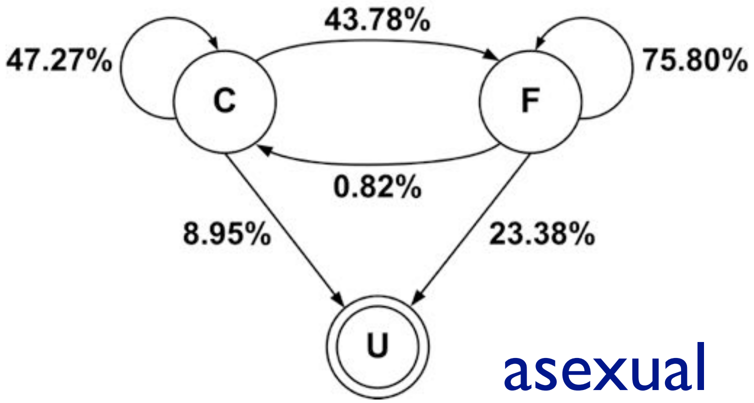
Gene Regulatory Networks







sexual



asexual

Yifei Wang

(&al 2014, 2015, 2020)

Regulation: The means by which a complex entity perpetuates a recognisable version of itself into the future.

Governance: explicit, deliberate regulation.

Regulation: The means by which a complex entity perpetuates a recognisable version of itself into the future.
Governance: explicit, deliberate regulation.

Government: An entity coordinating governance of a geographic region, possessing monopoly of force, & transnational obligations to defend human rights.

Regulation: The means by which a complex entity perpetuates a recognisable version of itself into the future.

Governance: explicit, deliberate regulation.

Government: An entity coordinating governance of a geographic region, possessing monopoly of force, & transnational obligations to defend human rights.

Governments typically provide "up regulation" (support, infrastructure) and "down regulation" (restrictions.)

Regulation: The means by which a complex entity perpetuates a recognisable version of itself into the future.

Governance: explicit, deliberate regulation.

Government: An entity coordinating governance of a geographic region, possessing monopoly of force, & transnational obligations to defend human rights.

Governments typically provide "up regulation" (support, infrastructure) and "down regulation" (restrictions).

'Restrictions' benefit innovation via stability, and sustainability.

Does Regulation Get in the Way?

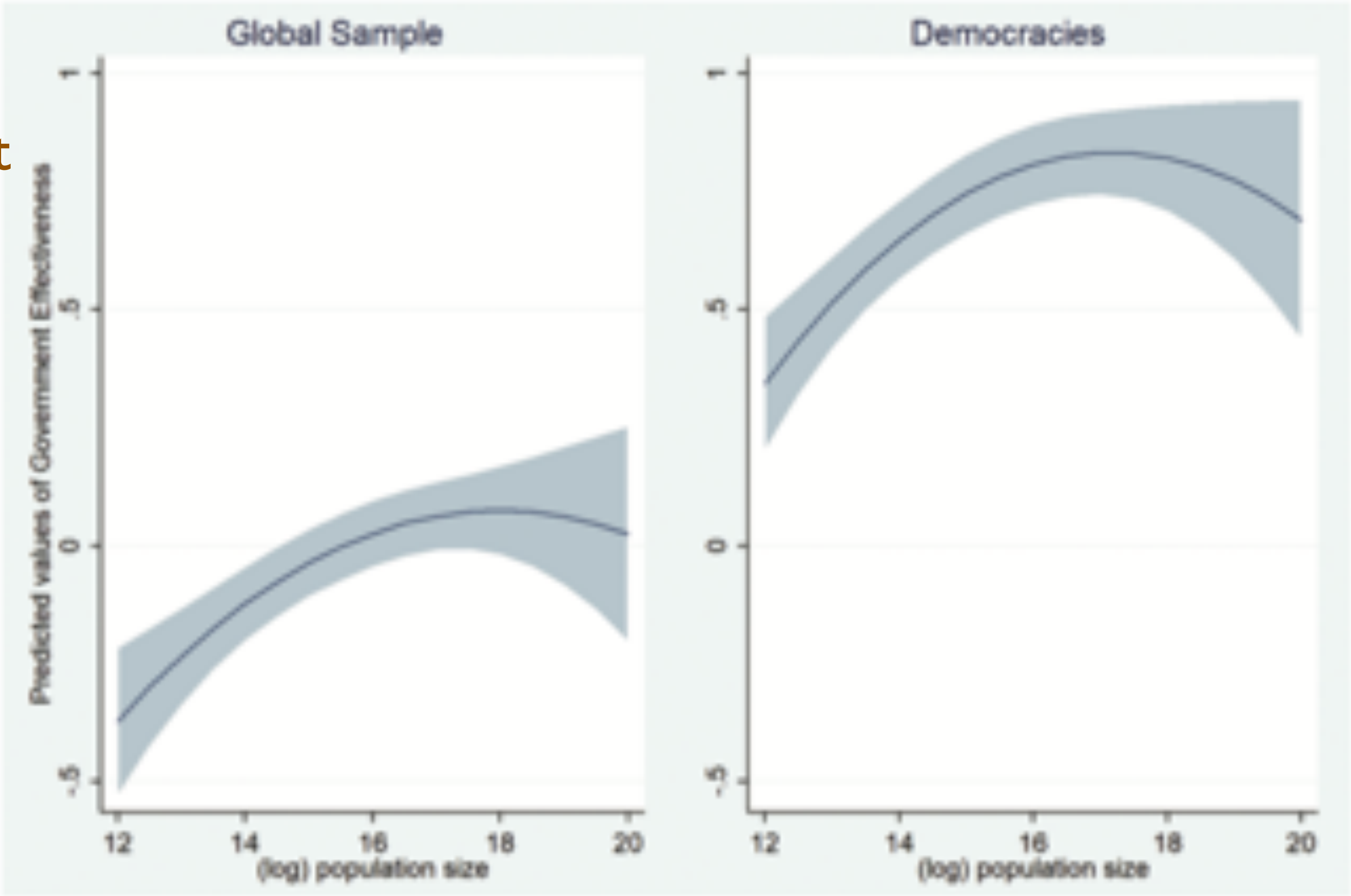
Journal of Public Administration Research and Theory, 2019, Vol. 29, No. 1

↑y axis = Government Effectiveness measure by the World Bank.

→x axis = log population

Peak: 66M – 485M

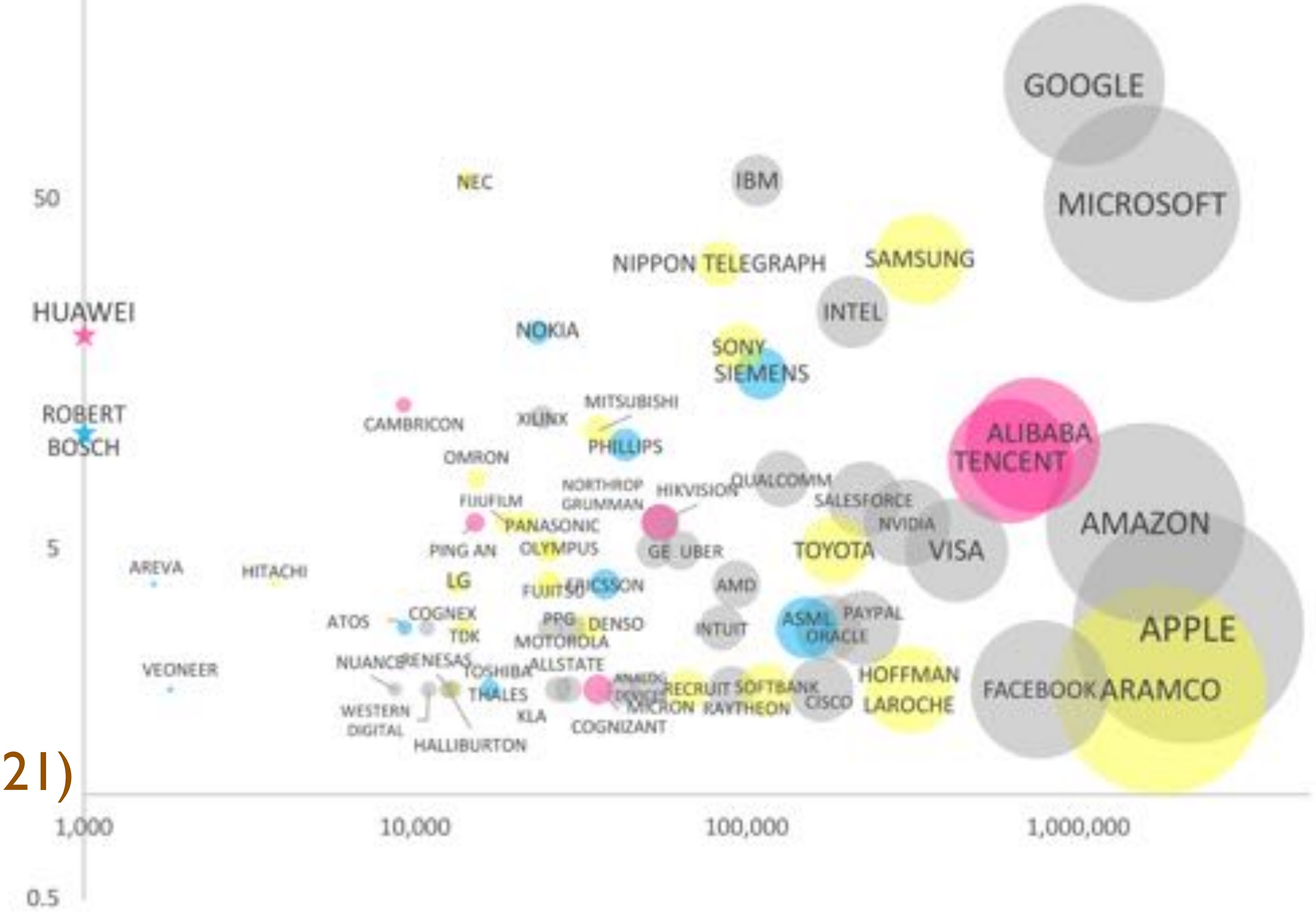
Jugl (2019)

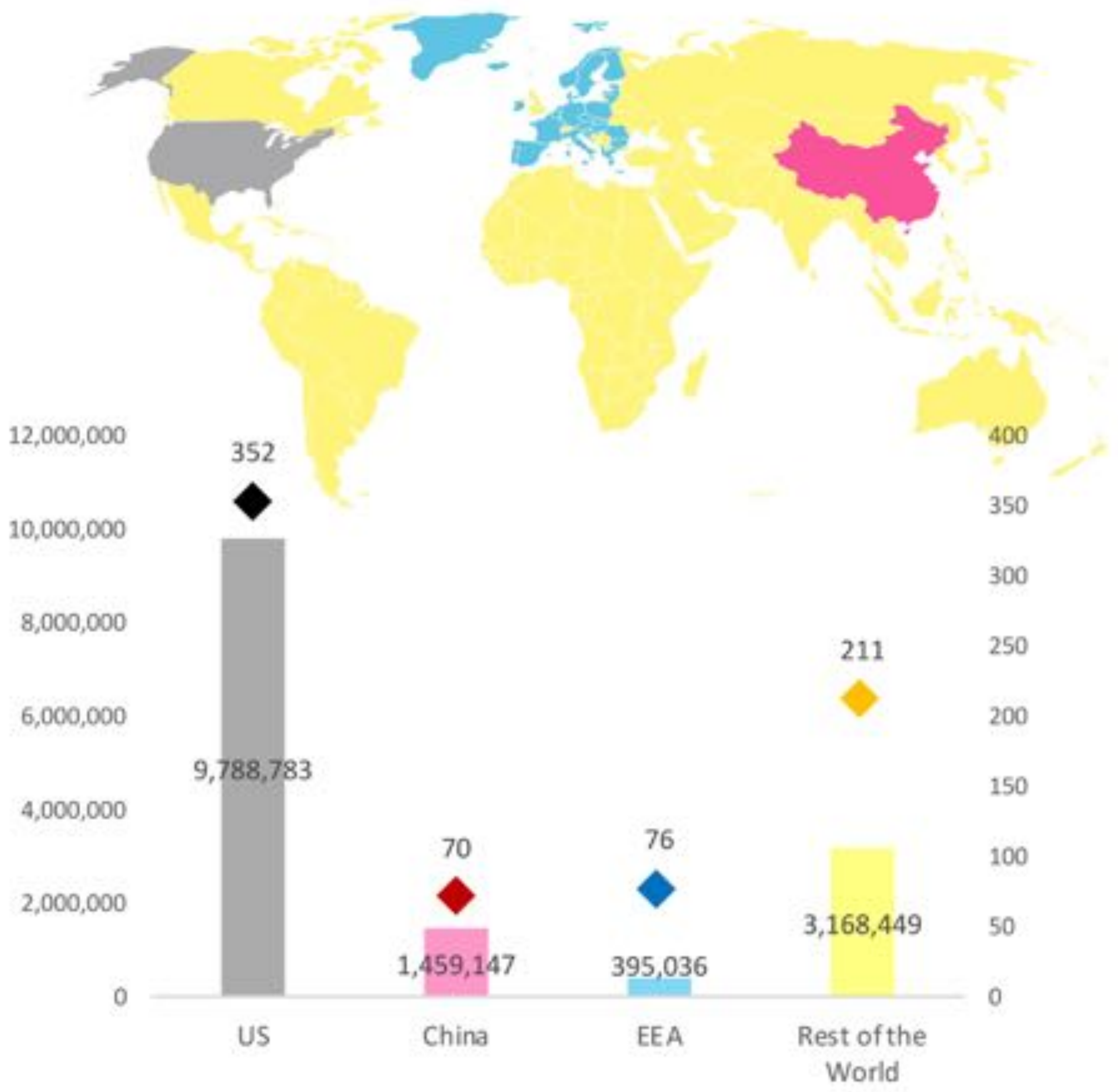


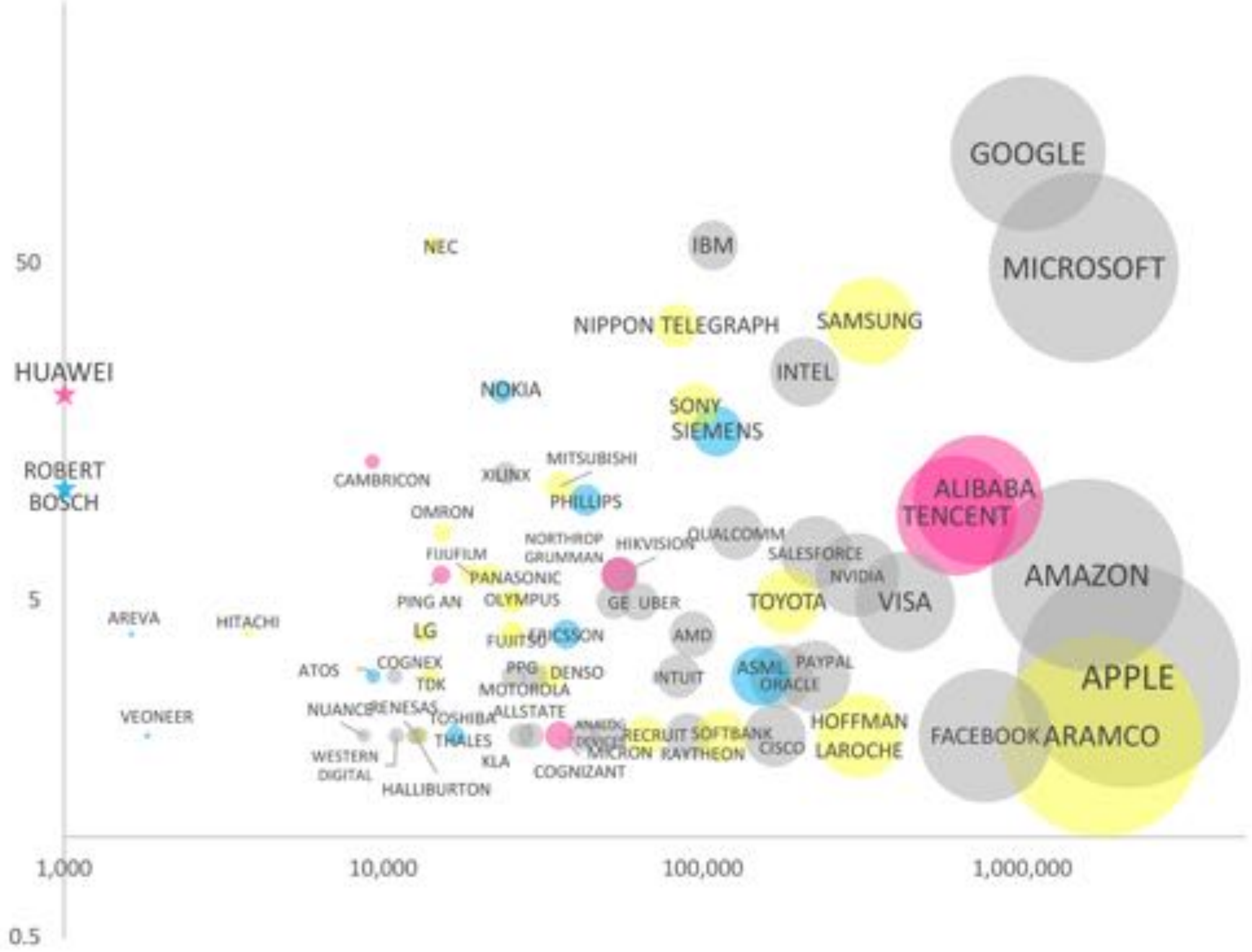
↑y axis = log 2019 AI patents in WIPO (G06N of the IPC classification dedicated to “Computer sys- tems based on specific computational models”)

→x axis = log Oct 2020 Market Capitalisation

Bryson & Malikova (2021)





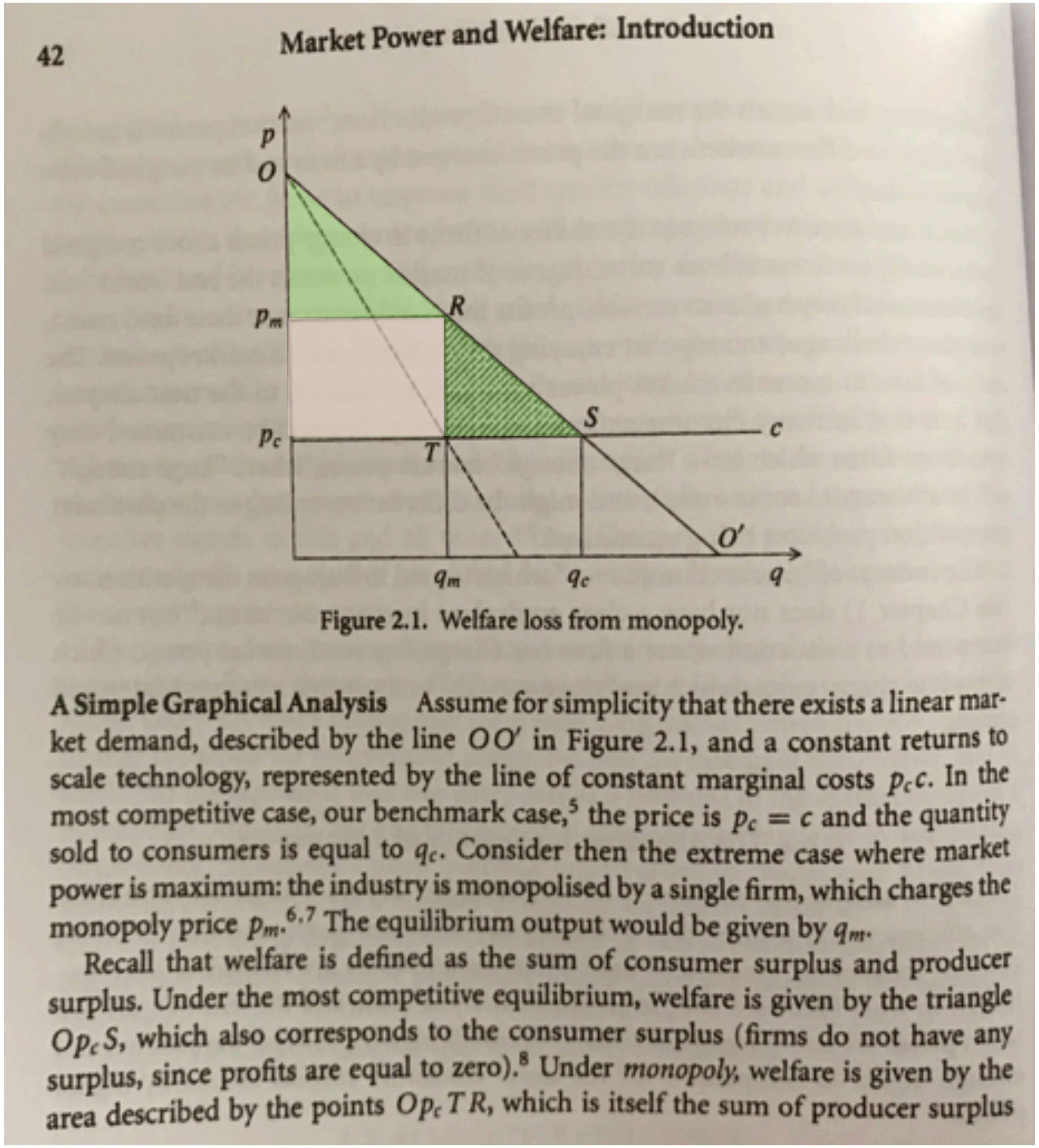


Bryson & Malikova (2021)

Is there an AI cold war? *Global Perspectives* 2(1)

y↑: price
x→: quantity sold
O↘O’: demand

fair-price welfare = consumer surplus + (producer surplus = 0) for full competition;
With market power, overall welfare declines, but producers get some surplus.



42 Market Power and Welfare: Introduction

p
O
p_m
R
p_c
T
S
c
O′
q_m
q_c
q

Figure 2.1. Welfare loss from monopoly.

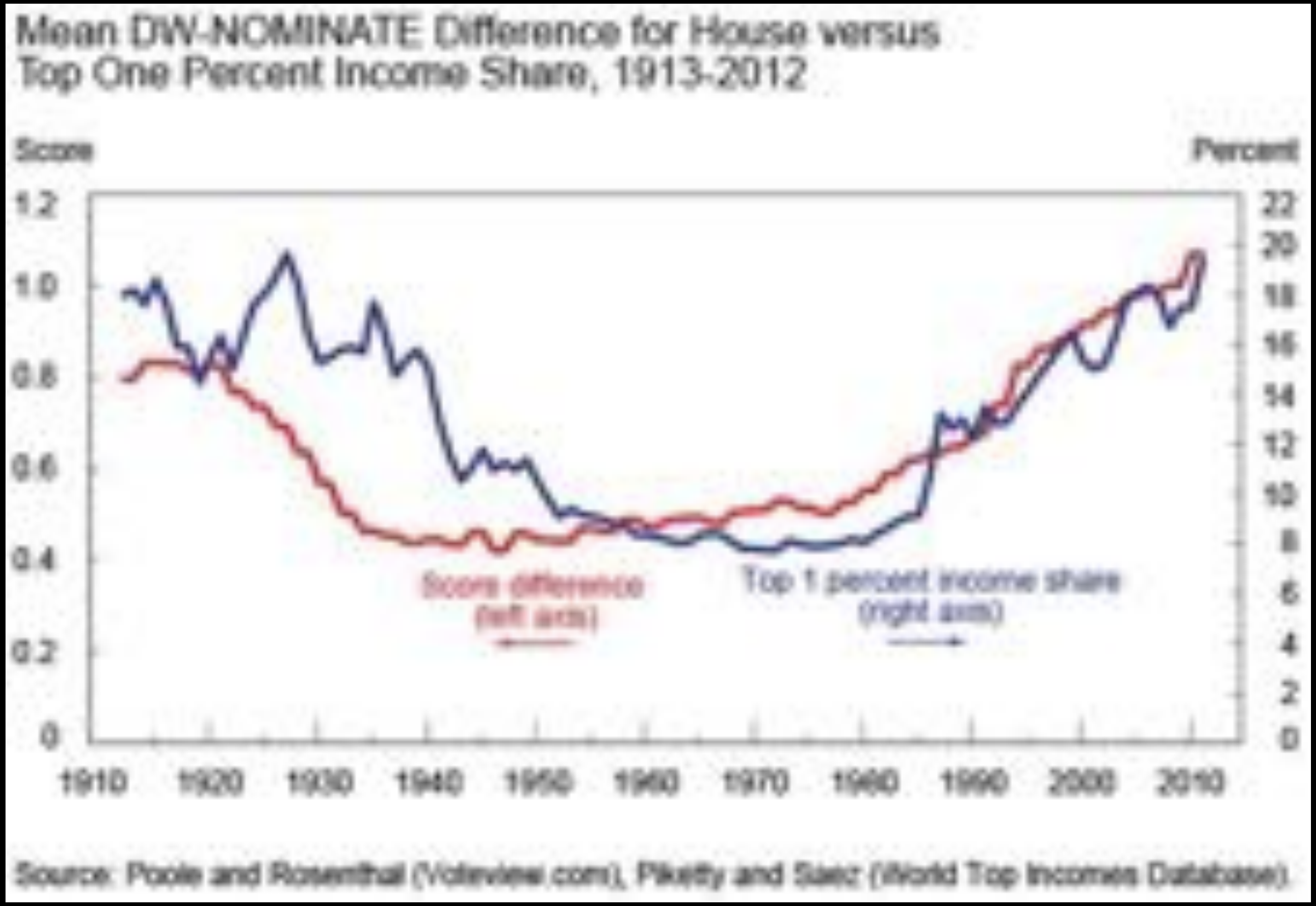
A Simple Graphical Analysis Assume for simplicity that there exists a linear market demand, described by the line OO' in Figure 2.1, and a constant returns to scale technology, represented by the line of constant marginal costs $p_c c$. In the most competitive case, our benchmark case,[5] the price is $p_c = c$ and the quantity sold to consumers is equal to q_c. Consider then the extreme case where market power is maximum: the industry is monopolised by a single firm, which charges the monopoly price p_m.[6,7] The equilibrium output would be given by q_m.

Recall that welfare is defined as the sum of consumer surplus and producer surplus. Under the most competitive equilibrium, welfare is given by the triangle Op_cS, which also corresponds to the consumer surplus (firms do not have any surplus, since profits are equal to zero).[8] Under *monopoly*, welfare is given by the area described by the points Op_cTR, which is itself the sum of producer surplus

Competition Policy: Theory & Practice

Massimo Motta (2004)

Regulatory Capture, Inequality, and Political Polarisation

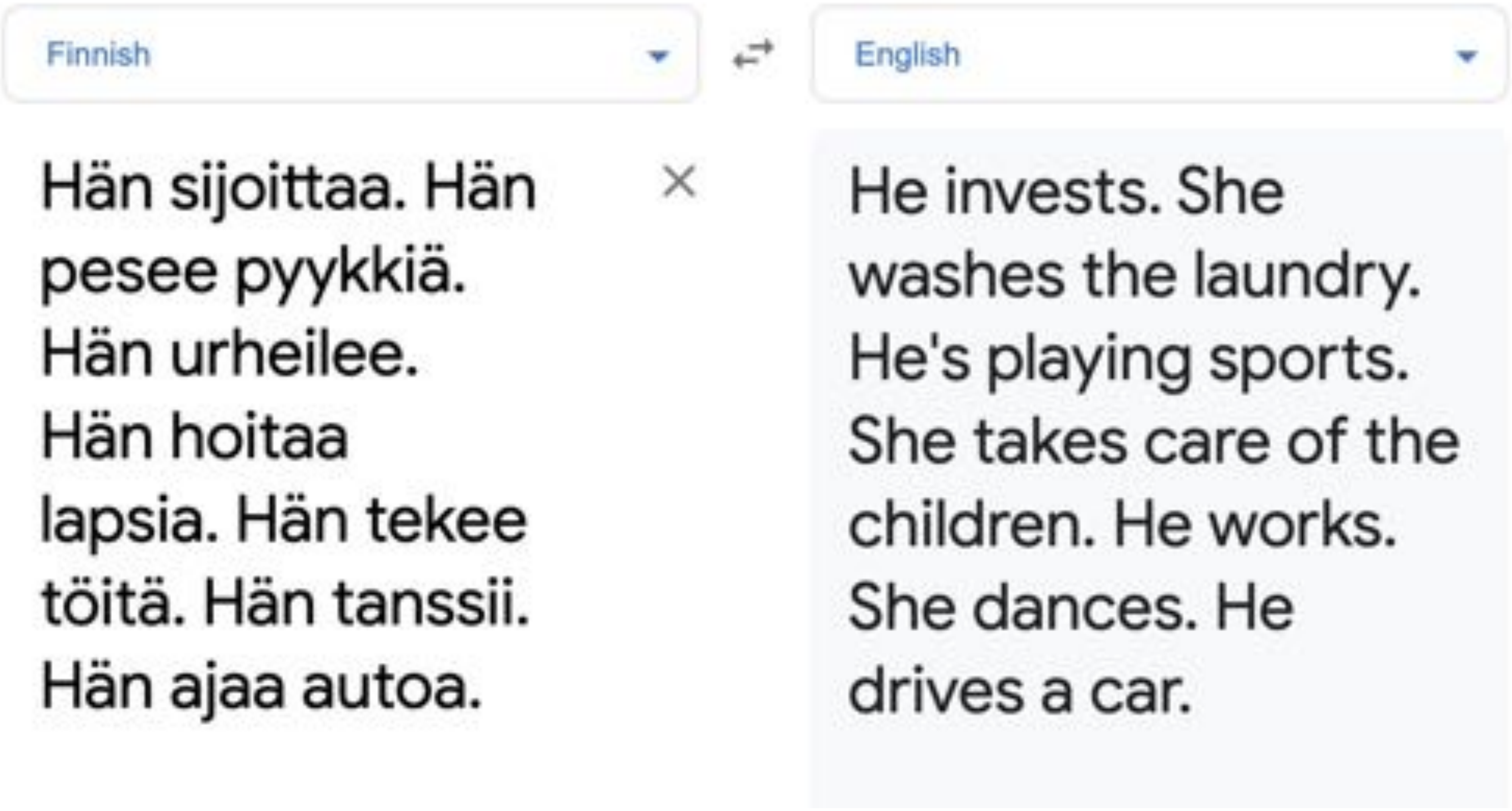


Mean DW-NOMINATE Difference for House versus Top One Percent Income Share, 1913-2012

Source: Poole and Rosenthal (Voteview.com), Piketty and Saez (World Top Incomes Database).

- Inadequate governance of organizations or sectors leads to regulatory capture and inequality.
- Inequality leads to social unrest, loss of social mobility, decline in innovation, general insecurity.
- Political polarisation is correlated with inequality, may be caused by it (Stewart, Bryson & McCarty 2020.)

How to Regulate AI

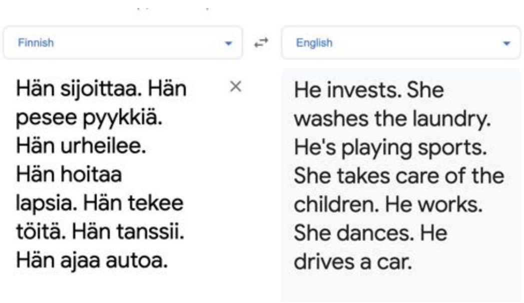


Hän sijoittaa. Hän pesee pyykkiä. Hän urheilee. Hän hoitaa lapsia. Hän tekee töitä. Hän tanssii. Hän ajaa autoa.

He invests. She washes the laundry. He's playing sports. She takes care of the children. He works. She dances. He drives a car.

@vuokko recently, though Aylin Caliskan did it first

Translator



Hän sijoittaa. Hän pesee pyykkiä. Hän urheilee. Hän hoitaa lapsia. Hän tekee töitä. Hän tanssii. Hän ajaa autoa.

He invests. She washes the laundry. He's playing sports. She takes care of the children. He works. She dances. He drives a car.

ML
simple, transparent algorithm

stereotyped output

XAI human readable hacks

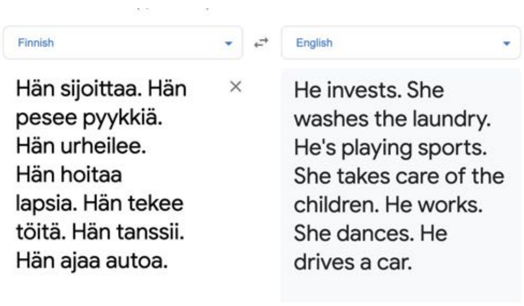
predefined fair output

Replicates lived experience

Tests of completeness documented in design plans

@vuokko recently, though Aylin Caliskan did it first

Translator



Hän sijoittaa. Hän pesee pyykkiä. Hän urheilee. Hän hoitaa lapsia. Hän tekee töitä. Hän tanssii. Hän ajaa autoa.

He invests. She washes the laundry. He's playing sports. She takes care of the children. He works. She dances. He drives a car.

the whole thing is the translator

ML
simple, transparent algorithm

stereotyped output

ML simple, transparent alg.

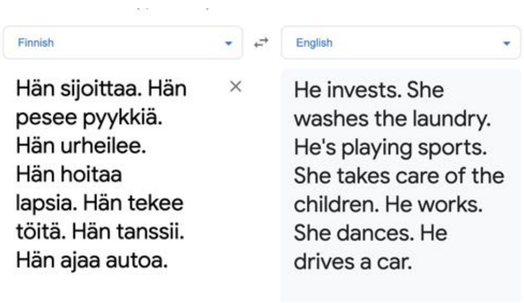
predefined fair output

Replicates lived experience

Tests of completeness documented in design plans

@vuokko recently, though Aylin Caliskan did it first

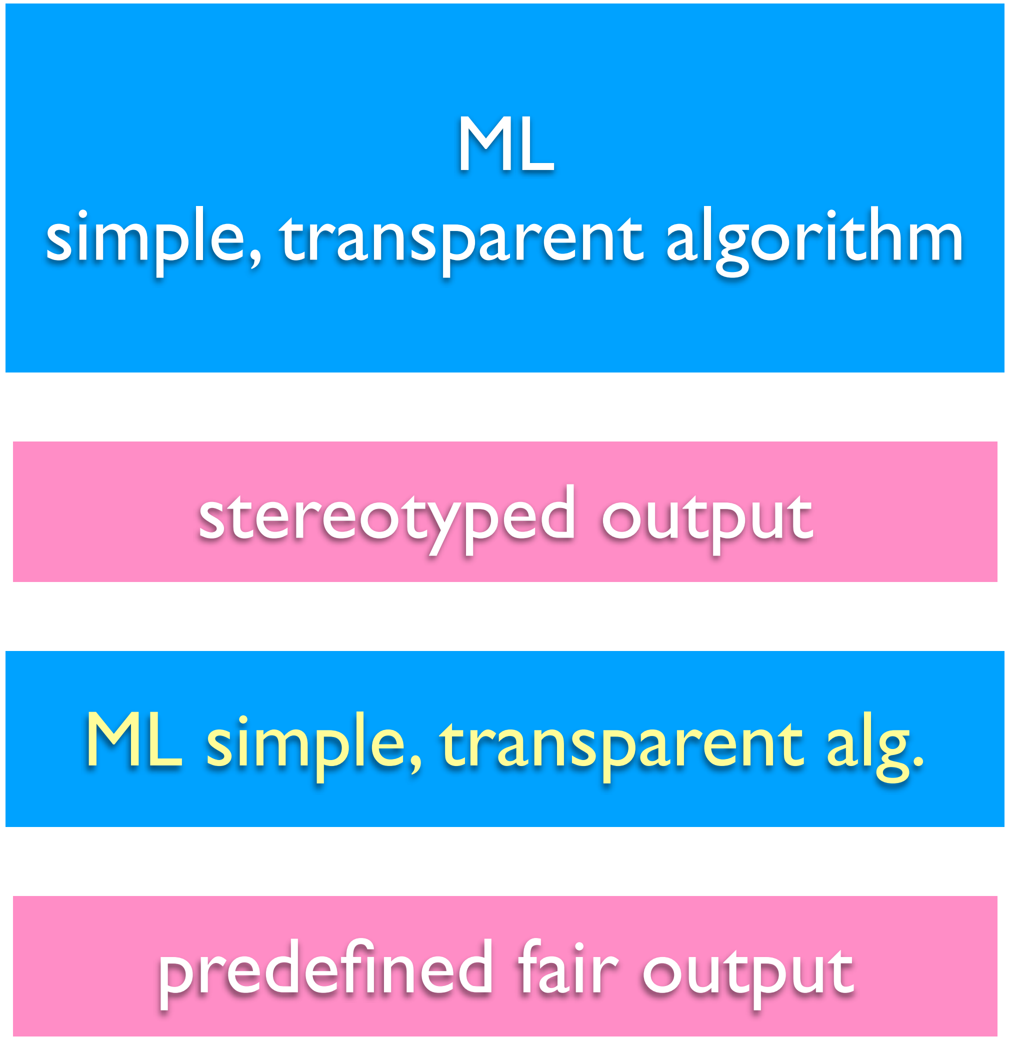
Translator



Hän sijoittaa. Hän pesee pyykkiä. Hän urheilee. Hän hoitaa lapsia. Hän tekee töitä. Hän tanssii. Hän ajaa autoa.

He invests. She washes the laundry. He's playing sports. She takes care of the children. He works. She dances. He drives a car.

the whole thing is the translator



ML
simple, transparent algorithm

stereotyped output

ML simple, transparent alg.

predefined fair output

Each stage should be auditable and replicable.

Each stage demonstrably meets criteria.

Accountability for AI is possible, but requires reliable enforcement – governance.

Can we trust a government?

No.

We have to actively make sure a government works*.

*Governments are a principle means by which we ensure everyone does what's fair, just, and sustainable.

Sustainability allows us to flourish securely.

Thanks!

→

for bubble
charts &
monopoly

→

Helena
Malikova

AI and Other Acts

Joanna J. Bryson

Hertie School
Centre for
Digital Governance

@j2bryson

The AI Regulation / Act

The Digital Services Act

The Digital Markets Act

Liability, GDPR,…

What Actually Matters

- Sufficient transparency for accountability.
- Liability / enforcement to prevent both negligence and malfeasance.
- Proportionality / minimal barriers to entry – ways for robust, agile economies of SMEs to thrive.

What Actually Matters

- Sufficient transparency for accountability.
- Liability / enforcement to prevent both negligence and malfeasance.
- Proportionality / minimal barriers to entry – ways for robust, agile economies of SMEs to thrive.

How to Get There

- Clarify / enshrine in law that all software is a manufactured product, even if that product is used to offer a service, or includes AI.
    - This gives you requirements on product safety, due diligence.
- Make all requirements proportional to corporations' own–as well as externally assessed–risk (not just those in the DSA).

Proportionality

- The DSA requires that large corporations

  1. assess for themselves risks posed by their products,
  2. propose remedies for these risks.

- This means the regulator only needs enough talent to check the work of the corporation.

- What work is “large” doing here? Real proportionality would let corporations do this to the extent they perceive risk.

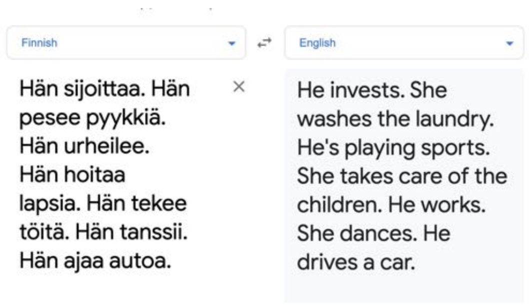
  3. Similarly for transparency – let companies put as much resource into transparency as they assess they expect to need for audits, liability.
  4. Requires resourcing for enforcement, so some risk is perceived.

Digital Systems Are Easily Transparent

- What we audit is not the micro details of how AI works, but how humans behave when they build, train, test deploy, and monitor it.
- Architecture documents of the system: design of its components, processes for development, use, and maintenance.
- Security documents for the system. Including logs; provenance of software & data libraries.
- Logs of every change to the code base – who made the change, when, and why. For ML, log also data libraries, and model parameters.
- Logs of testing before and during release; and performance – inputs and decisions – of operational systems.
- All benefit the developers, and are auditable (cf. DSA).

cf Bryson OUP 2020

Translator



Hän sijoittaa. Hän pesee pyykkiä. Hän urheilee. Hän hoitaa lapsia. Hän tekee töitä. Hän tanssii. Hän ajaa autoa.

He invests. She washes the laundry. He's playing sports. She takes care of the children. He works. She dances. He drives a car.

the whole thing is the translator

ML
simple, transparent algorithm

stereotyped output

ML simple, transparent alg.

predefined fair output

Each stage should be auditable and replicable.

Each stage demonstrably meets criteria.

Accountability for AI is possible, but requires reliable enforcement – governance.

What Is the AI Act for?

- Seems to be more or less specific to the subset of software and that citizens interact with directly or are individually affected by.
  - AI is the subset of ICT that people over identify with.
  - Fine, make it about how to have these conversations.
- Don’t define the regulated systems (“AI”) by how they work. Define the affected systems by their outcomes.
- Don’t try to create hard boundaries around what throws you into a particular regulatory regime (“levels”).
  - Maybe have some trigger thresholds, but allow corporations to assess their own risk as being at a potentially higher level, seek auditing / certification for liability defence under the regulation.

Thanks!

OECD Principles of AI

Endorsed by 44 world governments
22 May 2019, + the G20 same year.

AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.

AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.

There should be transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them [the AI systems] and can challenge outcomes.

AI systems must function in a robust, secure and safe way throughout their lifetimes, and potential risks should be continually assessed and managed.

Organisations and individuals developing, deploying or operating AI systems should be held accountable for their proper functioning in line with the above principles.

OECD Principles of AI

Endorsed by 44 world governments 22 May 2019, + the G20 same year.

Human-centred

AI should benefit people and the planet by driving inclusive growth, sustainable development and well-being.

“Fair”

AI systems should be designed in a way that respects the rule of law, human rights, democratic values and diversity, and they should include appropriate safeguards – for example, enabling human intervention where necessary – to ensure a fair and just society.

Transparent

There should be transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them [the AI systems] and can challenge outcomes.

Safe

AI systems must function in a robust, secure and safe way throughout their lifetimes, and potential risks should be continually assessed and managed.

Accountable

Organisations and individuals developing, deploying or operating systems should be held accountable for their proper functioning in line with the above principles.

cf Floridi &al. 2018

UK Principles of Robotics (2011)

Asimov's Laws revised for Manufacturer Responsibility

1. Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.
2. Humans, not robots, are responsible agents. Robots should be designed & operated as far as is practicable to comply with existing laws & fundamental rights & freedoms, including privacy.
3. Robots are products. They should be designed using processes which assure their safety and security. [devops]
4. Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.
5. The person with legal responsibility for a robot should be attributed. [like automobile titles]

Owner / Operator Respon-sibility

cf Bryson *AISBQ* 2000; Bryson; Prescott; Boden & al (special issue) *Connection Science*, 2017

UK Principles of Robotics (2011)

Human-centred

1. Robots are multi-use tools. Robots should not be designed solely or primarily to kill or harm humans, except in the interests of national security.

(non) Lethal

2. Humans, not robots, are responsible agents. Robots should be designed & operated as far as is practicable to comply with existing laws & fundamental rights & freedoms, including privacy.

Just

Ethical

3. Robots are products. They should be designed using processes which assure their safety and security. [devops]

Safe

4. Robots are manufactured artefacts. They should not be designed in a deceptive way to exploit vulnerable users; instead their machine nature should be transparent.

Transparent

5. The person with legal responsibility for a robot should be attributed. [like automobile titles]

Accountable

cf Bryson *AISBQ* 2000; Bryson; Prescott; Boden & al (special issue) *Connection Science*, 2017

Kilian Gross

European
Commission

SHAPING EUROPE'S DIGITAL FUTURE

AI is good …

- For citizens
- For business
- For the public interest

… but creates some risks

- For the safety of consumers and users
- For fundamental rights

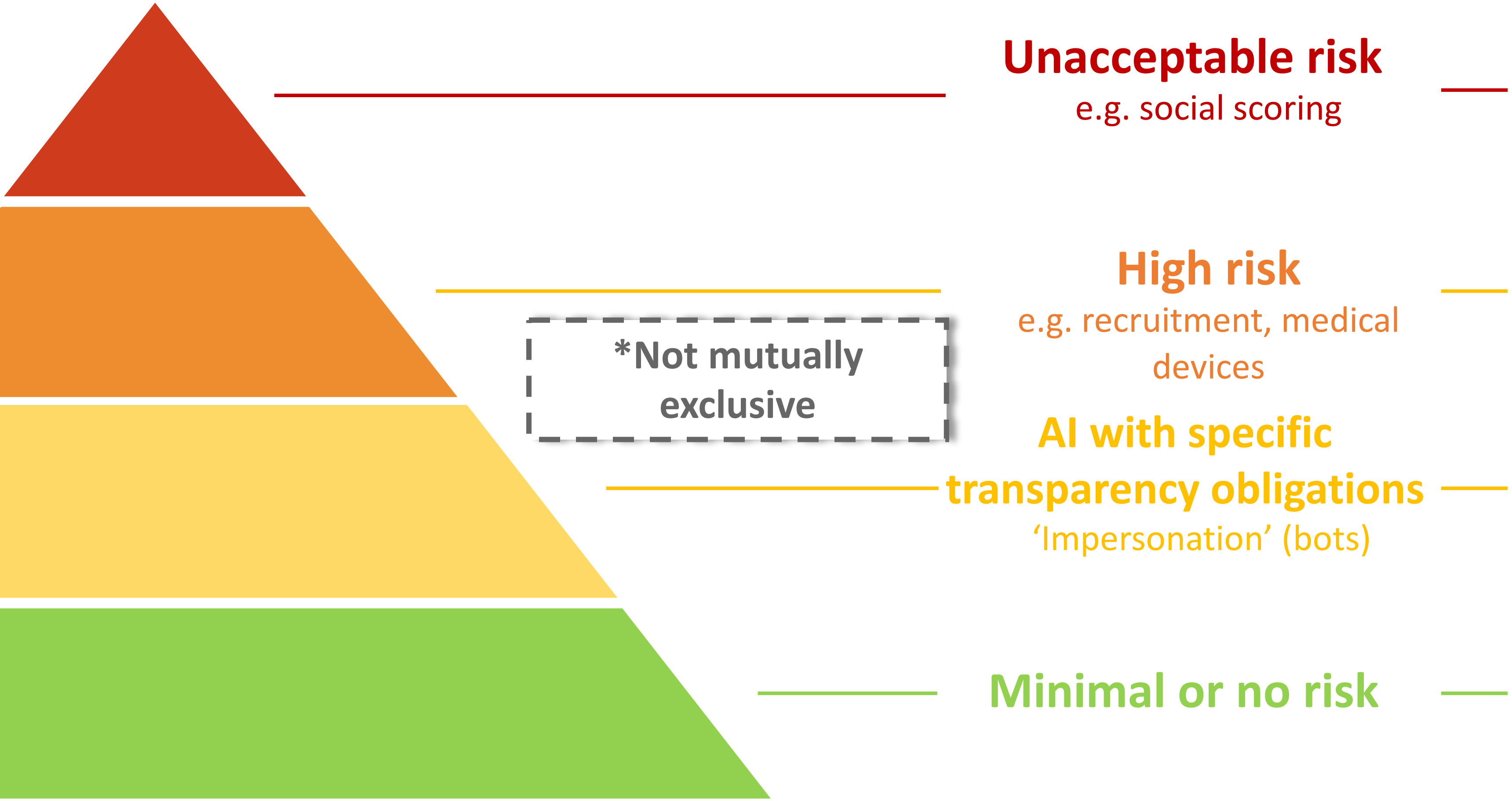
Definition and technological scope of the regulation (Art. 3)

Definition of Artificial Intelligence

- Definition of AI should be **as neutral as possible** in order to cover techniques which are not yet known/developed
- **Overall aim is to cover all AI**, including traditional symbolic AI, Machine learning, as well as hybrid systems
- **Annex I**: list of AI techniques and approaches should provide for legal certainty (adaptations over time may be necessary)

> *"a software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with"*

A risk-based approach to regulation



*Not mutually exclusive

Unacceptable risk
e.g. social scoring

Prohibited

High risk
e.g. recruitment, medical devices

Permitted subject to compliance with AI requirements and ex-ante conformity assessment

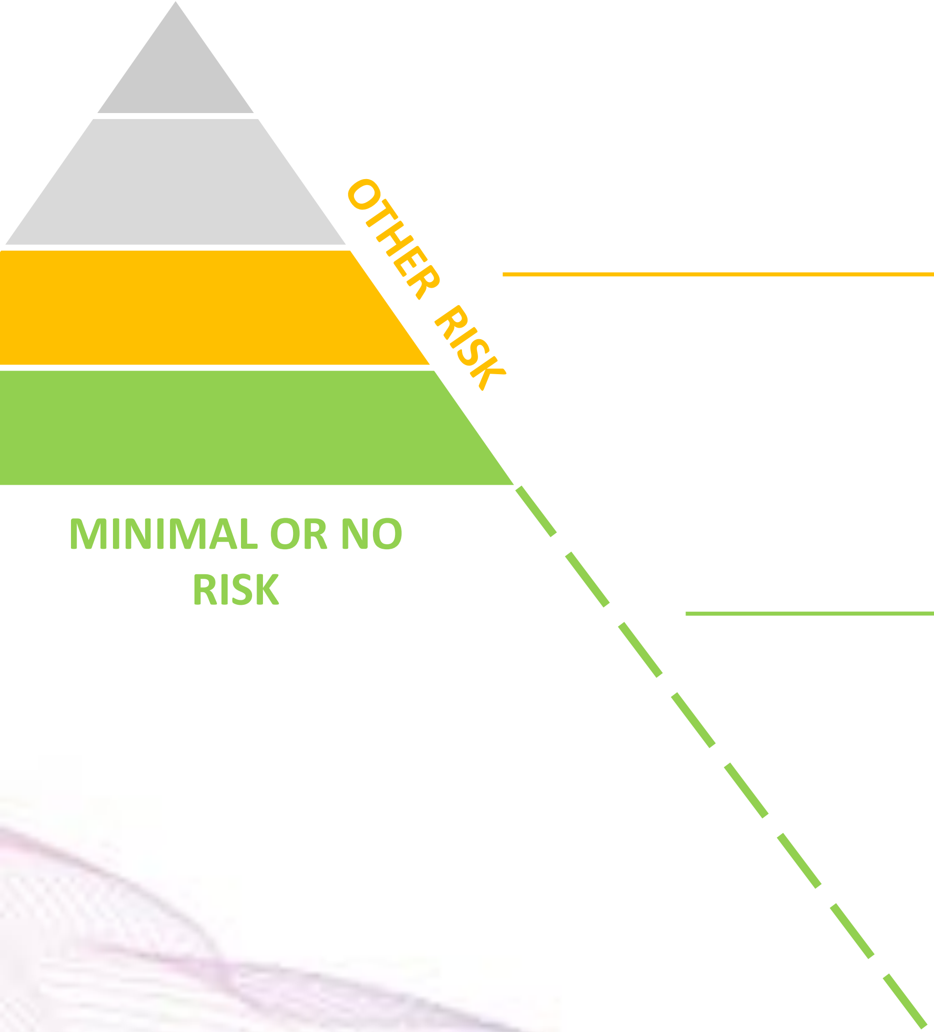
AI with specific transparency obligations
'Impersonation' (bots)

Permitted but subject to information/transparency Obligations

Minimal or no risk

Permitted with no restrictions

Most AI systems will not be high-risk (Titles IV, IX)



New transparency obligations for certain AI systems (Art. 52)

- **Notify humans** that they are **interacting with an AI system** unless this is evident
- Notify humans that emotional recognition or biometric categorisation systems are applied to them
- Apply **label to deep fakes** (unless necessary for the exercise of a fundamental right or freedom or for reasons of public interests)

Possible voluntary codes of conduct for AI with specific transparency requirements (Art. 69)

- No mandatory obligations
- Commission and Board to encourage drawing up of codes of conduct intended to foster the **voluntary application of requirements to low-risk AI systems**

High-risk Artificial Intelligence Systems (Title III, Annexes II and III)

Certain applications in the following fields:

1 **SAFETY COMPONENTS OF REGULATED PRODUCTS**

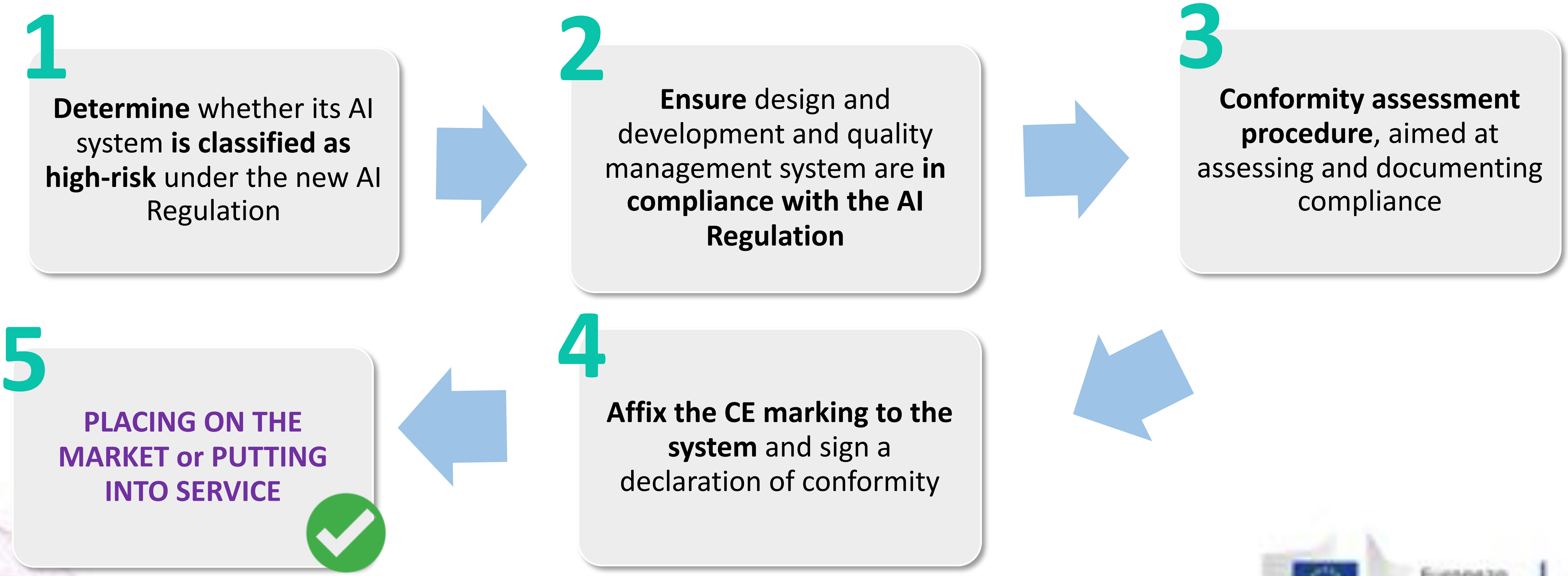
(e.g. medical devices, machinery) which are subject to third-party assessment under the relevant sectorial legislation

2 **CERTAIN (STAND-ALONE) AI SYSTEMS IN THE FOLLOWING FIELDS**

- ✓ Biometric identification and categorisation of natural persons
- ✓ Management and operation of critical infrastructure
- ✓ Education and vocational training
- ✓ Employment and workers management, access to self-employment
- ✓ Access to and enjoyment of essential private services and public services and benefits
- ✓ Law enforcement
- ✓ Migration, asylum and border control management
- ✓ Administration of justice and democratic processes

CE marking and process (Title III, chapter 4, art. 49.)

CE marking is an indication that a product complies with the requirements of a relevant Union legislation regulating the product in question. In order to affix a CE marking to a high-risk AI system, a provider shall undertake **the following steps:**



1 **Determine** whether its AI system **is classified as high-risk** under the new AI Regulation

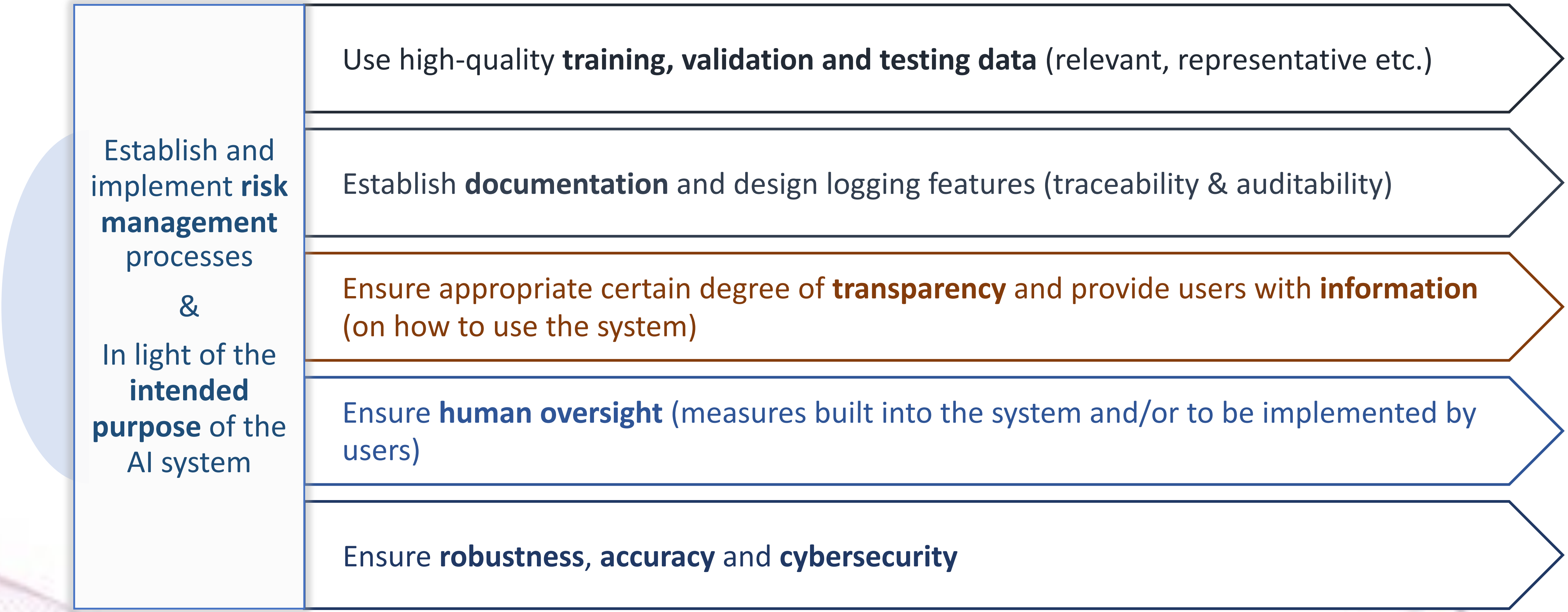
2 **Ensure** design and development and quality management system are **in compliance with the AI Regulation**

3 **Conformity assessment procedure**, aimed at assessing and documenting compliance

4 **Affix the CE marking to the system** and sign a declaration of conformity

5 **PLACING ON THE MARKET or PUTTING INTO SERVICE**

Requirements for high-risk AI (Title III, chapter 2)



Establish and implement **risk management** processes

&

In light of the **intended purpose** of the AI system

Use high-quality **training, validation and testing data** (relevant, representative etc.)

Establish **documentation** and design logging features (traceability & auditability)

Ensure appropriate certain degree of **transparency** and provide users with **information** (on how to use the system)

Ensure **human oversight** (measures built into the system and/or to be implemented by users)

Ensure **robustness**, **accuracy** and **cybersecurity**

AI that contradicts EU values is prohibited (Title II, Article 5)

X **Subliminal manipulation** resulting in physical/ psychological harm

Example: An **inaudible sound** is played in truck drivers' cabins to push them to **drive longer than healthy and safe**. AI is used to find the frequency maximising this effect on drivers.

X **Exploitation of children or mentally disabled persons** resulting in physical/psychological harm

Example: A doll with an integrated **voice assistant** encourages a minor to **engage in progressively dangerous behavior** or challenges in the guise of a fun or cool game.

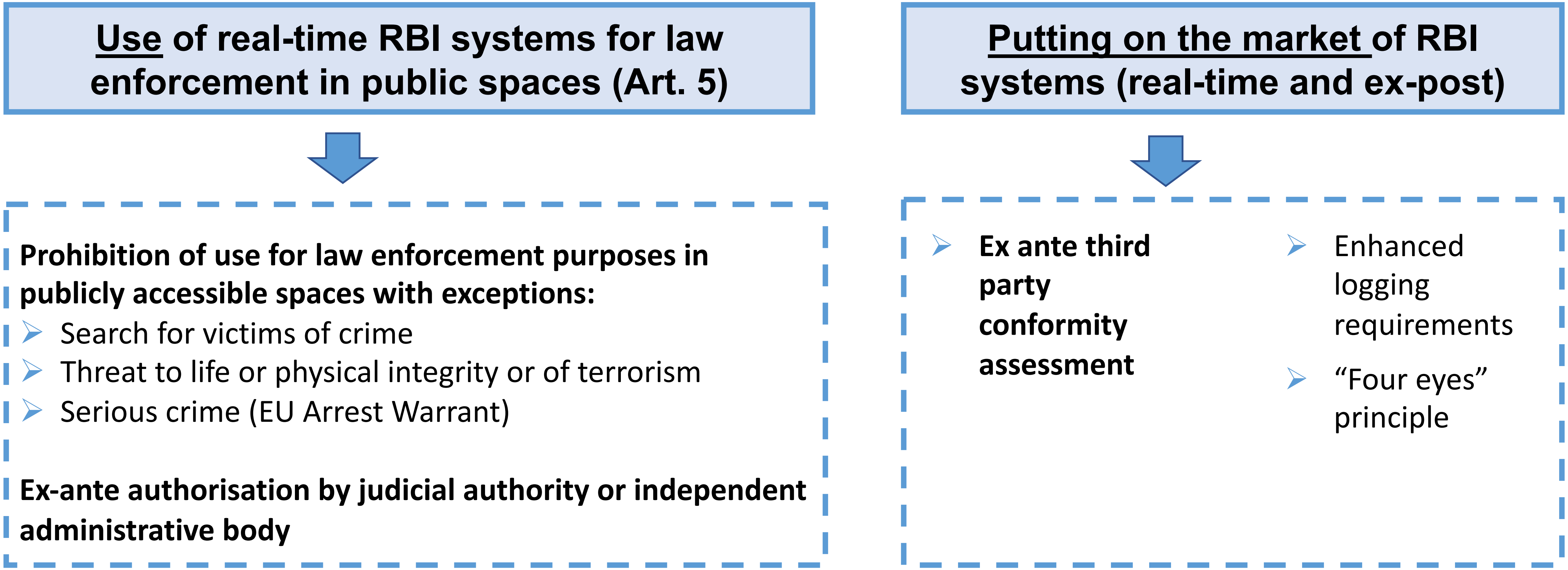
X **General purpose social scoring**

Example: An AI system **identifies at-risk children** in need of social care **based on insignificant or irrelevant social 'misbehavior'** of parents, e.g. missing a doctor's appointment or divorce.

X **Remote biometric identification for** law enforcement purposes in publicly accessible spaces (with exceptions)

Example: All faces captured live by video cameras checked, in real time, against a database to identify a terrorist.

Remote biometric identification (RBI) (Title II, Art. 5, Title III)



Use of real-time RBI systems for law enforcement in public spaces (Art. 5)

Prohibition of use for law enforcement purposes in publicly accessible spaces with exceptions:

- Search for victims of crime
- Threat to life or physical integrity or of terrorism
- Serious crime (EU Arrest Warrant)

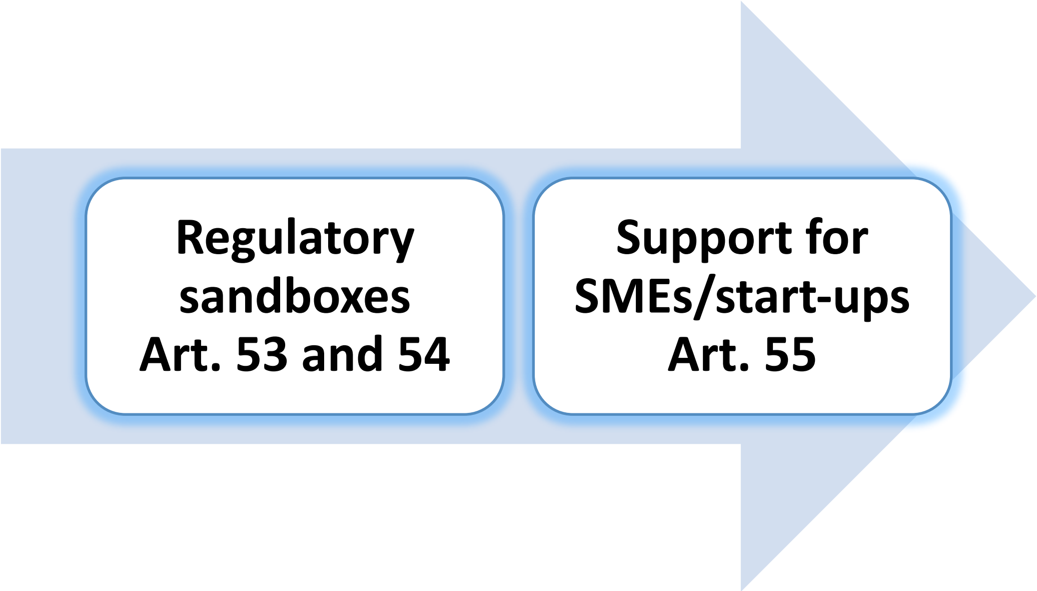
Ex-ante authorisation by judicial authority or independent administrative body

Putting on the market of RBI systems (real-time and ex-post)

- **Ex ante third party conformity assessment**
- Enhanced logging requirements
- "Four eyes" principle

No additional rules foreseen for use of real-time and post RBI systems: existing data protection rules apply

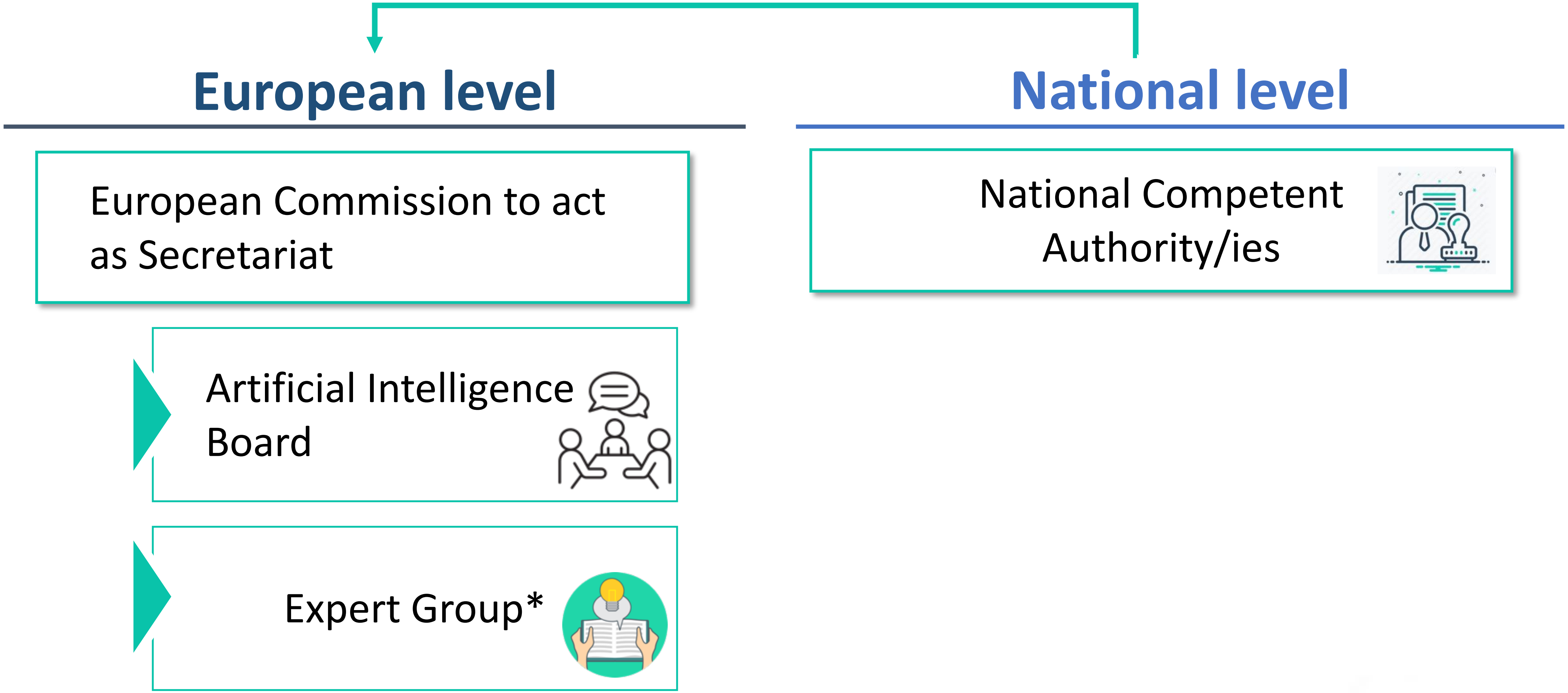
Supporting innovation (Title V)



Regulatory sandboxes Art. 53 and 54

Support for SMEs/start-ups Art. 55

The governance structure (Titles VI and VII)



European level

European Commission to act as Secretariat

Artificial Intelligence Board

Expert Group*

National level

National Competent Authority/ies

*Not foreseen in the regulation but the Commission intends to introduce it in the implementation process

Thank you

Gregor Strojin

si2021.eu
Slovensko predsedovanje Svetu Evropske unije
Slovenian Presidency of the Council of the European Union

REGULATION OF ARTIFICIAL INTELLIGENCE

ETHICAL AND FUNDAMENTAL RIGHTS ASPECTS
EUROPEAN UNION AND INTERNATIONAL PERSPECTIVE

CAHAI (CoE)

July 20, 2021

Gregor Strojin

CAHAI (Council of Europe) - Chair
gregor.strojin@gmail.com

CAHAI - MANDATE

Under the authority of the Committee of Ministers, the CAHAI is instructed to:

- *examine the **feasibility** and **potential elements** on the basis of broad multi-stakeholder consultations, of a legal framework for the development, design and application of artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law.*

When fulfilling this task, the Ad hoc Committee shall:

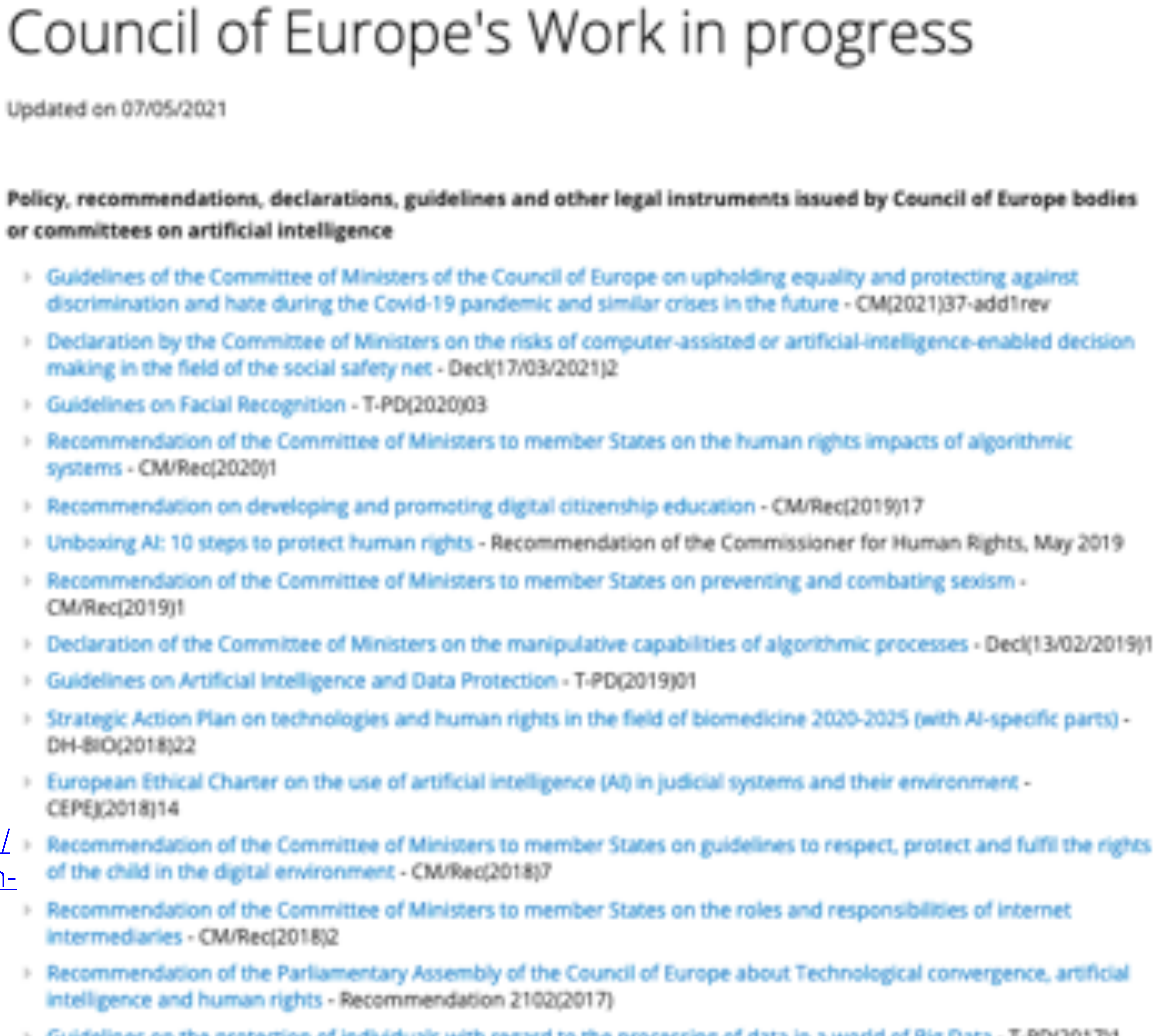
- *take into account the standards of the Council of Europe relevant to the design, development and application of digital technologies, in the fields of human rights, democracy and the rule of law, in particular on the basis of existing legal instruments;*

- *take into account relevant existing universal and regional international legal instruments, work undertaken by other Council of Europe bodies as well as ongoing work in other international and regional organisations;*

- *take due account of a gender perspective, building cohesive societies and promoting and protecting rights of persons with disabilities in the performance of its tasks.*

https://www.coe.int/en/web/artificial-intelligence/cahai

OR



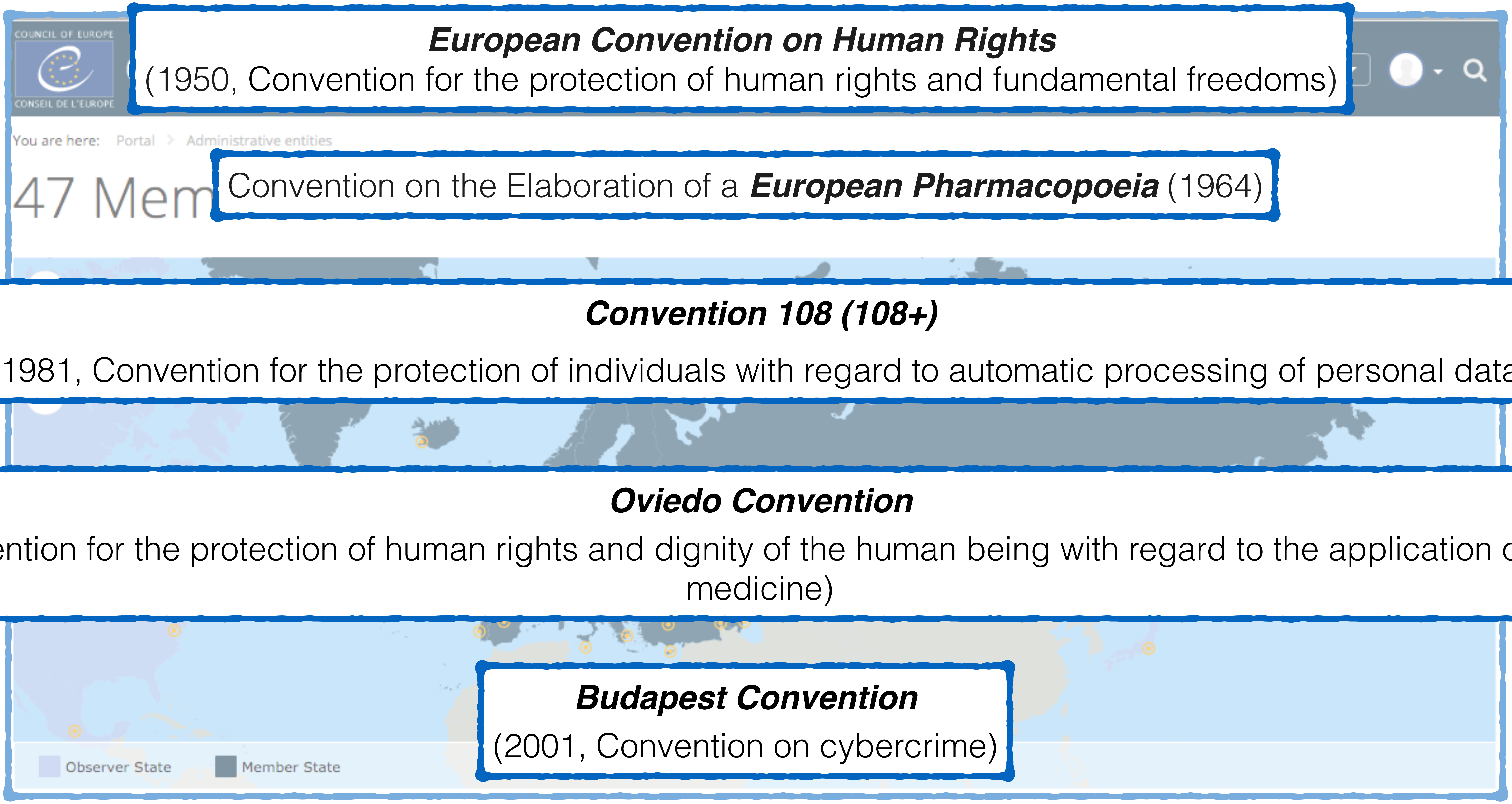
Council of Europe's Work in progress

Updated on 07/05/2021

Policy, recommendations, declarations, guidelines and other legal instruments issued by Council of Europe bodies or committees on artificial intelligence

- Guidelines of the Committee of Ministers of the Council of Europe on upholding equality and protecting against discrimination and hate during the Covid-19 pandemic and similar crises in the future - CM(2021)37-add1rev
- Declaration by the Committee of Ministers on the risks of computer-assisted or artificial-intelligence-enabled decision making in the field of the social safety net - Decl(17/03/2021)2
- Guidelines on Facial Recognition - T-PD(2020)03
- Recommendation of the Committee of Ministers to member States on the human rights impacts of algorithmic systems - CM/Rec(2020)1
- Recommendation on developing and promoting digital citizenship education - CM/Rec(2019)17
- Unboxing AI: 10 steps to protect human rights - Recommendation of the Commissioner for Human Rights, May 2019
- Recommendation of the Committee of Ministers to member States on preventing and combating sexism - CM/Rec(2019)1
- Declaration of the Committee of Ministers on the manipulative capabilities of algorithmic processes - Decl(13/02/2019)1
- Guidelines on Artificial Intelligence and Data Protection - T-PD(2019)01
- Strategic Action Plan on technologies and human rights in the field of biomedicine 2020-2025 (with AI-specific parts) - DH-BIO(2018)22
- European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment - CEPEJ(2018)14
- Recommendation of the Committee of Ministers to member States on guidelines to respect, protect and fulfil the rights of the child in the digital environment - CM/Rec(2018)7
- Recommendation of the Committee of Ministers to member States on the roles and responsibilities of internet intermediaries - CM/Rec(2018)2
- Recommendation of the Parliamentary Assembly of the Council of Europe about Technological convergence, artificial intelligence and human rights - Recommendation 2102(2017)
- Guidelines on the protection of individuals with regard to the processing of data in a world of Big Data - T-PD(2017)1

https://www.coe.int/en/web/artificial-intelligence/work-in-progress



European Convention on Human Rights
(1950, Convention for the protection of human rights and fundamental freedoms)

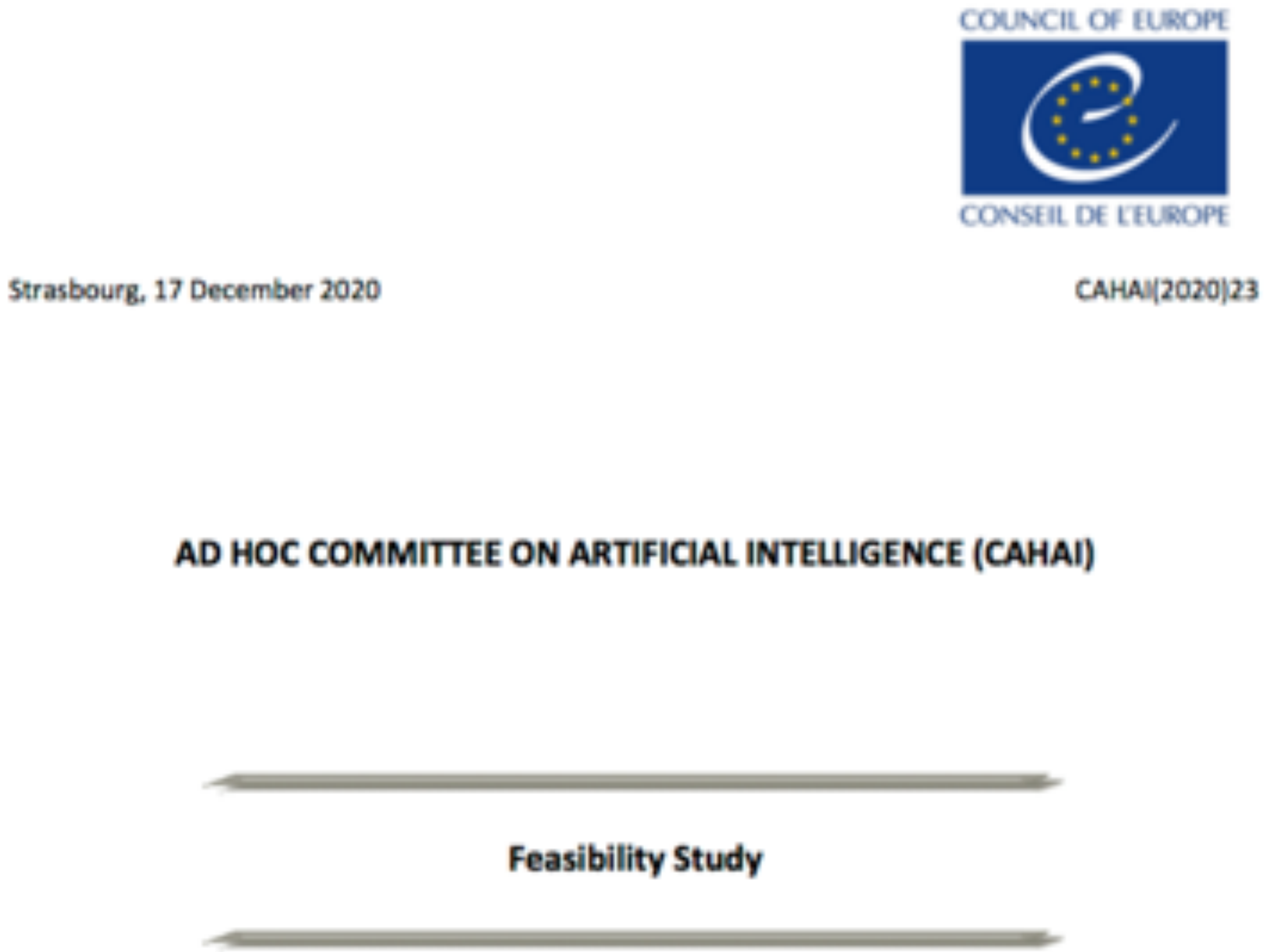
Convention on the Elaboration of a ***European Pharmacopoeia*** (1964)

Convention 108 (108+)

(1981, Convention for the protection of individuals with regard to automatic processing of personal data)

Oviedo Convention
(1997, Convention for the protection of human rights and dignity of the human being with regard to the application of biology and medicine)

Budapest Convention
(2001, Convention on cybercrime)



COUNCIL OF EUROPE

CONSEIL DE L'EUROPE

Strasbourg, 17 December 2020

CAHAI(2020)23

AD HOC COMMITTEE ON ARTIFICIAL INTELLIGENCE (CAHAI)

Feasibility Study

https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

OR

No legal vacuum, but …

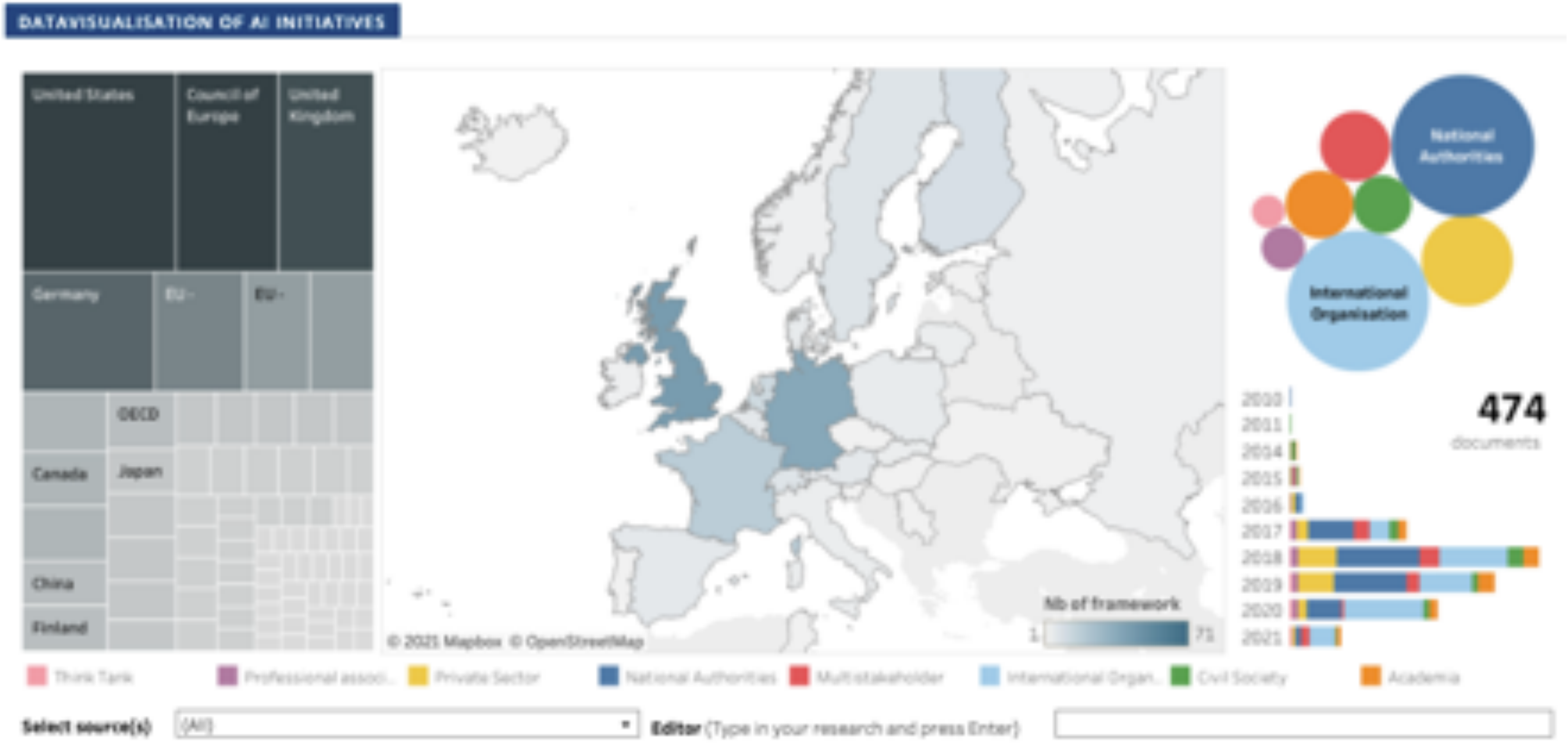
(see chapters 3 & 5)

- Substantive and procedural gaps
- Uneven protection levels
- Uncertainties affect development and implementation
- Soft law approach has major limitations

https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da

OR

AI initiatives



https://www.coe.int/en/web/artificial-intelligence/national-initiatives

OR

KEY VALUES, RIGHTS AND PRINCIPLES (chapter 7)

- Human **dignity**
- **Prevention of harm** to human rights, democracy and the rule of law
- Human **freedom** and Human **autonomy**
- **Non-Discrimination**, **Gender equality**, **Fairness** and **Diversity**
- **Transparency** and **Explainability** of AI systems
- **Data protection** and the right to **privacy**
- **Accountability** and **responsibility**
- **Democracy**
- **Rule of Law**

7.1.1. Human Dignity *(example)*

Key substantive rights:

- The right to **human dignity**, the right to **life** (Art. 2 ECHR), and the right to **physical and mental integrity**.
- The **right to be informed** of the fact that one is interacting with an AI system rather than with a human being, in particular when the risk of confusion arises and can affect human dignity.
- The **right to refuse interaction** with an AI system whenever this can adversely impact human dignity.

Key obligations:

- Member States should ensure that, **where tasks risk violating human dignity** if carried out by machines rather than human beings, these **tasks are reserved for humans**.
- Member States should require AI deployers to **inform human beings** of the fact that they are interacting with an AI system rather than with a human being whenever confusion may arise

APPROPRIATE LEGAL FRAMEWORK (1/2)

An **appropriate legal framework** will likely consist of a **combination of binding and non-binding** legal instruments, that complement each other.

A binding instrument, a convention or framework convention, of horizontal character, could **consolidate general common principles** – contextualised to apply to the AI environment and using a risk-based approach – and include more granular provisions in line with the rights, principles and obligations identified in this feasibility study.

Any binding document, whatever its shape, should not be overly prescriptive so as to secure its **future-proof** nature. Moreover, it should ensure that **socially beneficial AI innovation can flourish**, all the while **adequately tackling the specific risks** posed by the design, development and application of AI systems.

APPROPRIATE LEGAL FRAMEWORK (2/2)

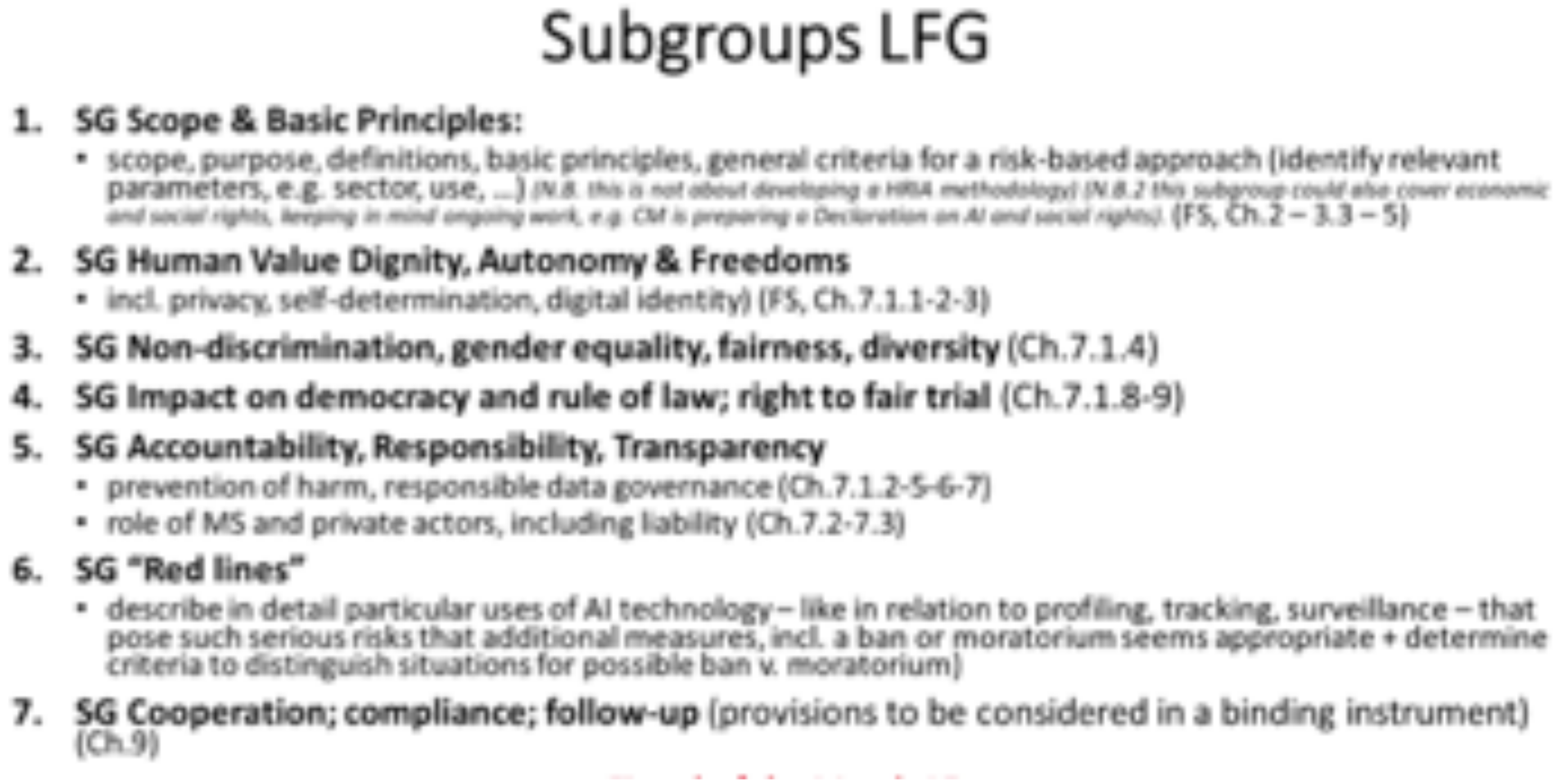
This instrument could be combined with additional binding or non-binding **sectoral Council of Europe instruments** to address challenges brought by AI systems in specific sectors.

This **combination** would also allow **legal certainty** for AI stakeholders to be enhanced, and provide the required legal **guidance to private actors** wishing to undertake **self-regulatory** initiatives.

Moreover, by establishing **common norms at an international level**, **transboundary trust** in AI products and services would be ensured, thereby guaranteeing that the benefits generated by AI systems can travel across national borders.

It is important that any legal framework includes **practical mechanisms to mitigate risks** arising from AI systems, as well as appropriate **follow-up mechanisms** and processes and measures for international co-operation.

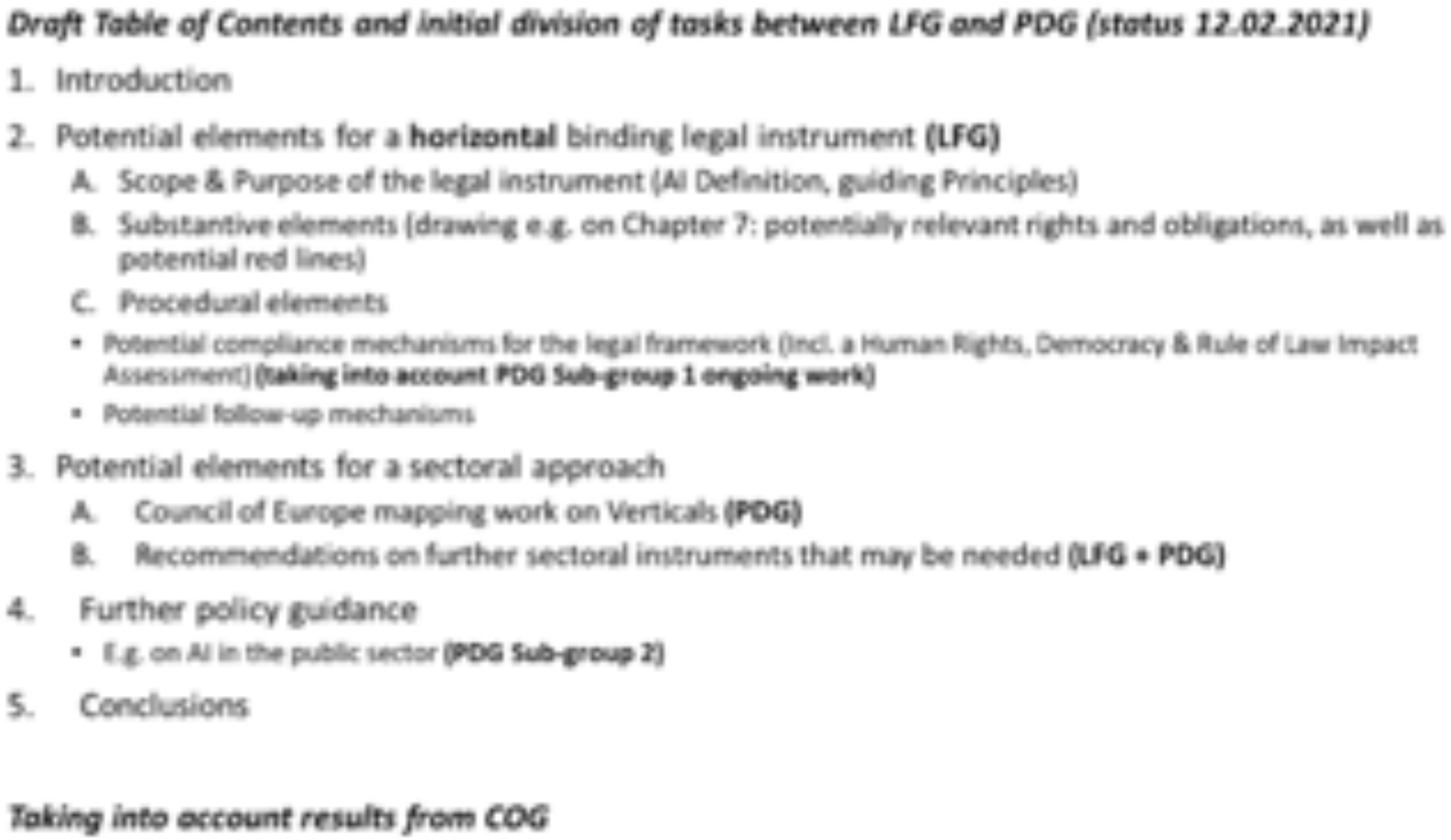
LFG - internal division of work



Subgroups LFG

1. **SG Scope & Basic Principles:**
   - scope, purpose, definitions, basic principles, general criteria for a risk-based approach (identify relevant parameters, e.g. sector, use, …) *(N.B. this is not about developing a HRIA methodology) (N.B.2 this subgroup could also cover economic and social rights, keeping in mind ongoing work, e.g. CM is preparing a Declaration on AI and social rights).* (FS, Ch.2 – 3.3 – 5)
2. **SG Human Value Dignity, Autonomy & Freedoms**
   - incl. privacy, self-determination, digital identity) (FS, Ch.7.1.1-2-3)
3. **SG Non-discrimination, gender equality, fairness, diversity** (Ch.7.1.4)
4. **SG Impact on democracy and rule of law; right to fair trial** (Ch.7.1.8-9)
5. **SG Accountability, Responsibility, Transparency**
   - prevention of harm, responsible data governance (Ch.7.1.2-5-6-7)
   - role of MS and private actors, including liability (Ch.7.2-7.3)
6. **SG "Red lines"**
   - describe in detail particular uses of AI technology – like in relation to profiling, tracking, surveillance – that pose such serious risks that additional measures, incl. a ban or moratorium seems appropriate + determine criteria to distinguish situations for possible ban v. moratorium)
7. **SG Cooperation; compliance; follow-up** (provisions to be considered in a binding instrument) (Ch.9)

LFG and PDG in 2021



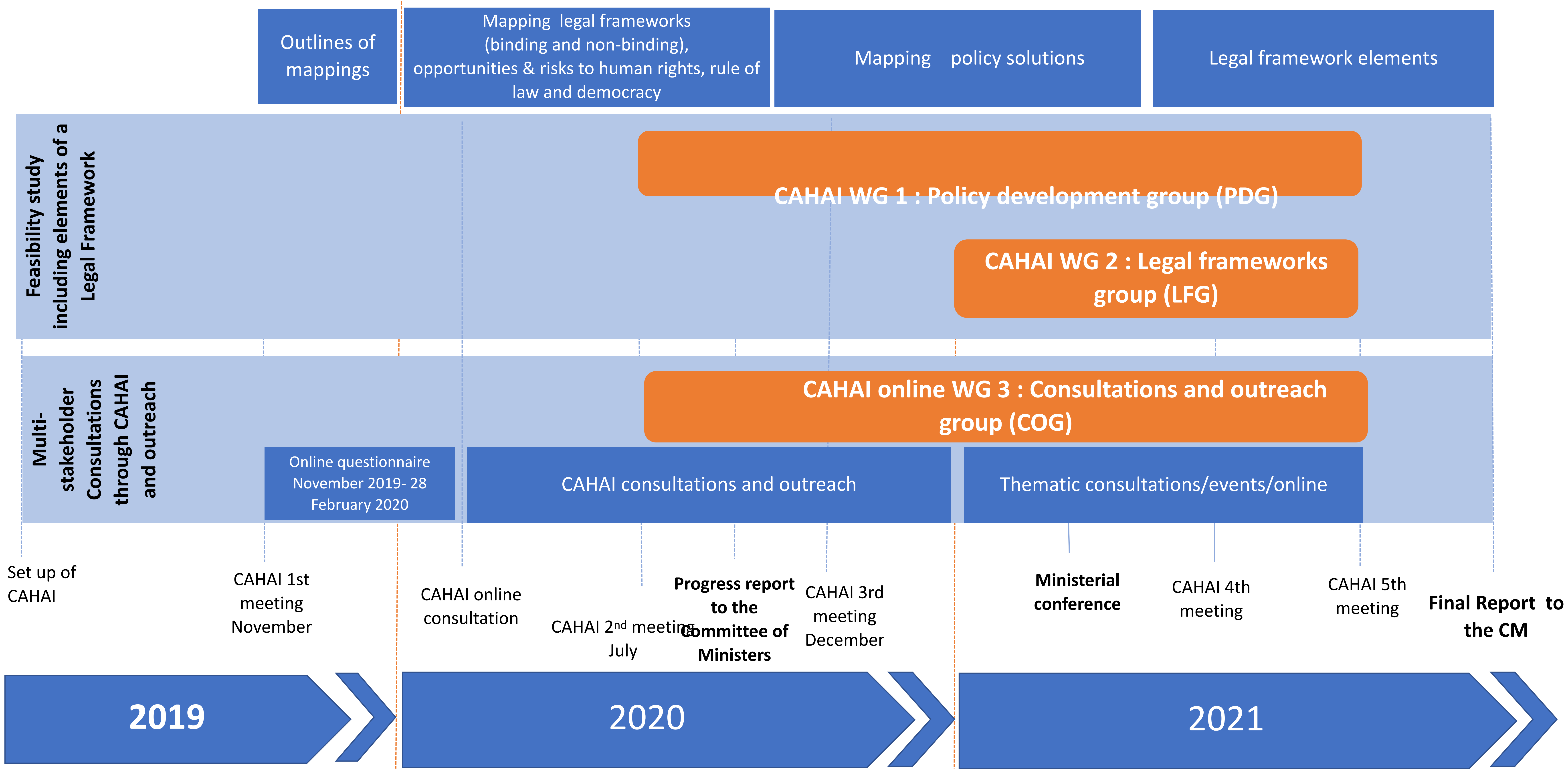
Draft Table of Contents and initial division of tasks between LFG and PDG (status 12.02.2021)

1. Introduction
2. Potential elements for a **horizontal** binding legal instrument **(LFG)**
   - A. Scope & Purpose of the legal instrument (AI Definition, guiding Principles)
   - B. Substantive elements (drawing e.g. on Chapter 7: potentially relevant rights and obligations, as well as potential red lines)
   - C. Procedural elements
   - Potential compliance mechanisms for the legal framework (Incl. a Human Rights, Democracy & Rule of Law Impact Assessment) **(taking into account PDG Sub-group 1 ongoing work)**
   - Potential follow-up mechanisms
3. Potential elements for a sectoral approach
   - A. Council of Europe mapping work on Verticals **(PDG)**
   - B. Recommendations on further sectoral instruments that may be needed **(LFG + PDG)**
4. Further policy guidance
   - E.g. on AI in the public sector **(PDG Sub-group 2)**
5. Conclusions

Taking into account results from COG

CAHAI - ROADMAP

Key deliverables and proposed roadmap of CAHAI (2019 –2021)



Outlines of mappings

Mapping legal frameworks (binding and non-binding), opportunities & risks to human rights, rule of law and democracy

Mapping policy solutions

Legal framework elements

Feasibility study including elements of a Legal Framework

CAHAI WG 1 : Policy development group (PDG)

CAHAI WG 2 : Legal frameworks group (LFG)

Multi-stakeholder Consultations through CAHAI and outreach

CAHAI online WG 3 : Consultations and outreach group (COG)

Online questionnaire November 2019- 28 February 2020

CAHAI consultations and outreach

Thematic consultations/events/online

Set up of CAHAI

CAHAI 1st meeting November

CAHAI online consultation

CAHAI 2nd meeting July

Progress report to the Committee of Ministers

CAHAI 3rd meeting December

Ministerial conference

CAHAI 4th meeting

CAHAI 5th meeting

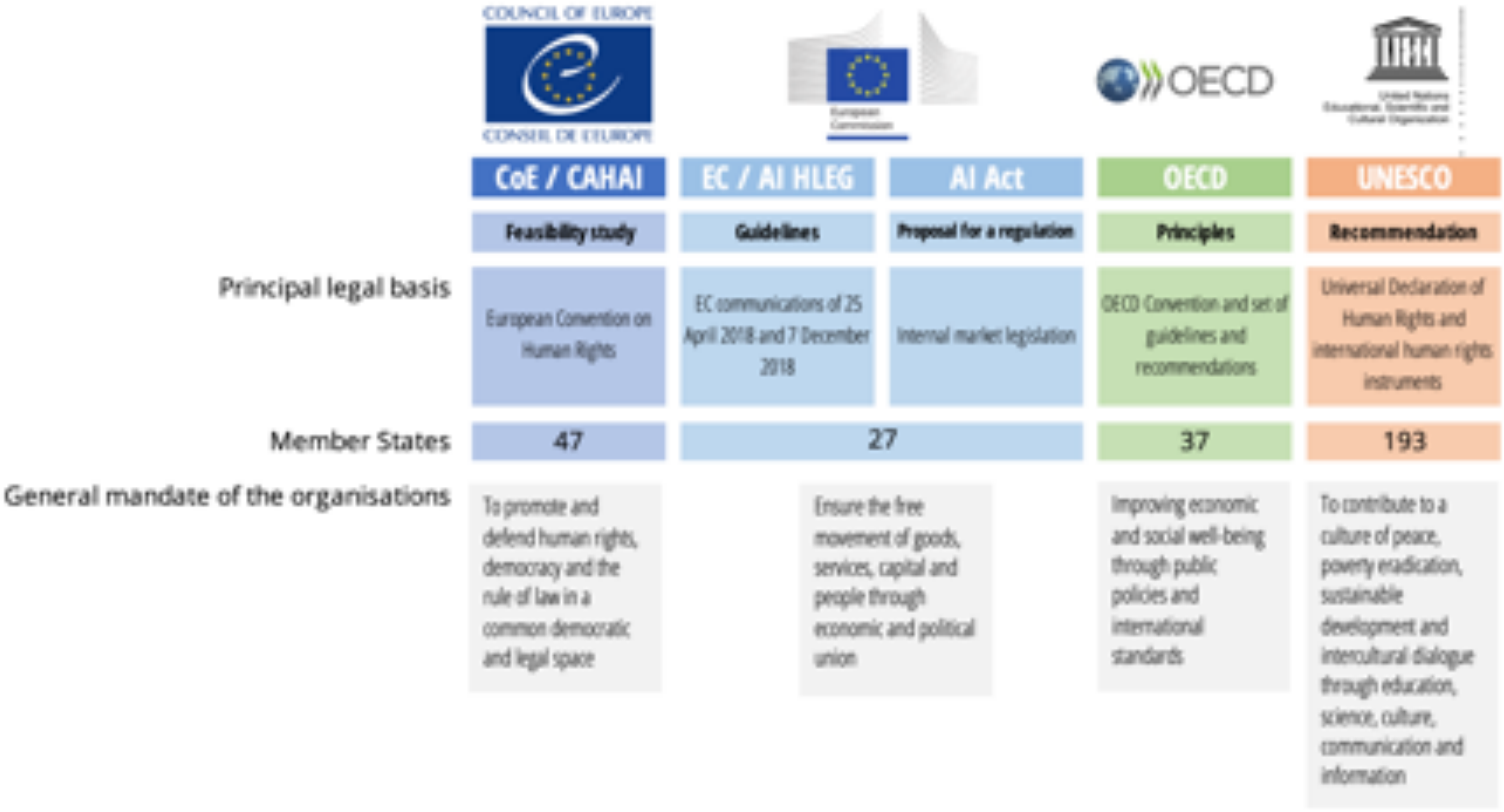
Final Report to the CM

2019

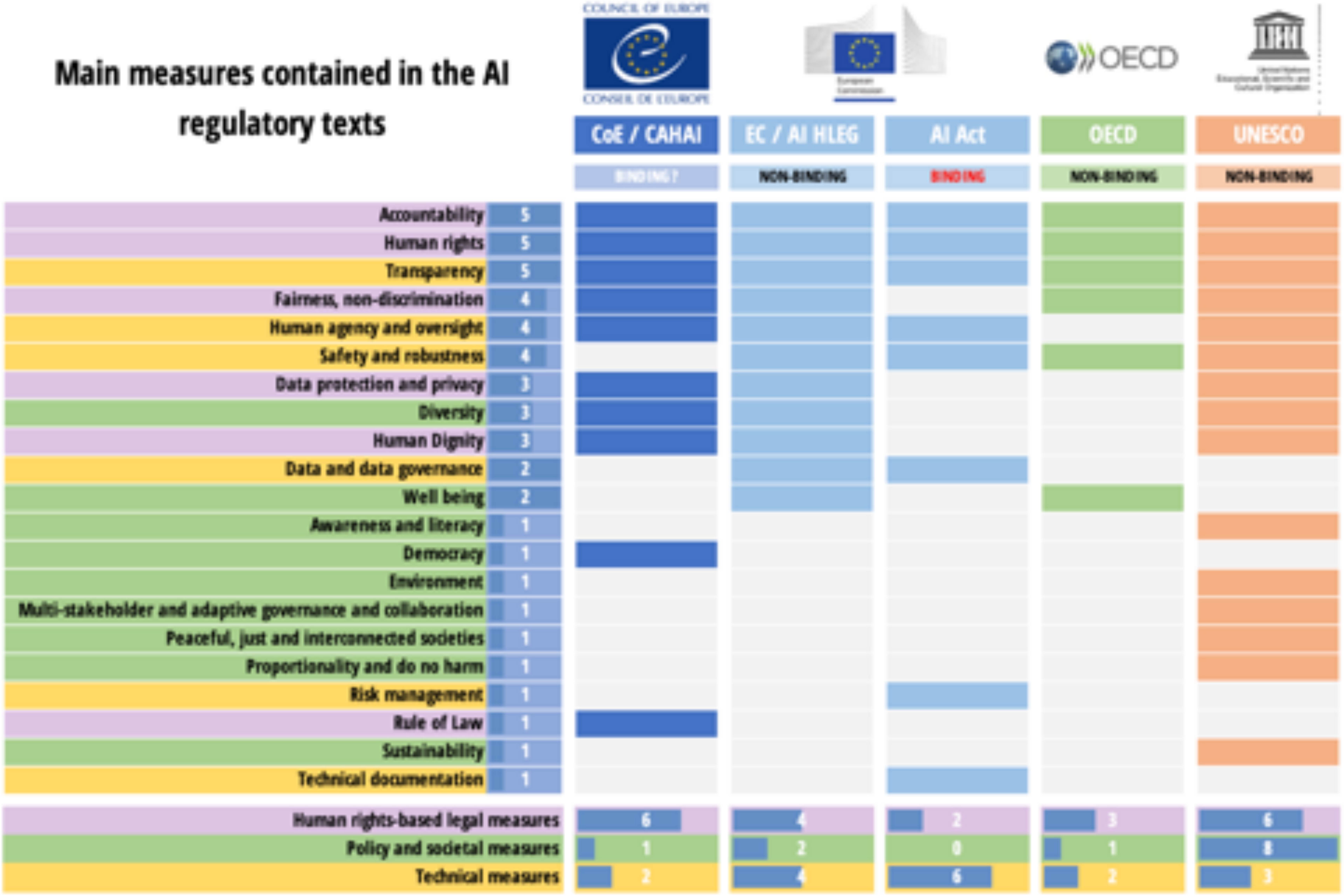
2020

2021

REGULATION OF AI
General missions of intergovernmental organisations


COUNCIL OF EUROPE
CONSEIL DE L'EUROPE
European Commission
OECD
United Nations Educational, Scientific and Cultural Organization

	CoE / CAHAI	EC / AI HLEG	AI Act	OECD	UNESCO
	Feasibility study	Guidelines	Proposal for a regulation	Principles	Recommendation
Principal legal basis	European Convention on Human Rights	EC communications of 25 April 2018 and 7 December 2018	Internal market legislation	OECD Convention and set of guidelines and recommendations	Universal Declaration of Human Rights and international human rights instruments
Member States	47	27		37	193
General mandate of the organisations	To promote and defend human rights, democracy and the rule of law in a common democratic and legal space	Ensure the free movement of goods, services, capital and people through economic and political union		Improving economic and social well-being through public policies and international standards	To contribute to a culture of peace, poverty eradication, sustainable development and intercultural dialogue through education, science, culture, communication and information



Main measures contained in the AI regulatory texts

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE
OECD

		CoE / CAHAI	EC / AI HLEG	AI Act	OECD	UNESCO
		BINDING?	NON-BINDING	BINDING	NON-BINDING	NON-BINDING
Accountability	5	■	■	■	■	■
Human rights	5	■	■	■	■	■
Transparency	5	■	■	■	■	■
Fairness, non-discrimination	4	■	■		■	■
Human agency and oversight	4	■	■	■		■
Safety and robustness	4		■	■	■	■
Data protection and privacy	3	■	■			■
Diversity	3	■	■			■
Human Dignity	3	■	■			■
Data and data governance	2		■	■		
Well being	2		■		■	
Awareness and literacy	1					■
Democracy	1	■				
Environment	1					■
Multi-stakeholder and adaptive governance and collaboration	1					■
Peaceful, just and interconnected societies	1					■
Proportionality and do no harm	1					■
Risk management	1			■		
Rule of Law	1	■				
Sustainability	1					■
Technical documentation	1			■		
Human rights-based legal measures		6	4	2	3	6
Policy and societal measures		1	2	0	1	8
Technical measures		2	4	6	2	3

Dr. David Leslie

The Alan Turing Institute

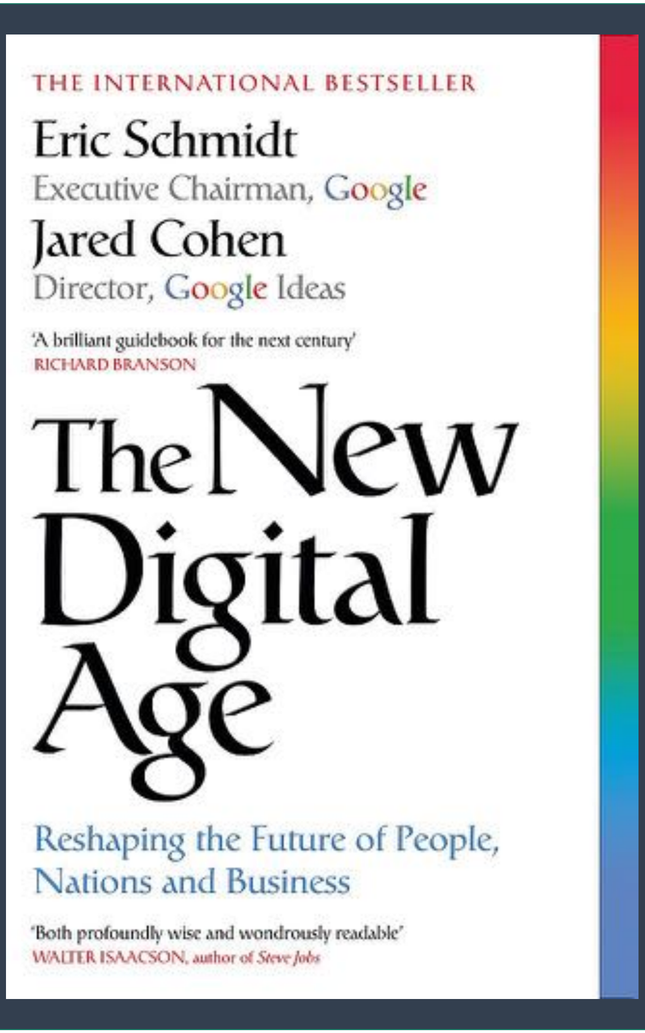
Dr. David Leslie, Ethics Lead

AI and the Civilizational Tipping Point: *Finding our way with fundamental rights and freedoms*

Are we at a tipping point?

A glance back at ancient history:

> 'As in a social contract,' 'users will voluntarily relinquish things they value in the physical world—privacy, security, personal data—in order to gain the benefits that come with being connected to the virtual world.'
>
> (Schmidt and Cohen, 2013)



THE INTERNATIONAL BESTSELLER

Eric Schmidt
Executive Chairman, Google
Jared Cohen
Director, Google Ideas

'A brilliant guidebook for the next century'
RICHARD BRANSON

The New Digital Age

Reshaping the Future of People, Nations and Business

'Both profoundly wise and wondrously readable'
WALTER ISAACSON, author of *Steve Jobs*

Are we at a tipping point?

Coming ubiquity of “intelligent” cyber-physical systems

From the internet of things…

Are we at a tipping point?

Coming ubiquity of “intelligent” cyber-physical systems

To the internet of bodies…

Are we at a tipping point?

Coming ubiquity of “intelligent” cyber-physical systems

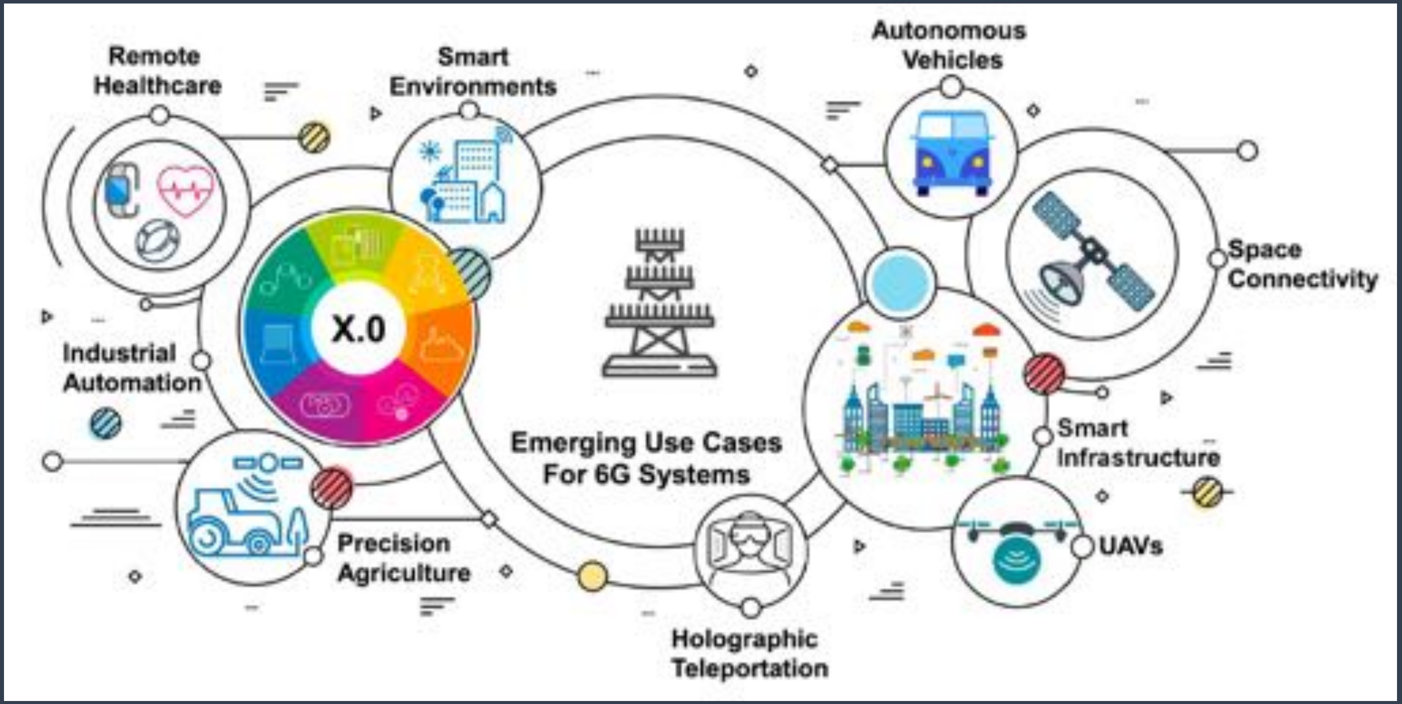


Image from: (Akyildyz et al., 2020)

To the internet of everything…

Are we at a tipping point?

Coming ubiquity of “intelligent” cyber-physical systems

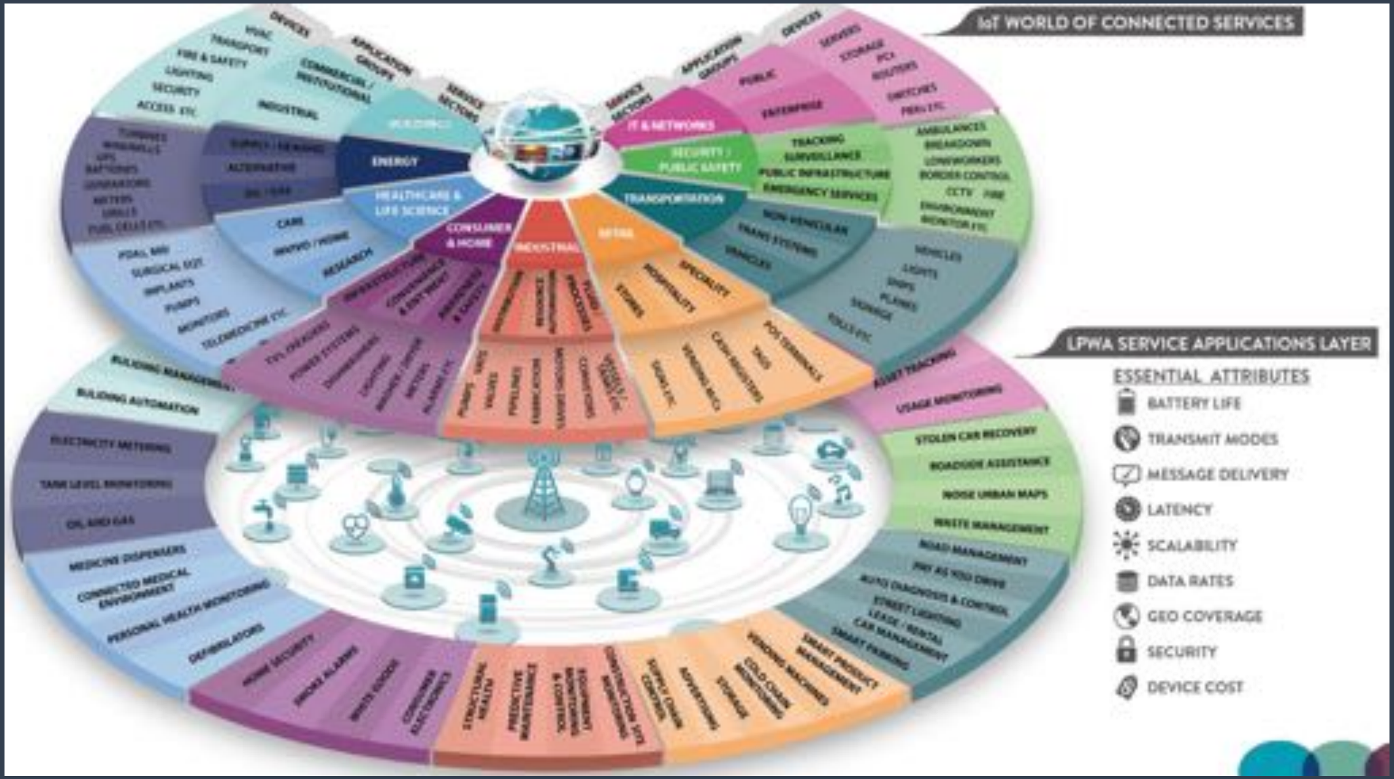
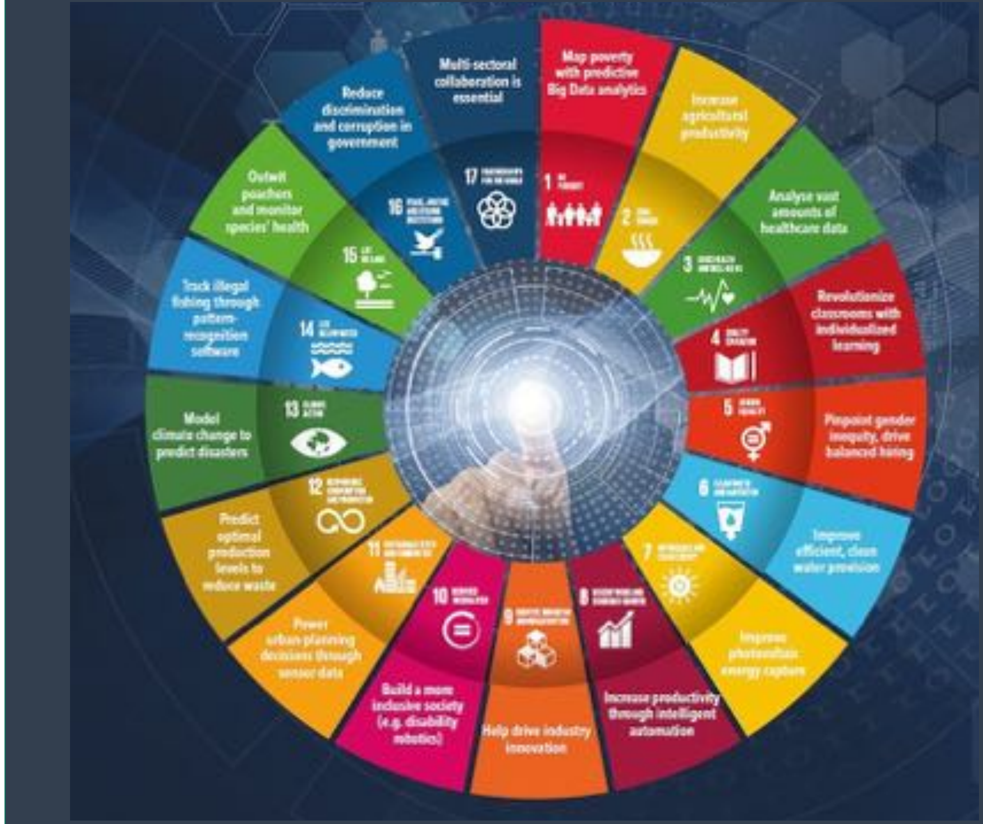


Image from: (Beecham, 2020)

To the internet of everything…

What shape will the digital society of tomorrow take?

Summum bonum…


Map poverty with predictive Big Data analytics
Increase agricultural productivity
Analyse vast amounts of healthcare data
Revolutionize classrooms with individualized learning
Pinpoint gender inequity, drive balanced hiring
Improve efficient, clean water provision
Improve photovoltaic energy capture
Increase productivity through intelligent automation
Help drive industry innovation
Build a more inclusive society (e.g. disability robotics)
Power urban planning decisions through sensor data
Predict optimal production levels to reduce waste
Model climate change to predict disasters
Track illegal fishing through pattern-recognition software
Outwit poachers and monitor species' health
Reduce discrimination and corruption in government
Multi-sectoral collaboration is essential
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17

Or Summum malum

Ethical hazards of pervasive AI and the IoE:

Loss of agency and social connection

There are potentially dehumanising consequences of integrating AI into ubiquitous cyber-physical systems. **Individuals may be disempowered and feel like they have been manipulated or 'reduced to statistics.' Crucial human connection, trust and empathy may be lost through automation and curation.**

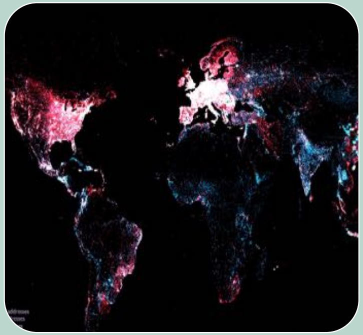
Harmful & poor-quality outcomes

Algorithmic models are only as good as the data on which they are trained and tested ('Garbage in, garbage out'). Inaccuracies and measurement errors across data collection and recording can taint datasets. This is intensified in pervasive sensor monitoring. Using poor quality data may have grave consequence for individual wellbeing and the public welfare.

Entrenched Bias and discrimination

Drawing insights from existing data distributions, supervised machine learning models, when they work reliably, make accurate out-of-sample predictions by replicating the social and cultural patterns of the past—regardless of whether these patterns are **inequitable or discriminatory**. Ubiquitous analytics of social data may augment discrimination and structural injustices.

Ethical hazards of pervasive AI and the IoE:



Widening global & local digital divides

Uneven global and domestic distribution of access to and the benefits of pervasive AI promises a hyper-exacerbation of extant dynamics of societal inequality. Infrastructural requirements for balanced progress in the distribution of intelligent cyber-physical systems demand a level of social equality orders of magnitude greater than currently exists.

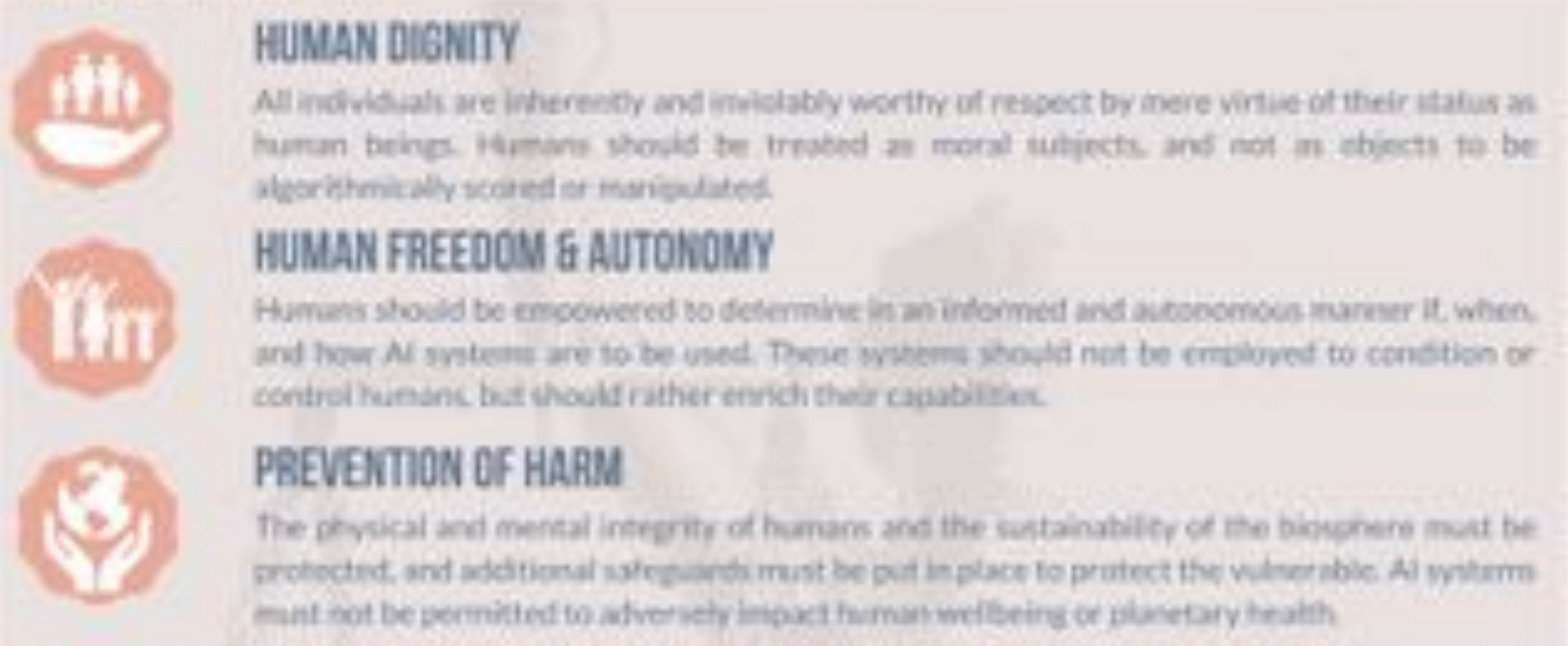
Data integrity, privacy and security

With the multiplication of sites of behavioural, social and environmental measurement processing on low energy networked devices, issues of data integrity and infrastructural security will intensify in kind. This will escalate risks of hacking at scale, cyber-terrorism, and privacy violation and magnify their consequences.

Biospheric harm

The environmental costs of mass, real-time information processing and AI/ML system training are potentially prohibitive. A connected IoT, where large-scale industrial, agricultural, transportation, health and infrastructural processes and products are smartified will pose risks to biospheric sustainability by virtue of the magnification of energy consumption.

Responding with principles, finding our way with fundamental rights and freedoms:



HUMAN DIGNITY

All individuals are inherently and inviolably worthy of respect by mere virtue of their status as human beings. Humans should be treated as moral subjects, and not as objects to be algorithmically scored or manipulated.

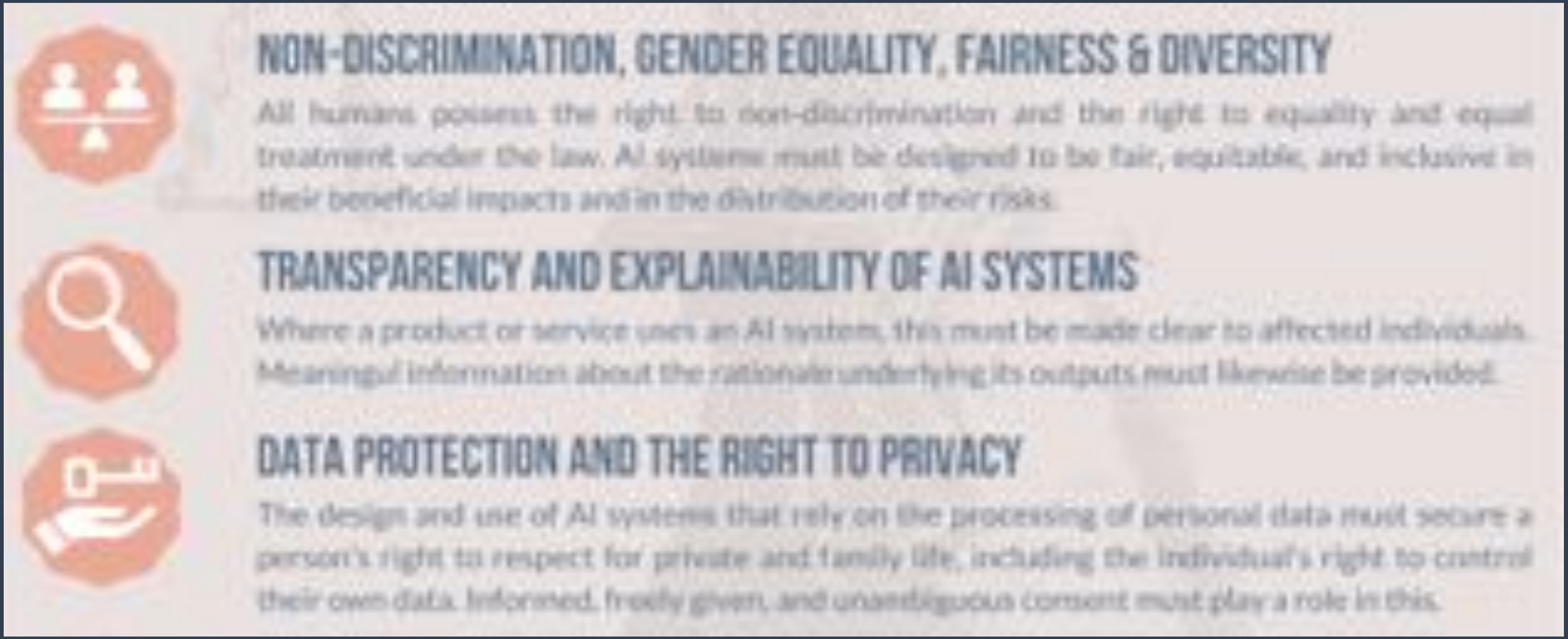
HUMAN FREEDOM & AUTONOMY

Humans should be empowered to determine in an informed and autonomous manner if, when, and how AI systems are to be used. These systems should not be employed to condition or control humans, but should rather enrich their capabilities.

PREVENTION OF HARM

The physical and mental integrity of humans and the sustainability of the biosphere must be protected, and additional safeguards must be put in place to protect the vulnerable. AI systems must not be permitted to adversely impact human wellbeing or planetary health.

Responding with principles, finding our way with fundamental rights and freedoms:



NON-DISCRIMINATION, GENDER EQUALITY, FAIRNESS & DIVERSITY

All humans possess the right to non-discrimination and the right to equality and equal treatment under the law. AI systems must be designed to be fair, equitable, and inclusive in their beneficial impacts and in the distribution of their risks.

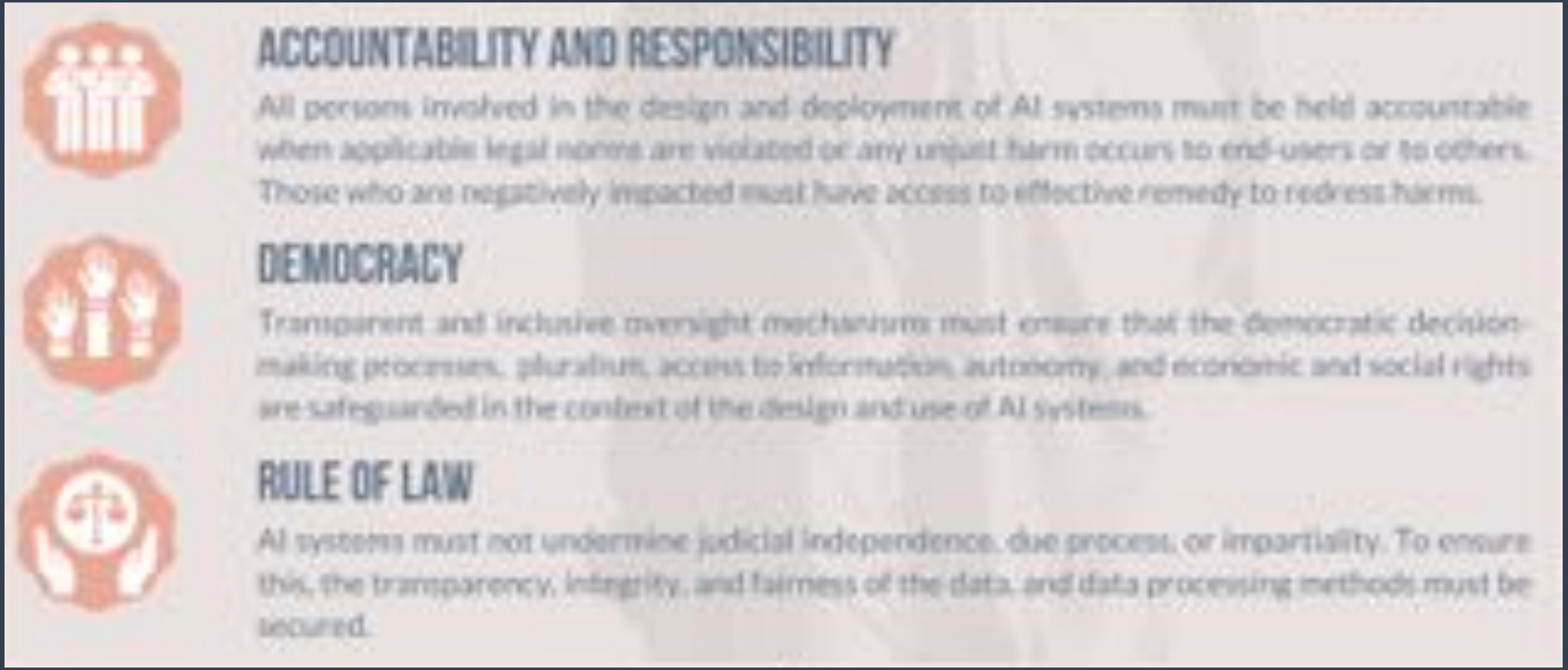
TRANSPARENCY AND EXPLAINABILITY OF AI SYSTEMS

Where a product or service uses an AI system, this must be made clear to affected individuals. Meaningful information about the rationale underlying its outputs must likewise be provided.

DATA PROTECTION AND THE RIGHT TO PRIVACY

The design and use of AI systems that rely on the processing of personal data must secure a person's right to respect for private and family life, including the individual's right to control their own data. Informed, freely given, and unambiguous consent must play a role in this.

Responding with principles, finding our way with fundamental rights and freedoms:



ACCOUNTABILITY AND RESPONSIBILITY

All persons involved in the design and deployment of AI systems must be held accountable when applicable legal norms are violated or any unjust harm occurs to end-users or to others. Those who are negatively impacted must have access to effective remedy to redress harms.

DEMOCRACY

Transparent and inclusive oversight mechanisms must ensure that the democratic decision-making processes, pluralism, access to information, autonomy, and economic and social rights are safeguarded in the context of the design and use of AI systems.

RULE OF LAW

AI systems must not undermine judicial independence, due process, or impartiality. To ensure this, the transparency, integrity, and fairness of the data, and data processing methods must be secured.

Thank you!

turing.ac.uk
@turinginst

Karine Perset

OECD.AI
Policy Observatory

OECD.AI and the OECD AI Network of Experts

From principles to practical implementation

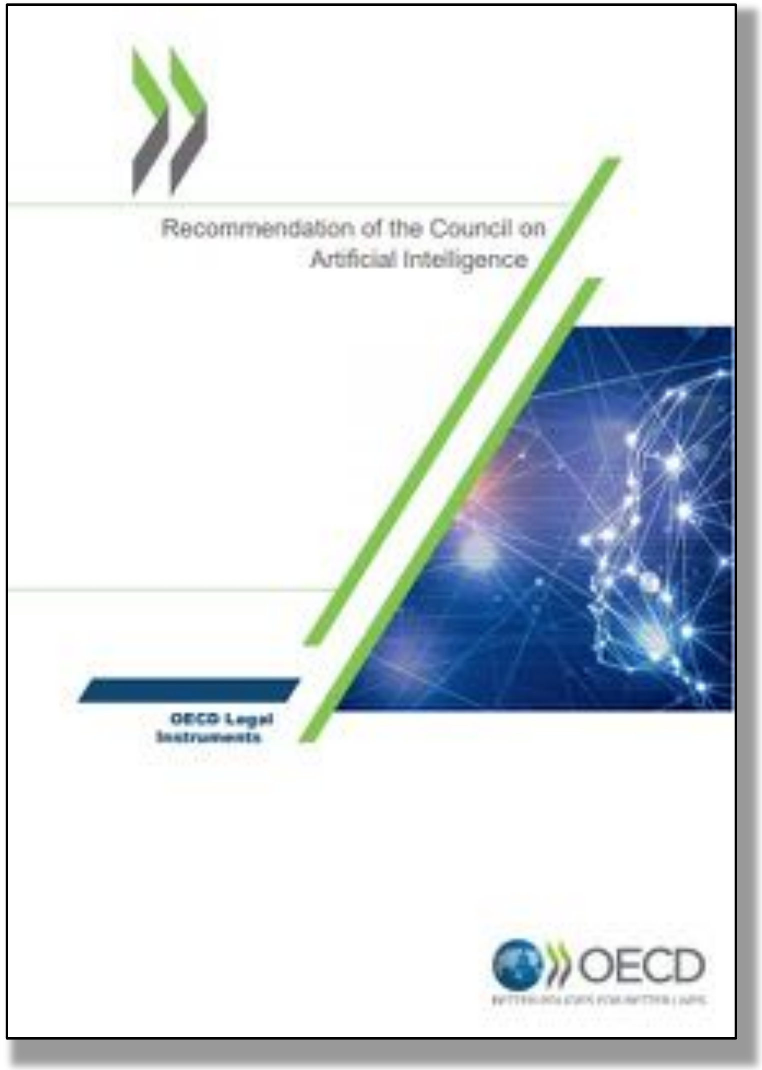
REGULATION OF ARTIFICIAL INTELLIGENCE – ETHICAL AND FUNDAMENTAL RIGHTS ASPECT
Conference of the Slovenian Presidency of the Council of the EU
20 July 2021

Karine Perset, Head of OECD.AI Policy Observatory and network of experts, OECD Digital Economy Policy Division

OECD

OECD AI Principles

- Goal: foster policy ecosystem for trustworthy AI that benefits people and planet.
- Inter-governmental standard. Adopted May 2019 by 37 OECD + 9 partner countries.
- “G20 AI Principles” in June 2019.
- Proposal for principles developed by first multi-stakeholder AI expert group @ OECD.
- Non-binding yet strong political commitment to implement & OECD monitoring.



Recommendation of the Council on Artificial Intelligence

OECD Legal Instruments

OECD

OECD AI Principles

OECD.AI Policy Observatory

10 Principles, covering two areas:

Principles for responsible stewardship of trustworthy AI

1.1. Inclusive growth, sustainable development and well-being

1.2. Human-centred values and fairness

1.3. Transparency and explainability

1.4. Robustness, security and safety

1.5. Accountability

National policies and international cooperation for trustworthy AI

2.1. Investing in AI research and development

2.2. Fostering a digital ecosystem for AI

2.3. Providing an enabling policy environment for AI

2.4. Building human capacity and preparing for labour transition

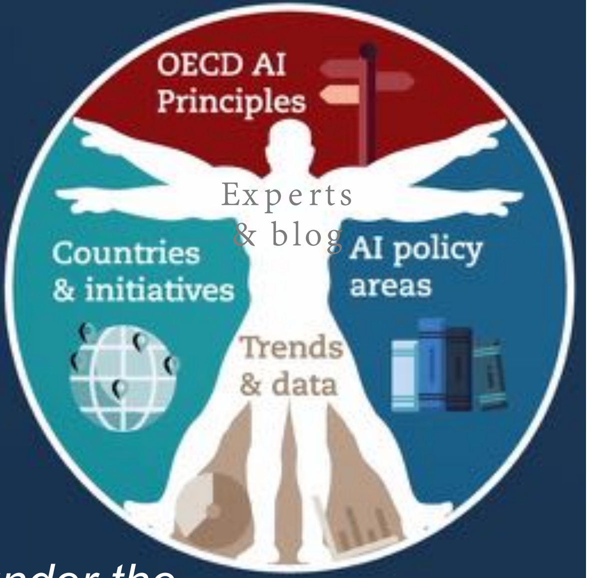
2.5. International cooperation

RESOURCES

OECD AI Policy Observatory (OECD.AI)

A platform to share & shape public policies for responsible, trustworthy & beneficial AI

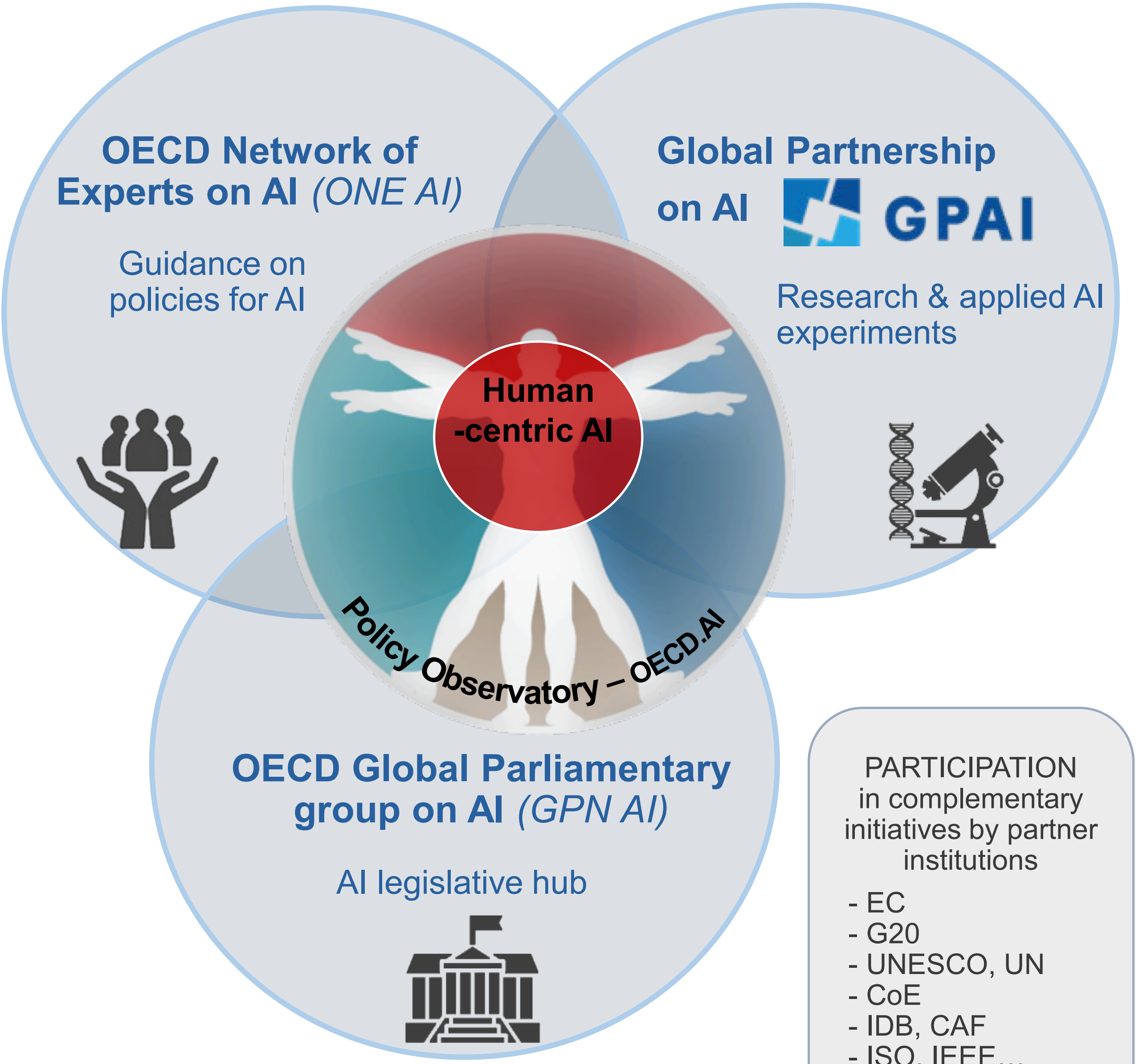
<u>5 pillars</u>:
- Network of experts and AI Wonk blog
- AI Principles & implementation
- AI policy areas
- AI trends & data, w JSI
- Country policies, w EC



GlobalPolicy.AI

Cooperation between 8 IGOs
Stay tuned for 14-15 launch under the Slovenian EU Presidency!!

EXPERTS AND DIALOGUE



OECD Network of Experts on AI *(ONE AI)*

Guidance on policies for AI

Global Partnership on AI GPAI

Research & applied AI experiments

Human -centric AI

Policy Observatory – OECD.AI

OECD Global Parliamentary group on AI *(GPN AI)*

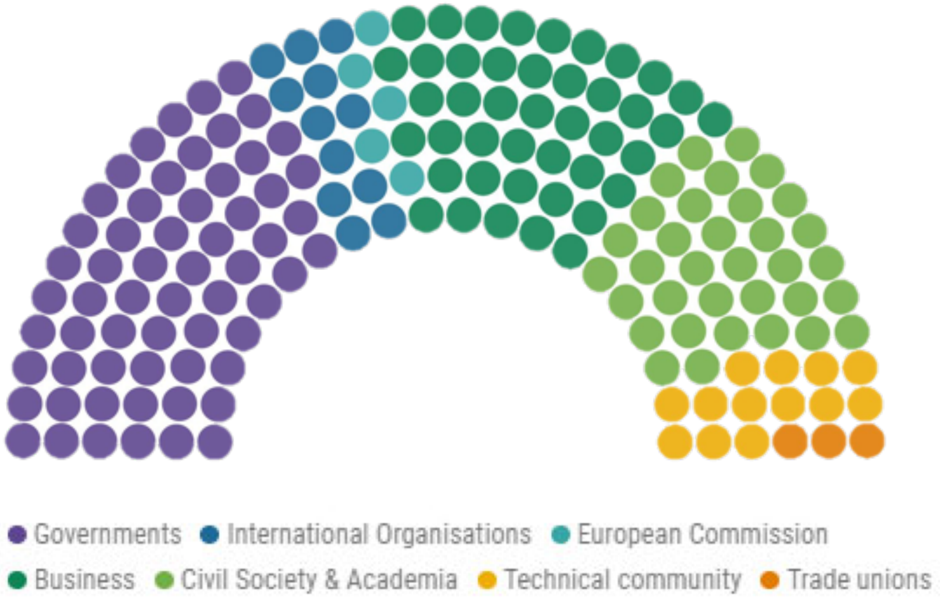
AI legislative hub

PARTICIPATION
in complementary initiatives by partner institutions
- EC
- G20
- UNESCO, UN
- CoE
- IDB, CAF
- ISO, IEEE…

OECD Network of Experts on AI (ONE AI)

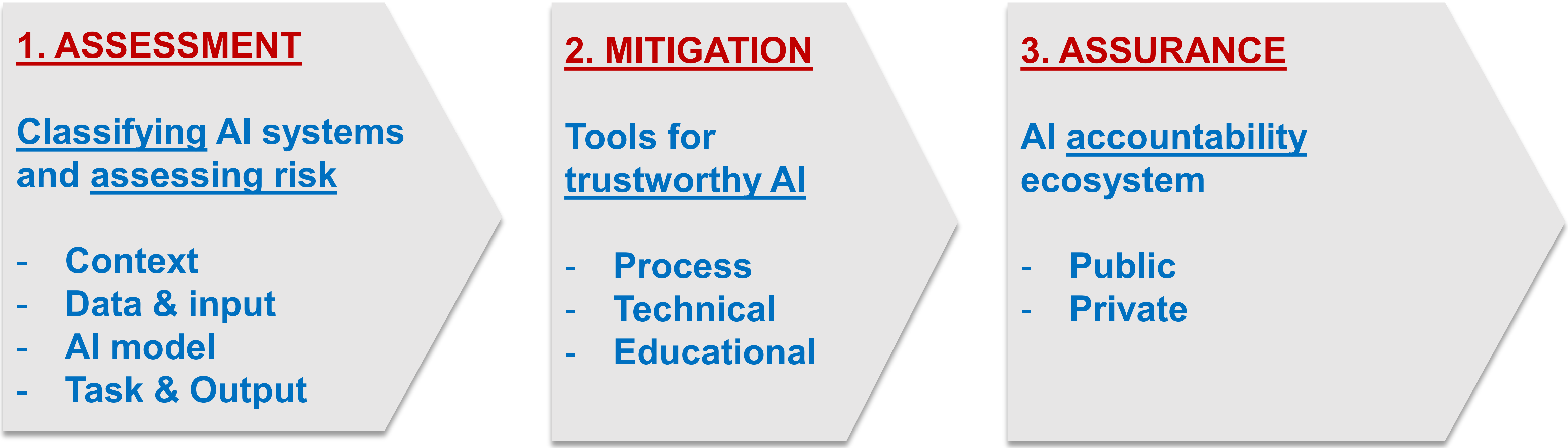
OECD.AI Policy Observatory

- OECD.AI Network of experts provides **AI-specific expertise** and advice on implementing the **OECD AI Principles**
- Launched in **February 2020**
- **200+ AI experts** from national governments, IGOs and the EC, business, civil society, academia, trade unions
- Facilitates **collaboration** between the OECD and other international initiatives on AI



Some of the focus areas of the OECD Network of Experts on AI (ONE AI)

OECD.AI Policy Observatory



1. ASSESSMENT

Classifying AI systems and assessing risk

- **Context**
- **Data & input**
- **AI model**
- **Task & Output**

2. MITIGATION

Tools for trustworthy AI

- **Process**
- **Technical**
- **Educational**

3. ASSURANCE

AI accountability ecosystem

- **Public**
- **Private**

PRINCIPLES

Values-based principles

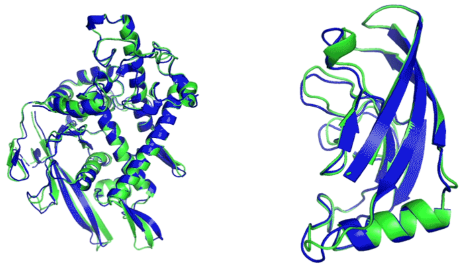
Socio-economic & environmental impacts
Human-centred values and fairness
Transparency, explainability
Robustness, security, safety
Accountability

National Policies

Investing in research
Compute, data, technologies
Enabling policy environment
Jobs, skills, transitions
International cooperation

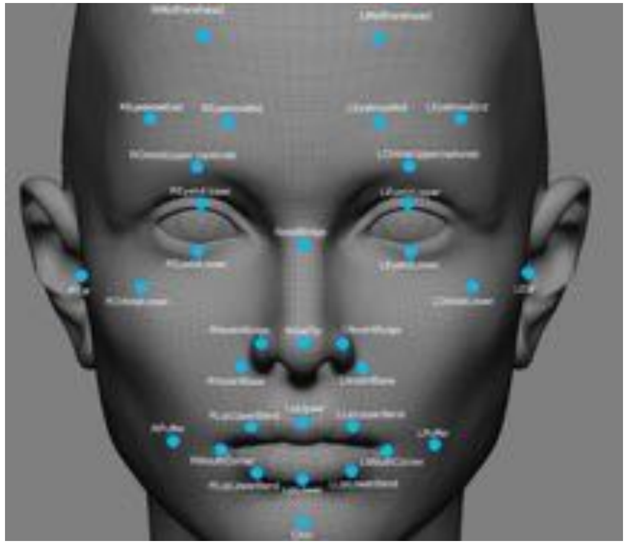
1. Assessing and classifying AI systems?

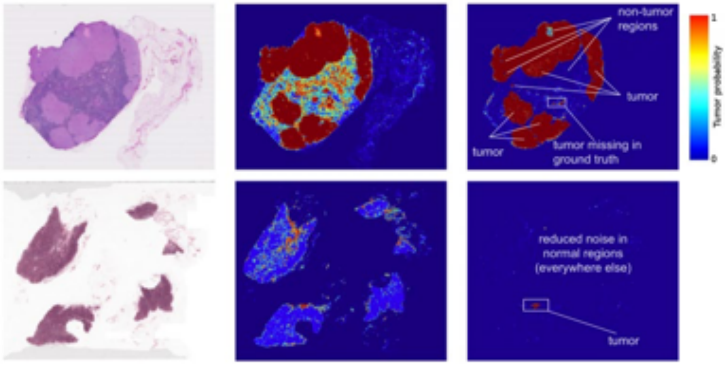
A variety of systems and policy implications


T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)

T1049 / 6y4f
93.3 GDT
(adhesin tip)

● Experimental result
● Computational predicti




non-tumor regions
tumor
tumor missing in ground truth
tumor
reduced noise in normal regions (everywhere else)
tumor
Tumor probability
1
0

OECD AI System Definition (OECD, 2019)

"An **AI system**, is a machine-based system that is capable of **influencing the environment** by producing an **output** (recommendations, predictions or decisions) for a given set of **objectives**.

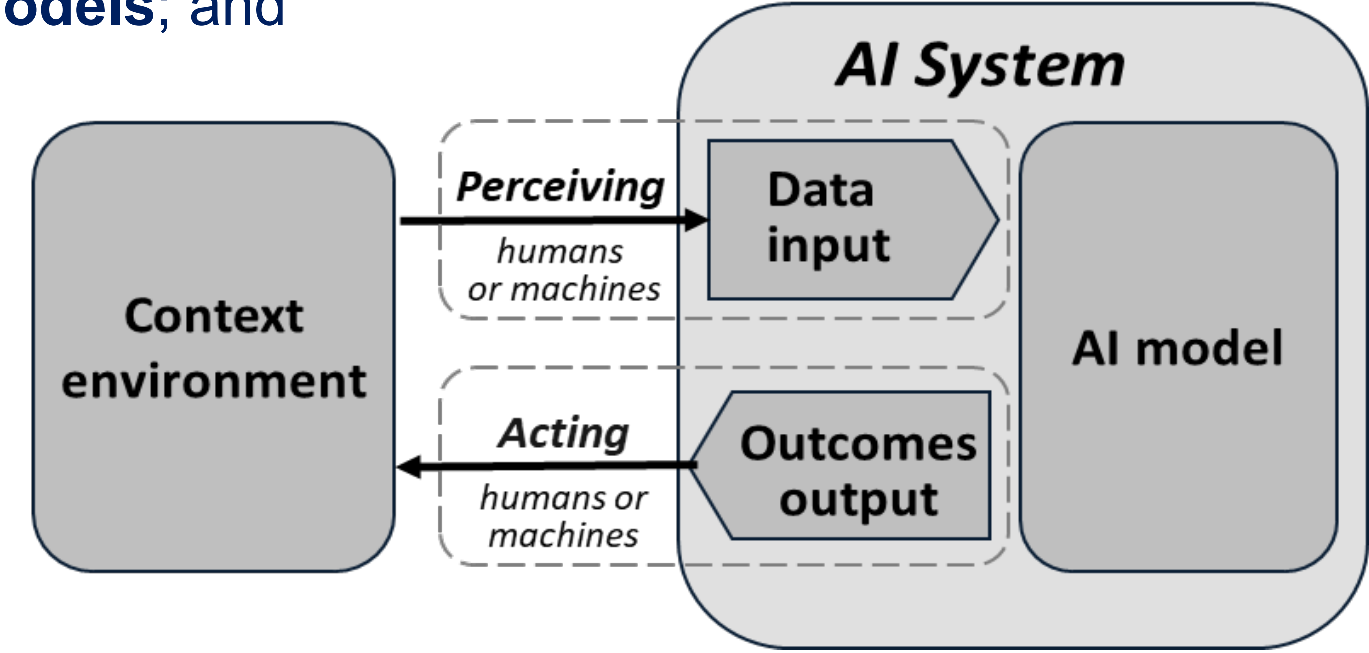
It uses machine and/or human-based inputs/data to:

i) **perceive** environments;

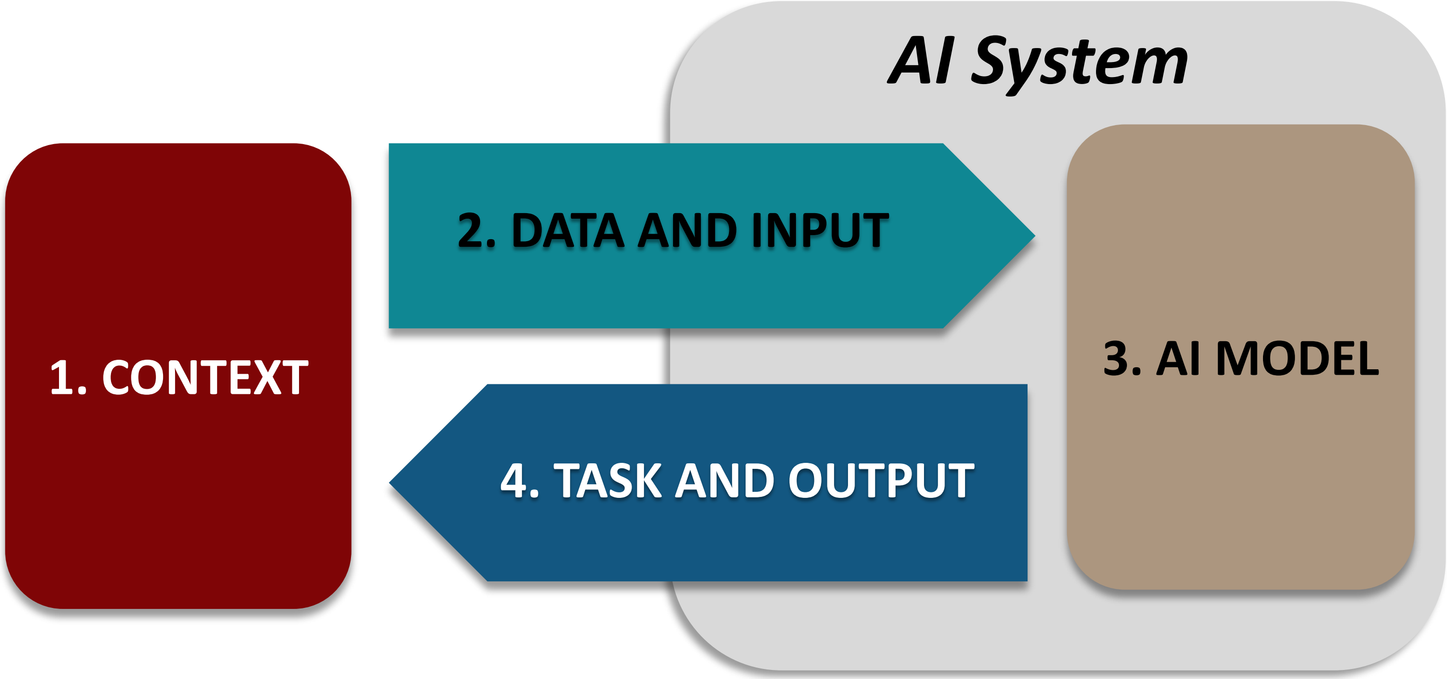
ii) abstract these perceptions **into models**; and

iii) **use** the models to formulate options for **outcomes**.

AI systems are designed to operate with **varying levels of autonomy**."



OECD framework to classify AI systems



Assessing the policy implications of different *types* of AI systems

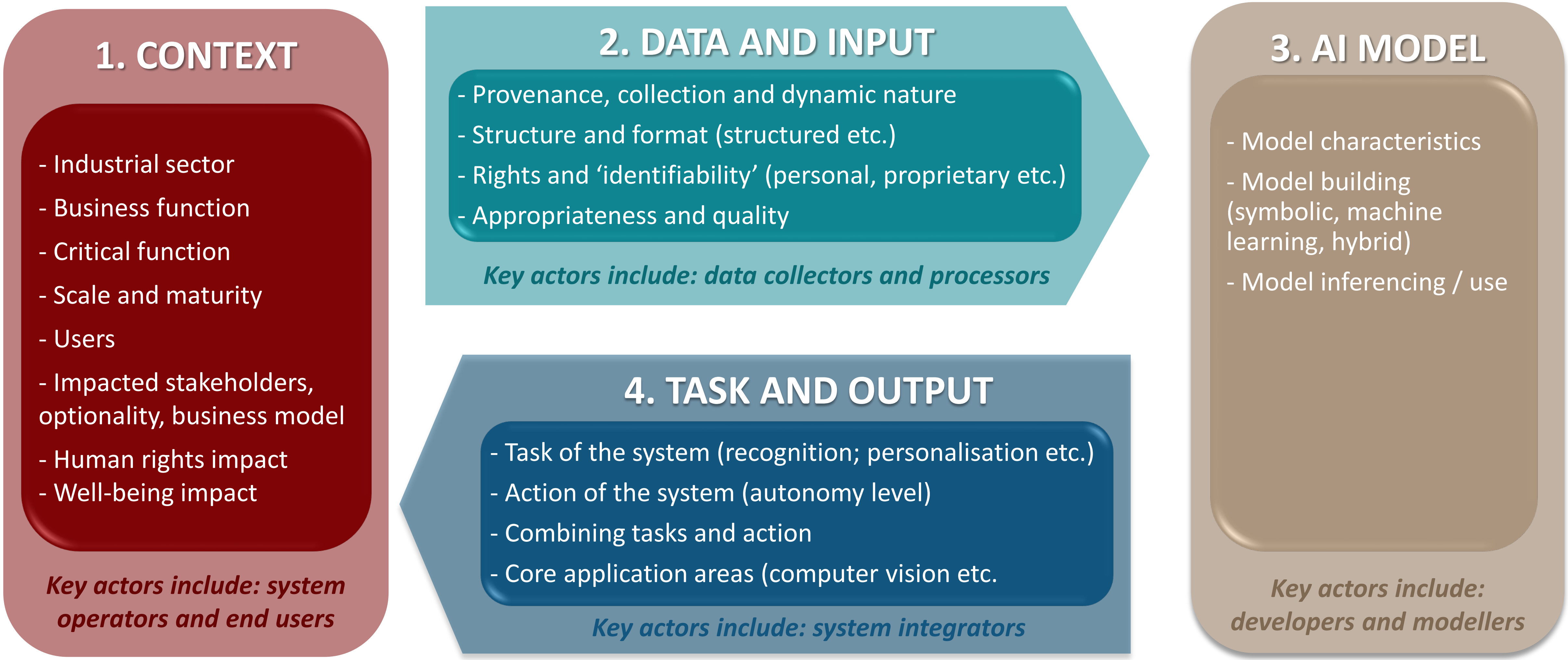
4 dimensions: **1. Context**, including sector (healthcare, etc.), impact and scale

2. Data and input, including data collection, personal nature of data

3. AI model (technologies), incl. model type and model building process

4. Task and output, incl. AI system's task (e.g., recognition, personalisation, etc.) and action autonomy

4 dimensions & 20 core criteria



1. CONTEXT

- Industrial sector
- Business function
- Critical function
- Scale and maturity
- Users
- Impacted stakeholders, optionality, business model
- Human rights impact
- Well-being impact

Key actors include: system operators and end users

2. DATA AND INPUT

- Provenance, collection and dynamic nature
- Structure and format (structured etc.)
- Rights and 'identifiability' (personal, proprietary etc.)
- Appropriateness and quality

Key actors include: data collectors and processors

3. AI MODEL

- Model characteristics
- Model building (symbolic, machine learning, hybrid)
- Model inferencing / use

Key actors include: developers and modellers

4. TASK AND OUTPUT

- Task of the system (recognition; personalisation etc.)
- Action of the system (autonomy level)
- Combining tasks and action
- Core application areas (computer vision etc.

Key actors include: system integrators

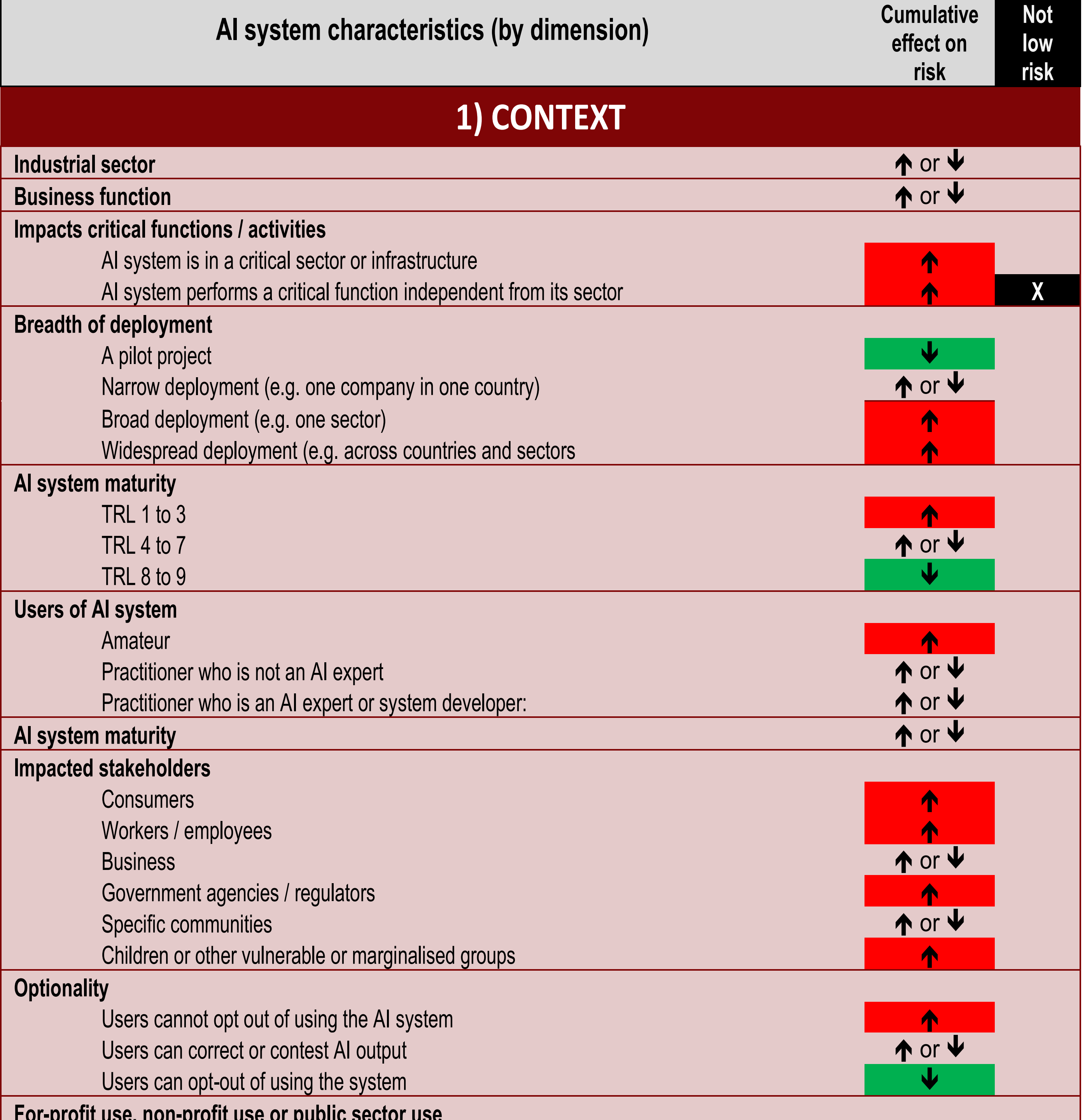
Work in progress:
Mapping systems' classification to risk

- Identified **characteristics that signify a system is not low risk,** including **potential risks for human rights,** building on Council of Europe's work
- Associated other characteristics with **positive, negative or neutral impact on risk** to obtain a preliminary cumulative effect

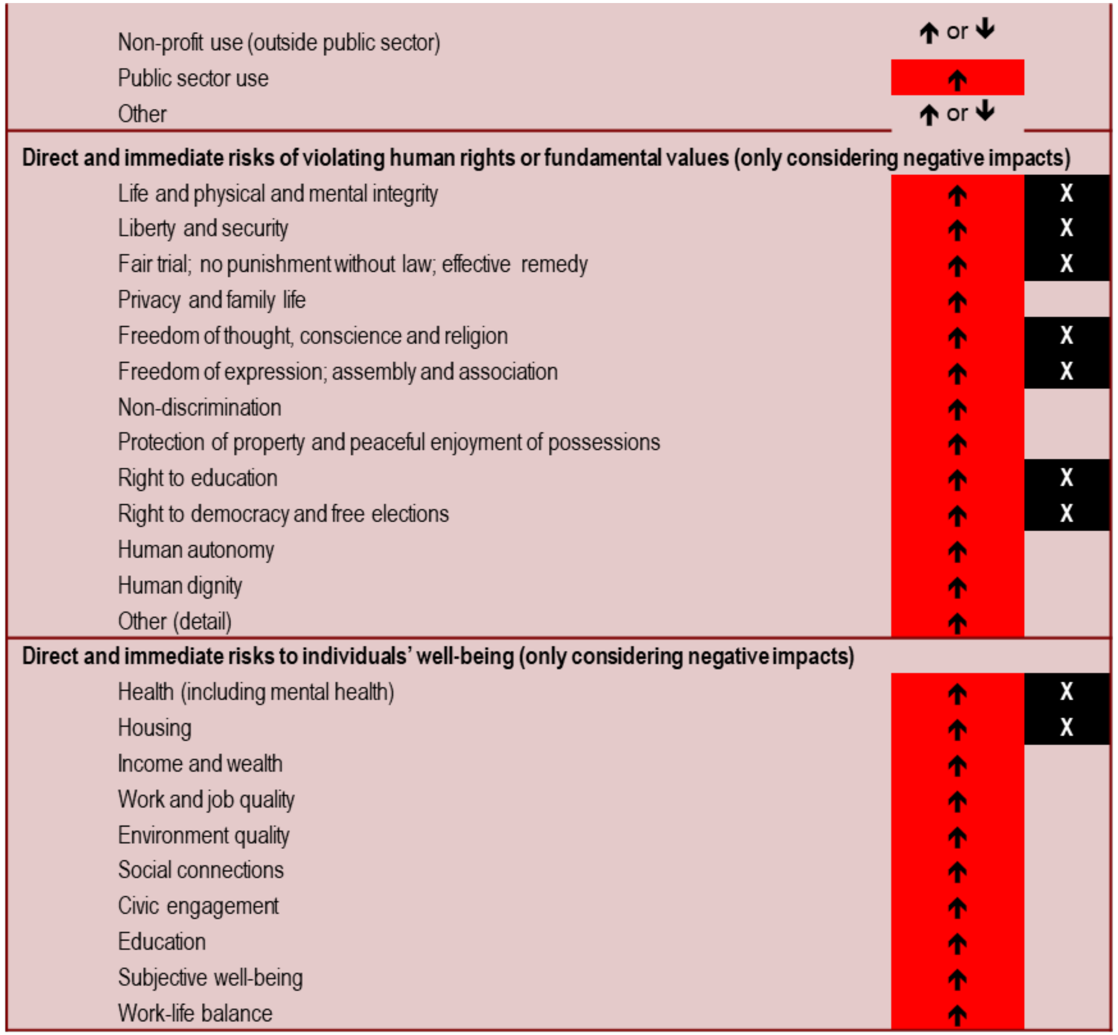
Next steps:

- Refine methodology and define output (*eg*. composite score)
- Test relevance & applicability

From classification to risk assessment (1)



AI system characteristics (by dimension)	Cumulative effect on risk	Not low risk
1) CONTEXT		
Industrial sector	↑ or ↓	
Business function	↑ or ↓	
Impacts critical functions / activities		
AI system is in a critical sector or infrastructure	↑	
AI system performs a critical function independent from its sector	↑	X
Breadth of deployment		
A pilot project	↓	
Narrow deployment (e.g. one company in one country)	↑ or ↓	
Broad deployment (e.g. one sector)	↑	
Widespread deployment (e.g. across countries and sectors	↑	
AI system maturity		
TRL 1 to 3	↑	
TRL 4 to 7	↑ or ↓	
TRL 8 to 9	↓	
Users of AI system		
Amateur	↑	
Practitioner who is not an AI expert	↑ or ↓	
Practitioner who is an AI expert or system developer:	↑ or ↓	
AI system maturity	↑ or ↓	
Impacted stakeholders		
Consumers	↑	
Workers / employees	↑	
Business	↑ or ↓	
Government agencies / regulators	↑	
Specific communities	↑ or ↓	
Children or other vulnerable or marginalised groups	↑	
Optionality		
Users cannot opt out of using the AI system	↑	
Users can correct or contest AI output	↑ or ↓	
Users can opt-out of using the system	↓	
For-profit use, non-profit use or public sector use		



Non-profit use (outside public sector)	↑ or ↓	
Public sector use	↑	
Other	↑ or ↓	
Direct and immediate risks of violating human rights or fundamental values (only considering negative impacts)		
Life and physical and mental integrity	↑	X
Liberty and security	↑	X
Fair trial; no punishment without law; effective remedy	↑	X
Privacy and family life	↑	
Freedom of thought, conscience and religion	↑	X
Freedom of expression; assembly and association	↑	X
Non-discrimination	↑	
Protection of property and peaceful enjoyment of possessions	↑	
Right to education	↑	X
Right to democracy and free elections	↑	X
Human autonomy	↑	
Human dignity	↑	
Other (detail)	↑	
Direct and immediate risks to individuals' well-being (only considering negative impacts)		
Health (including mental health)	↑	X
Housing	↑	X
Income and wealth	↑	
Work and job quality	↑	
Environment quality	↑	
Social connections	↑	
Civic engagement	↑	
Education	↑	
Subjective well-being	↑	
Work-life balance	↑	

Note: items marked "↑ or ↓" are to be assessed depending on the AI system usage and outcomes.

From classification to risk assessment (2, 3, 4)

Note: items marked "↑ or ↓" are to be assessed depending on the AI system usage and outcomes.

2) DATA AND INPUT	
Provenance of data and input	↑ or ↓
Detection and collection of data and input	↑ or ↓
Dynamic nature of data	
Static data	↓
Dynamic data updated from time-to-time	↑ or ↓
Dynamic real-time data	↑
Scale	↑ or ↓
Structure of data and input	↑ or ↓
Format of data and metadata	
Standardised data format	↑ or ↓
Non-standardised data format	↑
Standardised dataset metadata	↑ or ↓
Non-standardised dataset metadata	↑
Rights associated with data and input	
Proprietary data	↑
Public data	↑ or ↓
Personal data	↑
Identifiability of personal data	
Identified data	↑
Pseudonymised data	↑ or ↓
Unlinked pseudonymised data	↓
Anonymised data	↓
Aggregated data	↓
Data quality and appropriateness	
appropriateness of data for a particular problem	↓
(high) sample representativeness	↓
adequate sample size	↓
(high) completeness and coherence of sample	↓
(low) data noise	↓

3) AI MODEL	
AI model characteristics	
(High) transparency and explainability	↓
(High) safety, security, robustness	↓
(High) reproducibility	↓
Evolution during operation	↑
Evolution through uncontrolled learning	↑
Privacy-preserving properties, e.g. federated learning	↓

4) TASK AND OUTPUT	
Task of the system	
Recognition	↑ or ↓
Event detection	↑ or ↓
Forecasting	↑ or ↓
Personalisation	↑
Interaction support	↑
Goal-driven optimisation	↑ or ↓
Reasoning with knowledge structures	↑ or ↓
Action autonomy level	
High action autonomy	↑
Medium action autonomy	↑
Low action autonomy	↑ or ↓
No autonomy	↓
Displacement potential	
High displacement potential	↑
Core application areas	↑ or ↓

2. MITIGATION - Tools for trustworthy AI

Framework to evaluate different implementation approaches and help AI practitioners determine which tool fits their use case and how well it supports the OECD AI Principles for trustworthy AI.

Based on the framework, a frequently updated catalogue with interactive features and information on the latest tools will be built and made available on OECD.AI and as an API.

WG on implementing trustworthy AI: Identifying tools for trustworthy AI

1. Technical

- Toolkits / toolboxes / software tools
- Technical documentation
- Technical certification
- Technical standards
- Product development / lifecycle tools
- Technical validation tools

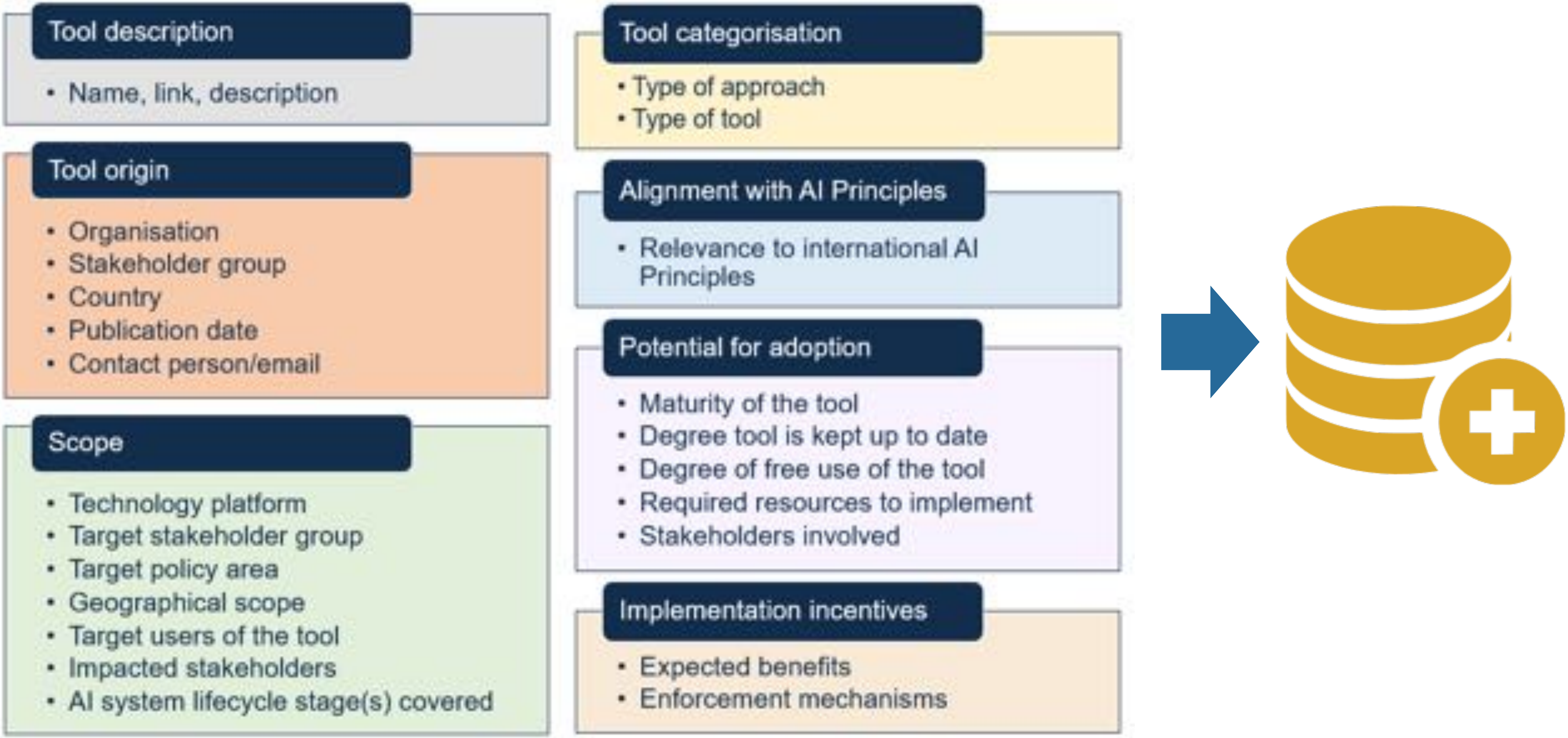
2. Procedural

- Guidelines
- Governance frameworks
- Product development / lifecycle tools
- Risk management tools
- Sector-specific codes of conduct
- Collective agreements
- Certification
- Process-related documentation
- Process standards

3. Educational

- Change management processes
- Capacity / awareness building
- Inclusive design guidance
- Educational materials / training programmes

A framework for trustworthy AI tools



Tool description

- Name, link, description

Tool origin

- Organisation
- Stakeholder group
- Country
- Publication date
- Contact person/email

Scope

- Technology platform
- Target stakeholder group
- Target policy area
- Geographical scope
- Target users of the tool
- Impacted stakeholders
- AI system lifecycle stage(s) covered

Tool categorisation

- Type of approach
- Type of tool

Alignment with AI Principles

- Relevance to international AI Principles

Potential for adoption

- Maturity of the tool
- Degree tool is kept up to date
- Degree of free use of the tool
- Required resources to implement
- Stakeholders involved

Implementation incentives

- Expected benefits
- Enforcement mechanisms

3. ASSURANCE - Tools for accountable AI

Public governance
Private governance
(Conformity assessments, auditing etc.)

- Back-up slides

OECD.AI
Policy Observatory

ONE AI working group on national AI policies

ONE AI Working Group on Policies for AI

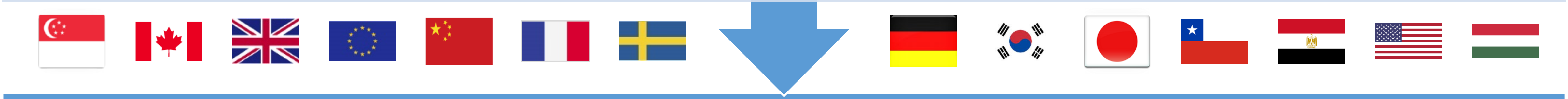
OECD.AI Policy Observatory

EC-OECD.AI database of national AI policies

Over 600 policy initiatives and national AI strategies from over 60 countries

+ AI experts' insights from governments and stakeholders

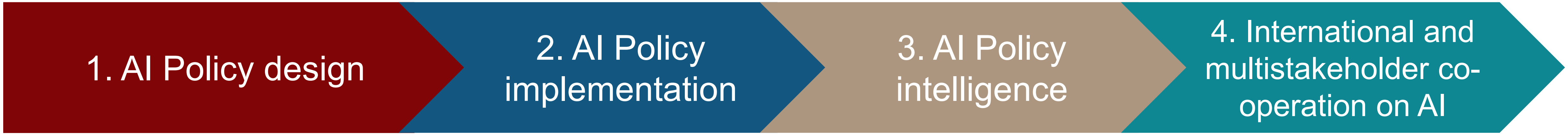
Present countries experiences in the design, implementation, and evaluation of AI policies, and challenges encountered



State of implementation of national AI policies

Evidence and lessons learned to date on implementing the five policy recommendations contained in the OECD AI Principles

AI policy cycle



1. AI Policy design

2. AI Policy implementation

3. AI Policy intelligence

4. International and multistakeholder co-operation on AI

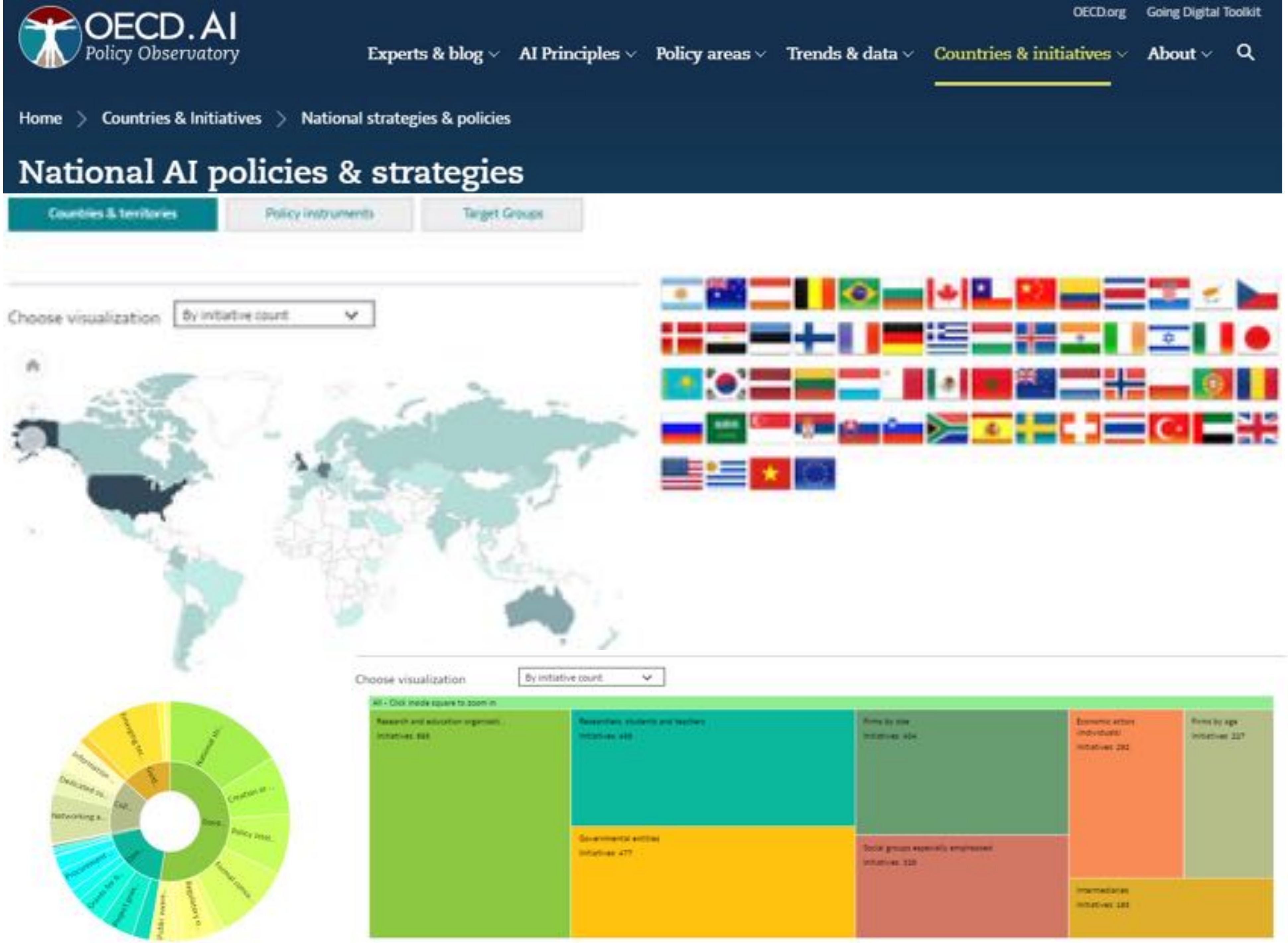
- National AI governance approaches (e.g. coordinating bodies, horizontal co-ordination, stakeholder participation and public consultations).

- Investing in AI R&D
- Data, compute, software and knowledge
- Regulation, testbeds, documentation.
- Automation, skills, jobs, education.
- Tools for trustworthy AI: codes of conduct, standards, capacity building.

- Translating AI policies into action plans and targets.
- Evaluating implementation of AI policies.
- Benchmarks and indicators (e.g. KPIs).

- International and multistakeholder co-operation (e.g. OECD, EC, Council of Europe, IDB, UNESCO, UN, WB, GPAI).
- Co-operation on standards development (e.g. ISO, IEEE).
- Multistakeholder initiatives.

National AI strategies



OECD.AI Policy Observatory

OECD.org Going Digital Toolkit

Experts & blog AI Principles Policy areas Trends & data Countries & initiatives About

Home > Countries & Initiatives > National strategies & policies

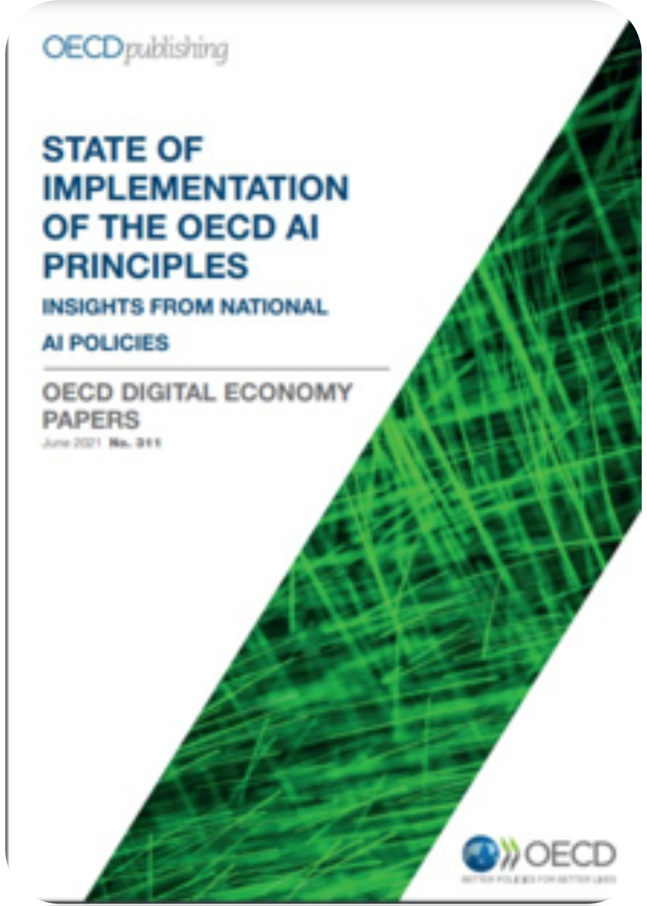
National AI policies & strategies

Countries & territories | Policy instruments | Target Groups

Choose visualization: By initiative count

Choose visualization: By initiative count

One joint EC/OECD database on national AI policies: two complementary reports launched 22 June 2021



OECDpublishing

STATE OF IMPLEMENTATION OF THE OECD AI PRINCIPLES

INSIGHTS FROM NATIONAL AI POLICIES

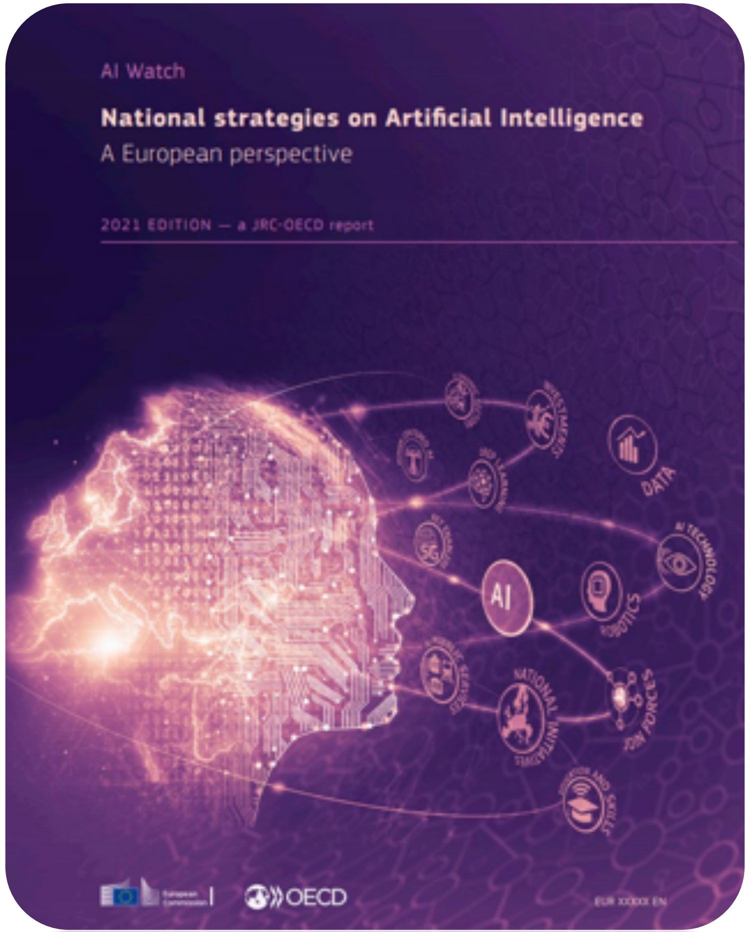
OECD DIGITAL ECONOMY PAPERS

June 2021 No. 311

OECD

BETTER POLICIES FOR BETTER LIVES

OECD Report



AI Watch

National strategies on Artificial Intelligence

A European perspective

2021 EDITION — a JRC-OECD report

European Commission | OECD

EUR XXXXX EN

EC-OECD Report

EC-OECD database of national AI policies and strategies

Globalpolicy.ai

International co-operation on trustworthy AI

COUNCIL OF EUROPE
CONSEIL DE L'EUROPE
European Commission
UNESCO
FRA
EUROPEAN UNION AGENCY FOR FUNDAMENTAL RIGHTS
IDB
OECD
UNITED NATIONS
WORLD BANK GROUP

What is Globalpolicy.ai?

- A **neutral portal** that gives access to information and resources on AI policy initiatives from inter-governmental organisations.
- Started in February 2020 to complement OECD.AI with a broader reach

Main objectives:

1. **Provide information and increase visibility**

   Provide an overview of each organisation's work on AI and to give access to relevant initiatives

2. **Help stakeholders navigate the AI landscape**

   Help policy makers and the wider public navigate the international AI governance landscape and access relevant resources

3. **Facilitate/promote co-operation**

   Provide a space where the initiatives can leverage each other's work, show progress towards common goals and pursue joint initiatives.

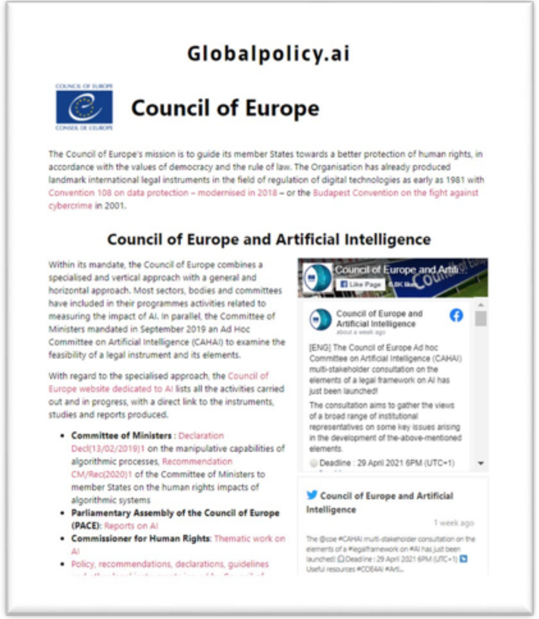
Progress to date

Platform

- ➢ Simple platform in FR & ENG
- ➢ Organisations listed, description at the top
- ➢ Live AI news map

Organisation pages

- ➢ Managed by each organisation
- ➢ Links to key projects
- ➢ Images, videos and other media
- ➢ Automated/live AI news relevant to each specific organisation on their pages
- ➢ Live RSS/social media feeds



Globalpolicy.ai

Council of Europe

The Council of Europe's mission is to guide its member States towards a better protection of human rights, in accordance with the values of democracy and the rule of law. The Organisation has already produced landmark international legal instruments in the field of regulation of digital technologies as early as 1981 with Convention 108 on data protection – modernised in 2018 – or the Budapest Convention on the fight against cybercrime in 2001.

Council of Europe and Artificial Intelligence

Within its mandate, the Council of Europe combines a specialised and vertical approach with a general and horizontal approach. Most sectors, bodies and committees have included in their programmes activities related to measuring the impact of AI. In parallel, the Committee of Ministers mandated in September 2019 an Ad Hoc Committee on Artificial Intelligence (CAHAI) to examine the feasibility of a legal instrument and its elements.

With regard to the specialised approach, the Council of Europe website dedicated to AI lists all the activities carried out and in progress, with a direct link to the instruments, studies and reports produced.

- **Committee of Ministers** : Declaration Decl(13/02/2019)1 on the manipulative capabilities of algorithmic processes, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems
- **Parliamentary Assembly of the Council of Europe (PACE)**: Reports on AI
- **Commissioner for Human Rights**: Thematic work on AI
- Policy, recommendations, declarations, guidelines

Council of Europe and Artifi...

Council of Europe and Artificial Intelligence
about a week ago

[ENG] The Council of Europe Ad hoc Committee on Artificial Intelligence (CAHAI) multi-stakeholder consultation on the elements of a legal framework on AI has just been launched!

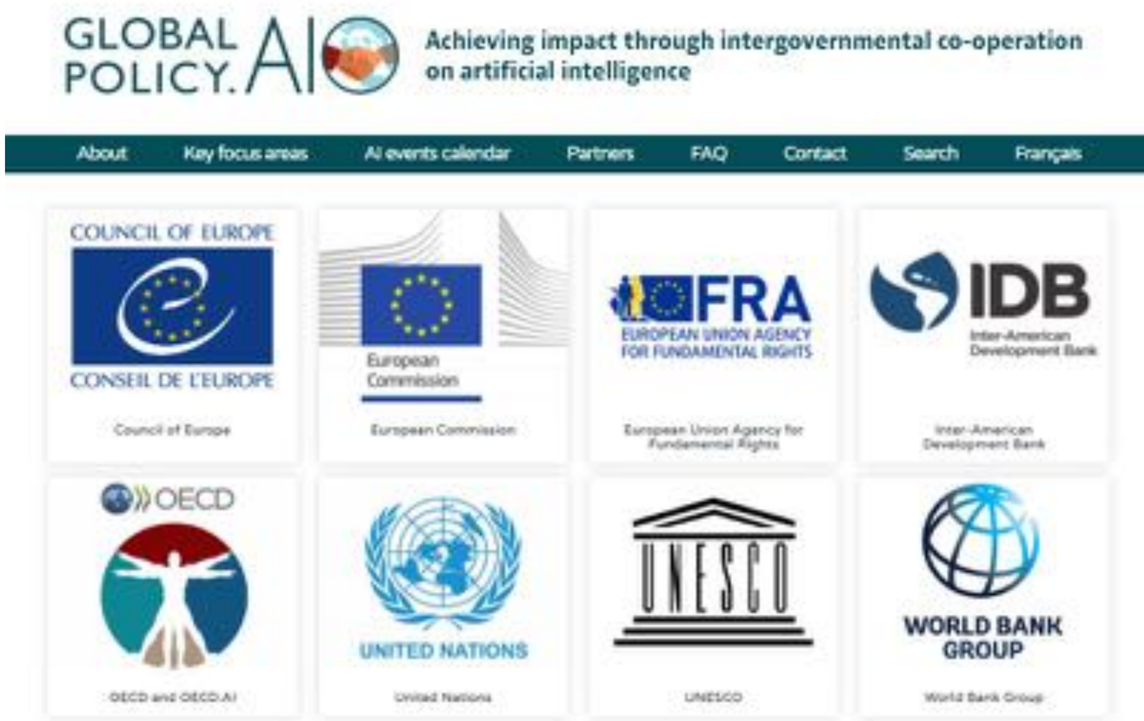
The consultation aims to gather the views of a broad range of institutional representatives on some key issues arising in the development of the-above-mentioned elements.

Deadline : 29 April 2021 6PM (UTC+1)

Council of Europe and Artificial Intelligence

1 week ago

The @coe #CAHAI multi-stakeholder consultation on the elements of a #legalframework on #AI has just been launched! Deadline : 29 April 2021 6PM (UTC+1) Useful resources #COE4AI #Arti...



GLOBAL POLICY.AI

Achieving impact through intergovernmental co-operation on artificial intelligence

Council of Europe

European Commission

European Union Agency for Fundamental Rights

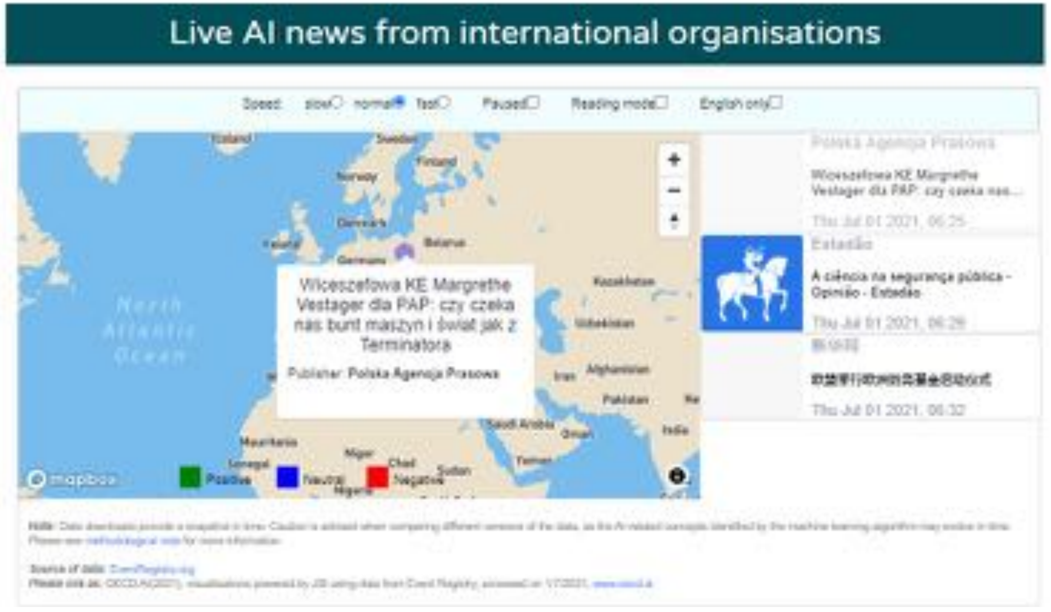
Inter-American Development Bank

OECD and OECD.AI

United Nations

UNESCO

World Bank Group



Live AI news from international organisations

Speed: slow ○ normal ● fast ○ Paused □ Reading mode □ English only □

Wiceszefowa KE Margrethe Vestager dla PAP: czy czeka nas bunt maszyn i świat jak z Terminatora

Publisher: Polska Agencja Prasowa

Polska Agencja Prasowa
Wiceszefowa KE Margrethe Vestager dla PAP: czy czeka nas...
Thu Jul 01 2021, 06:25

Estadão
A ciência na segurança pública - Opinião - Estadão
Thu Jul 01 2021, 06:29

[illegible]
[illegible]
Thu Jul 01 2021, 06:32

🟩 Positive 🟦 Neutral 🟥 Negative

Note: Data downloads provide a snapshot in time. Caution is advised when comparing different versions of the data, as the AI-related concepts identified by the machine learning algorithm may evolve in time. Please see methodological note for more information.

Source of data: EventRegistry.org
Please cite as: OECD.AI(2021), visualisations powered by JSI using data from Event Registry, accessed on 1/7/2021, www.oecd.ai

Globalpolicy.ai next steps

A neutral cooperative platform for IGOs

Soft launch	June 2021 at CoE Launch 14-15 September at EC / EU Slovenia event
Adding more functionalities	Addition of more functionalities
Potential partnerships	Explore potential partnerships to make the platform into a repository/hub of resources and use cases on AI for good e.g. mapping best practices of AI in different thematic areas, such as the UN SDGs and international AI Principles.