

The background is a vibrant blue with a futuristic, digital aesthetic. It features a large, glowing wireframe hand on the right side, composed of interconnected nodes and lines. The background is also filled with various circular patterns, some resembling clock faces or data wheels, and a network of glowing nodes connected by thin lines, suggesting a complex data structure or AI system. The overall tone is high-tech and innovative.

Responsible Machine Learning: Legal and Technical Aspects

Introduction of course team

- Doaa ABU ELYOUNES
dabuelyounes@sjd.law.harvard.edu
- Nicolas VAYATIS
nicolas.vayatis@ens-paris-saclay
- Marie Garin
mgarin@ens-paris-saclay.fr

Course objectives

- Bridging the gap between the technical aspects of ML and the legal and policy aspects.
- Discussing the social aspects of different technical choices.
- Learning how mechanism design can in practice create policy.
- Getting a glimpse of the regulation surrounding the technology domain, its opportunities and limitations.

Legal and technical definitions of AI

- Many AI and machine learning researchers tend to favor definitions of AI that emphasize technical functionality while policy-makers favor definitions that emphasize comparison to human thinking and behavior.
- From a technical perspective, the lack of precise and universally accepted definition of AI has helped the field to grow.
- In order for the definition to be suitable for policy making, it has to take into account both currently deployed AI technologies and future applications, it has to be accessible for non-technical experts, and it has to be flexible enough to be synchronized with other policies including allowing for oversight procedures.
- In the media, AI is often portrayed in a very futuristic and unrealistic way, and references to killer robots and humanoids are leading to misinformed public perceptions.

Sessions' description

- Session 1. How AI is regulated today and what the future holds (04.10.2023)
- Session 2. Fairness and non-discrimination- legal perspectives (11.10.2023)
- Session 3. Fairness and non-discrimination- technical perspectives (18.10.2023)
- Session 4. Fairness and non-discrimination- practical examples (25.10.2023)
- Session 5. Privacy and data protection- technical perspectives (08.11.2023)
- Session 6. Privacy and data protection- legal perspectives (15.11.2023)
- Session 7. Privacy and data protection- practical examples (22.11.2023)
- Session 8. Transparency and explainability- technical perspective (06.12.2023)
- Session 9. Transparency and explainability- legal perspective (13.12.2023)
- Session 10. Generative AI and impact on the environment (20.12.2023)
- Session 10. Presentations of final project (10.01.2023)
- Session 11. Presentations of final projects (13.01.2022)

Evaluation

- Technical homework, **30%**
- Critical analysis of two readings from the syllabus, **30%**
- Final legal assignment, role play debate in which students will have to work in teams and present one party in a legal case in front of judges, **40%**
- **Bonus factor: Attendance and active participation**

Introduction of students

- In 1-2 minutes please introduce yourself;
- Please mention your name and background;
- Please mention why you are interested in this course and if there is a specific topic you wish we would cover beyond what we covered in the outline of the course.

Regulating technology

<https://www.youtube.com/watch?v=3WS9DhSIWR0>

John Perry Barlow, A Declaration of the Independence of Cyberspace (1996)

“Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather. ... I declare the global social space we are building to be naturally independent of the tyrannies you seek to impose on us. You have no moral right to rule us nor do you possess any methods of enforcement we have true reason to fear.”

Lawrence Lessig, Code and other Laws of cyberspace (1999)

“Our choice is not between “regulation” and “no regulation.” The code regulates. It implements values, or not. It enables freedoms, or disables them. It protects privacy, or promotes monitoring. People choose how the code does these things. People write the code. Thus the choice is not whether people will decide how cyberspace regulates. People — coders — will. The only choice is whether we collectively will have a role in their choice — and thus in determining how these values regulate — or whether collectively we will allow the coders to select our values for us.”

Questions

- What do you think about the two statements?
- With which one you resonate more?
- The quotes were written at the middle and end of the 1990's, given the state of technological development today, do you think we need to be more strict about regulation or less so?
- Zooming on AI, can you think of an example for system that underline the debate on regulation?

Governing AI through principles

- More than 160 frameworks and guidelines for governing AI have been developed in the last 4 years.
- Principles are being developed by all actors in the field including private companies, standard bodies and international organizations.
- The rush to develop principles started after a series of articles exposed horrific discriminatory incidents caused by biased algorithms.

The advantages of governing through principles

- The efforts to govern cyberspace also begun in the form of principles.
- The general principles helped in laying the ground for specific rights.
- General principles can be used by civil society, the media and the judiciary for solving gaps in existing laws.
- While in the past most efforts to govern cyberspace focused on global governance and encouraged international collaborations, in recent years there has been a shift toward regional and national laws and there are more initiatives targeting changes that need to be done by the private sector.

Global Consensus on the Ethics of AI

- 193 countries agree to the first-ever global framework on Ethics of AI
 - deliver more effective rules & regulations to ensure the beneficial development of AI
 - control the downside risks
- The Recommendation asks Member States to:
 - ensure more inclusive, diverse and fair outcomes
 - enact strong enforcement mechanisms and remedial actions
 - improve their capacities to deal with this technologies.

Strengthen Human Rights and Democracy



Right to Equality

Right to Privacy

Protection against surveillance and curtailing of liberties

Freedom of expression and freedom of opinion

Right to Education

Right to Work

...

UNESCO Framework

Values

1. Respect, protection and promotion of human rights and fundamental freedoms and human dignity
2. Environment and ecosystem flourishing
3. Ensuring diversity and inclusiveness
4. Living in peaceful, just and interconnected societies



Principles

1. Proportionality and do no harm
2. Safety and security
3. Fairness and non-discrimination
4. Sustainability
5. Right to privacy, and data protection
6. Human oversight and determination
7. Transparency and explainability
8. Responsibility and accountability
9. Awareness and literacy
10. Multi-stakeholder and adaptive governance and collaboration

UNESCO Framework (contd.)

The Recommendation goes into details of actions to ensure accountability, responsibility, transparency and necessary regulations to ensure the rule of law.

Action-oriented cross-sectoral policy chapters of the Recommendation covers:



Data policy



Gender



Development
& international
cooperation



Environment &
ecosystems



Health
and social
well-being



Communication &
information



Education
& research



Economy
& Labour



Culture

POLICY IN PRACTICE: DATA POLICY (POLICY AREA 3)

Member States should...

- Develop data governance strategies, ensuring the **continual evaluation** of the **quality of training data** for AI system
- Put in place appropriate safeguards to **protect the right to privacy** in accordance with international law
- Ensure that individuals retain their rights over personal data and are protected by a framework

GENDER

FINANCING GENDER RELATED SCHEMES

INCLUDING GENDER ACTION PLAN

ADDRESSING WAGE AND EQUAL OPPORTUNITIES GAP

ENCOURAGING FEMALE ENGAGEMENT IN AI

INCREASING FEMALE PARTICIPATION IN STEM AND ICT

ERADICATING GENDER STEREOTYPING AND BIAIS



What are the challenges and disadvantages of governing through principles?

Moving from principles to practice

- Principles are too general and cannot be easily implemented in different domains.
- Principles can conflict with each other.
- There will always be tradeoffs that principles are not able to solve.
- Different groups might interpret principles differently, meaning that lists of broadly agreed-upon principles cannot recognize that important and legitimate differences in values exist across people and populations.

How can we move from principles to practice?

- Who are the stakeholders that are responsible for implementing the principles?
- What is their exact responsibilities?

Common avenues for translating principles into practice

- Governing AI in a sectoral manner
- Enacting new regulations, both transversal overarching ones and sectoral ones
- Adapting existing laws
- Adopting industry standards
- Instating periodical checkups and reporting requirements
- Adopting technology assessment tools

The AI Act - what seems to be at stake for the EU?

- Optimization, resource allocation, etc. is especially needed in high-impact sectors, including climate change, environment and health, the public sector, finance, mobility, home affairs and agriculture.
- AI brings about new risks
- The EU is committed to strive for a balanced approach.
- It is in the Union interest to preserve the EU's technological leadership

Twin objectives: Promoting the uptake of AI and of addressing the risks associated with certain uses of such technology. The AI Act seeks to implement the second objective: the development of an ecosystem of trust by proposing a legal framework for trustworthy AI

In the broader EU context

- The proposed Regulation is part of a tranche of proposals which must be understood in tandem, including:
- The Digital Services Act (with provisions on recommenders and research data access);
- The Digital Markets Act (with provisions on AI-relevant hardware, operating systems and software distribution);
- Announced product liability revision relating to AI
- The draft Data Governance Act (concerning data sharing frameworks)

The AI Act - specific objectives

- Ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values;
- Ensure legal certainty to facilitate investment and innovation in AI;
- Enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;
- Facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation

Key elements

- A single future-proof definition of AI
- Proportionality - imposes regulatory burdens only when an AI system is likely to pose high risks to fundamental rights and safety. (except transparency)
- risk-based approach / methodology to define “high-risk” AI
- Retain particularly harmful AI practices are prohibited as contravening Union values,
- Specific restrictions and safeguards are proposed in relation to certain uses of remote biometric identification systems for the purpose of law enforcement.
- “Throughout the whole AI systems’ lifecycle”
- Transparency obligations (flag AI is being used)

Other elements

- Reporting obligation for high-risk AI applications in a public EU-wide database + inform incidents

The Definition

Article 3 Definitions

For the purpose of this Regulation, the following definitions apply: (1) 'artificial intelligence system' (AI system) means 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;

ANNEX I - ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES referred to in Article 3, point 1

- (a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;
- (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;
- (c) Statistical approaches, Bayesian estimation, search and optimization methods.

Art. 5 Four levels of risk

- i) Unacceptable risks (Title II)
- ii) High risks (Title III)
- iii) Limited risks (Title IV)
- iv) Minimal risks (Title IX).



Title II: Prohibited practices (Unacceptable risks)

- Manipulative systems
 - subliminal systems
 - “in order to” materially distort behavior
 - and **individual** physical or psychological
- Social scoring
 - by or on behalf of public authorities
 - leading to...
- Biometric systems in publicly accessible spaces **by law enforcement**
 - 3 exceptions
 - Excludes public space *online!* (see recital 9)
 - No “placing on the market” prohibition

Title III: High Risk Systems (Art. 6 - annex II)

- AI systems that are products or safety components (broadly construed) of products already covered by certain Union health and safety harmonisation legislation (such as toys, machinery, lifts, or medical devices).
- 'Standalone' AI systems specified in an annex for use in eight fixed areas: (Comes from product regulation)

- Biometric identification - remote and 'post' (v. art. 5)
- Management and operation of critical infrastructure
- Educational and vocational training
- Employment, worker management and access to self-employment
- Access to and enjoyment of essential services and benefits
- Law enforcement
- Migration, asylum and border management
- Administration of justice and democracy

UNESCO technology assessment tools

The Recommendation acknowledges AI to be **continuously evolving**:

- The readiness of countries to develop, implement and use AI need to be assessed in a **dynamic** fashion
- Ensuring that AI systems and tools abide by the **ethical principles** agreed by Member States requires a **methodology** able to detect and address the challenges and shortcomings of AI.

Readiness Assessment Methodology (RAM)

Macro diagnostic tool

helps countries to understand where they stand on the **scale of preparedness to implement AI ethically and responsibly for all their citizens**, highlighting the needed institutional and regulatory changes and helping UNESCO **tailor capacity building efforts to country-specific needs**.



unesco






Ethical Impact Assessment (EIA)

Micro-level procurement tool

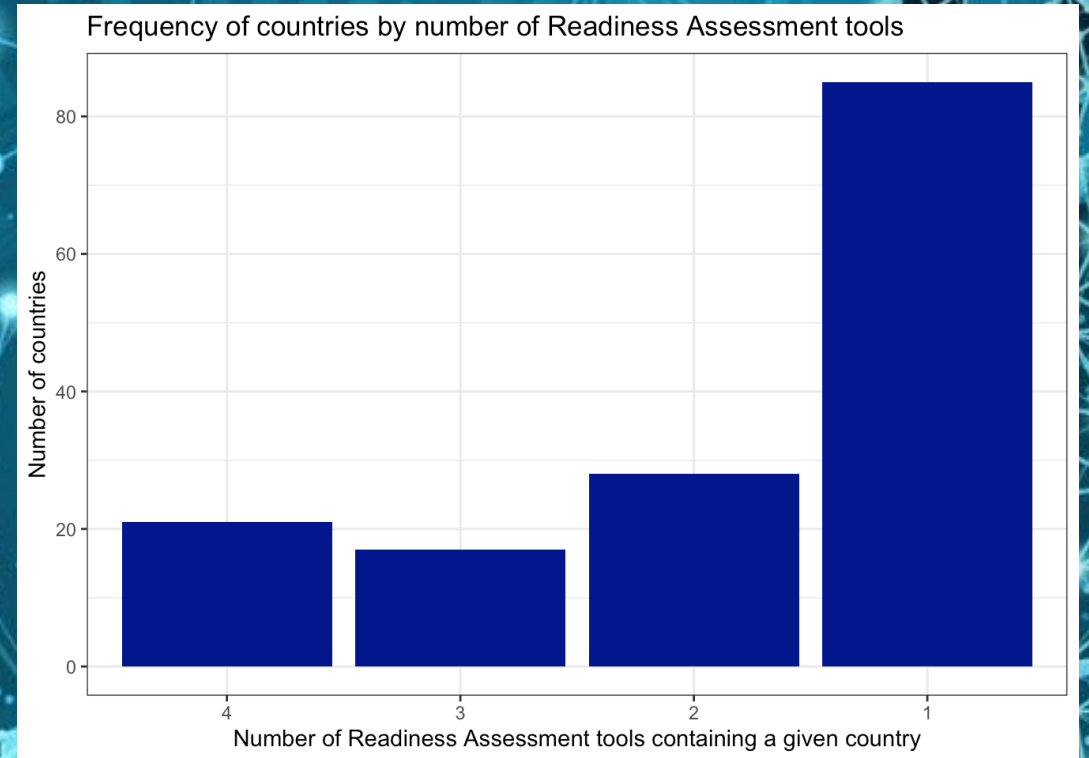
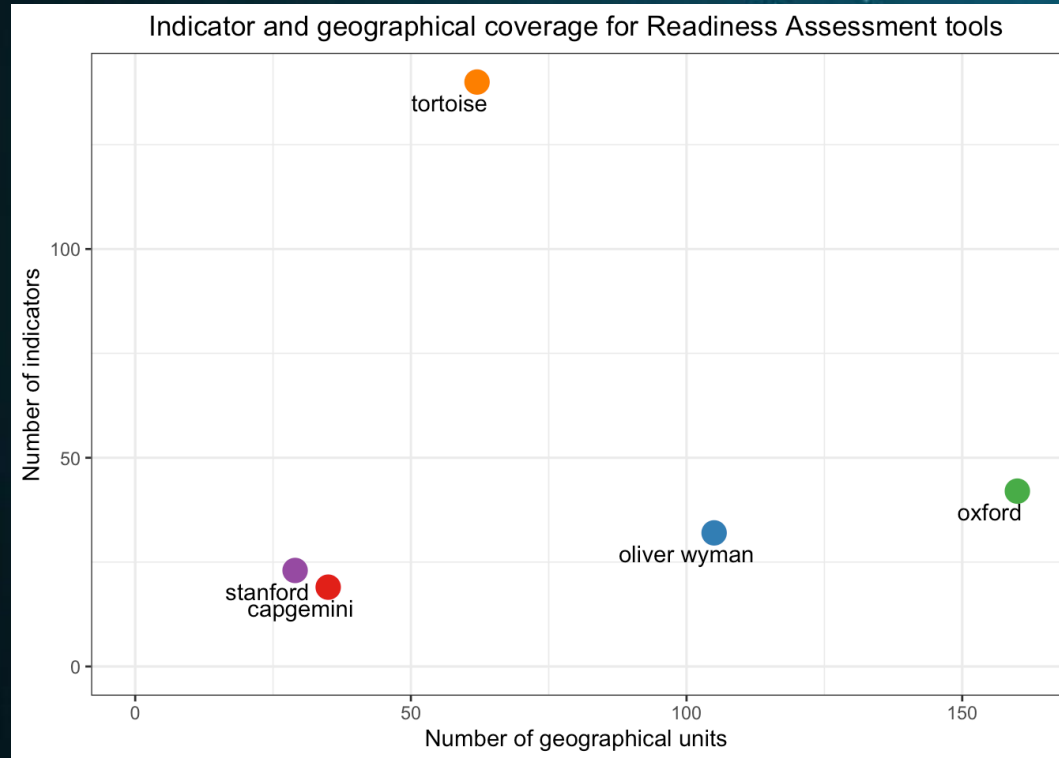
help government officers involved in the **procurement of AI systems** ensure that the AI systems that they will be purchasing are **aligned with the ethical standards set out in the UNESCO Recommendation**.

The Readiness Assessment Methodology — Background

A study conducted by the secretariat looked at five existing readiness assessment tools:

- The Government AI Readiness Index – Oxford Insights  OXFORD INSIGHTS
- The AI Index Report – Stanford HAI  **Stanford University**
Human-Centered
Artificial Intelligence
- The Global AI Index – Tortoise Media 
- The AI Readiness Benchmark – Capgemini Consulting 
- Global Cities AI Readiness Index – Oliver Wyman Forum 

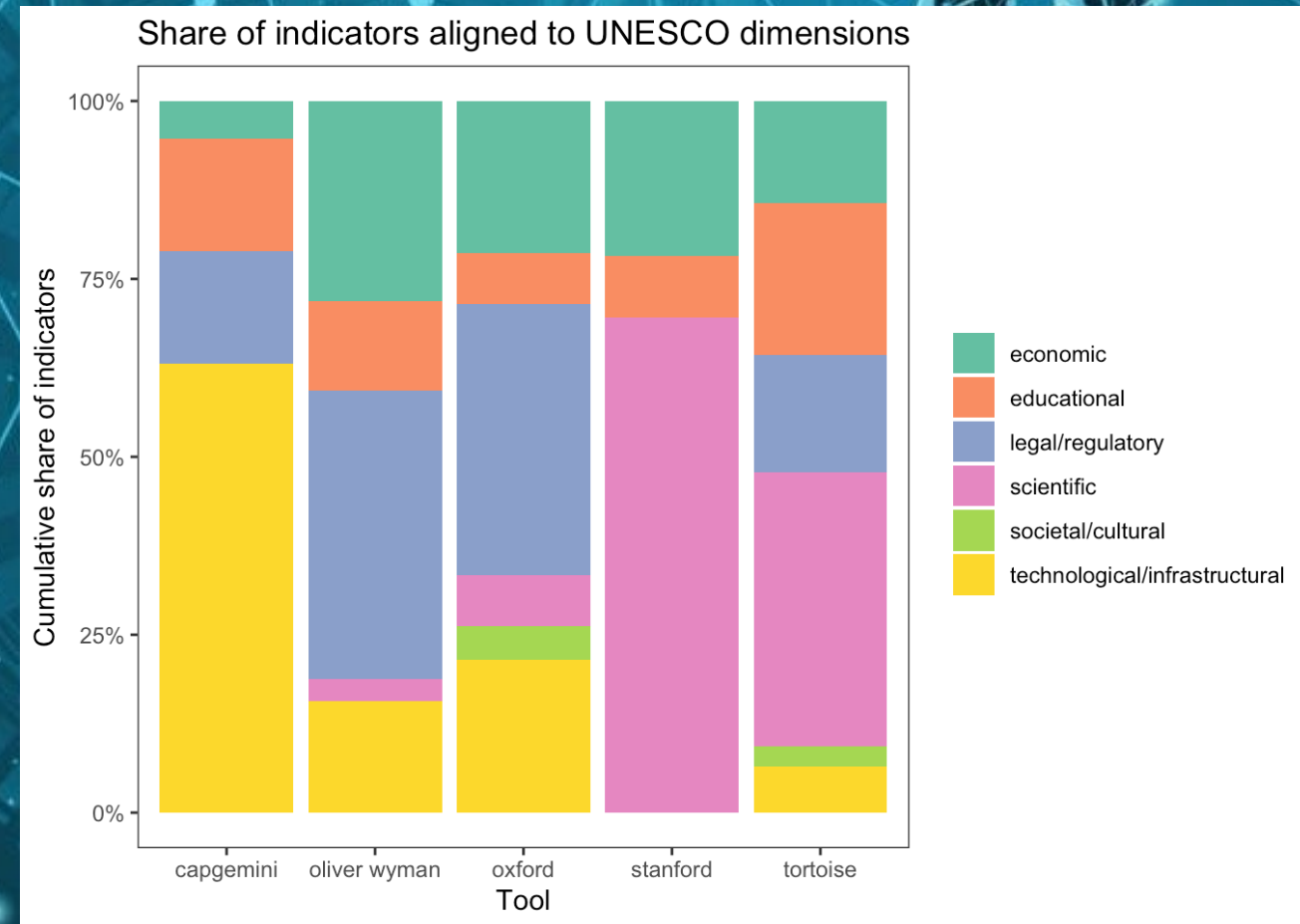
Gaps in Existing Tools — Coverage



- Existing tools cover a **very limited set of countries** (mainly from the **Global North**)
- **Only 21 countries worldwide** appear in **all four tools**

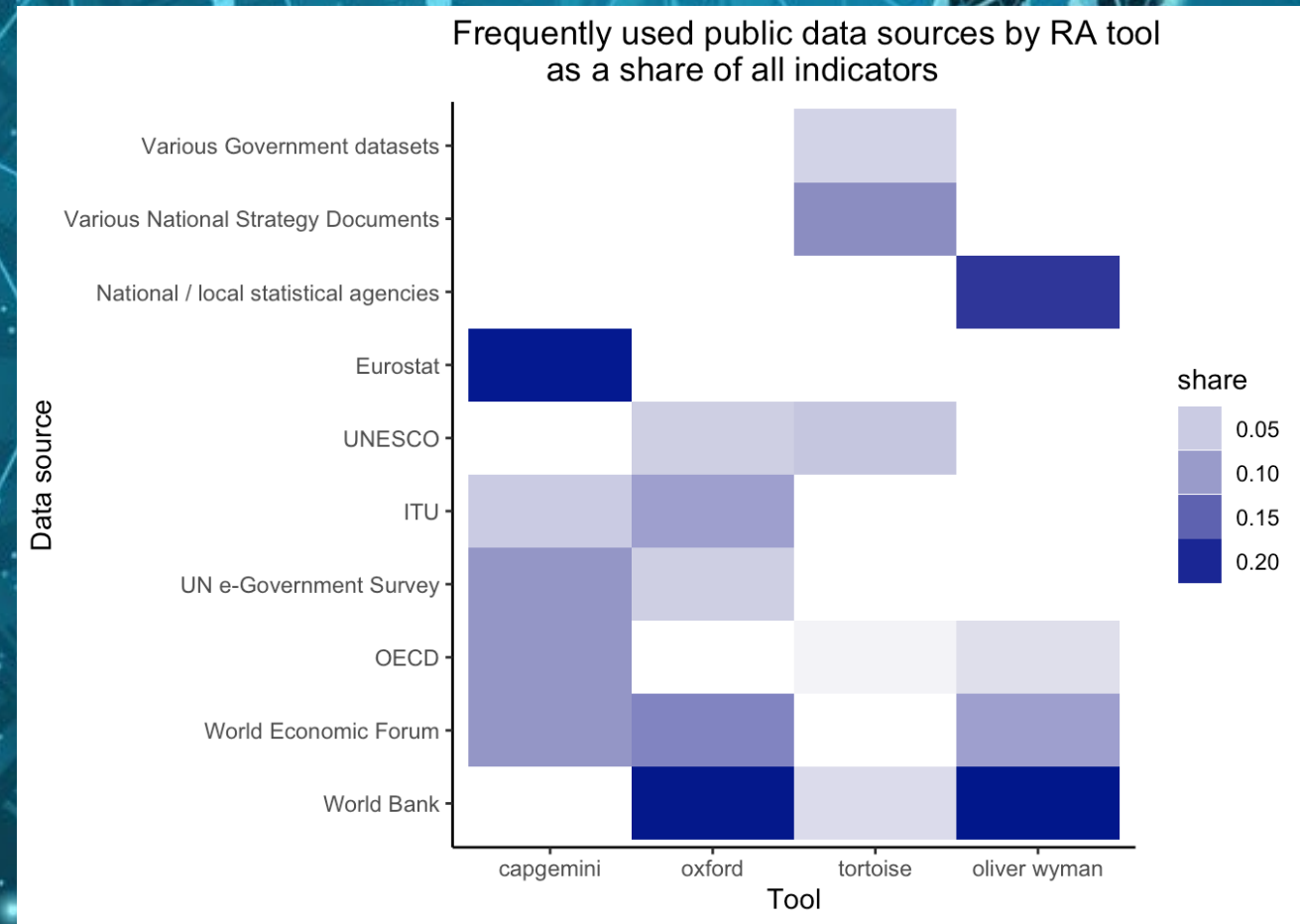
Gaps in Existing Tools — Alignment with UNESCO's Recommendation

- Existing tools mainly cover the **technical/infrastructural** and **scientific** dimensions
- Existing tools rarely assess **ethical** or **societal/cultural** aspects
- Most existing tools include **generic indicators** that are not specific to AI.

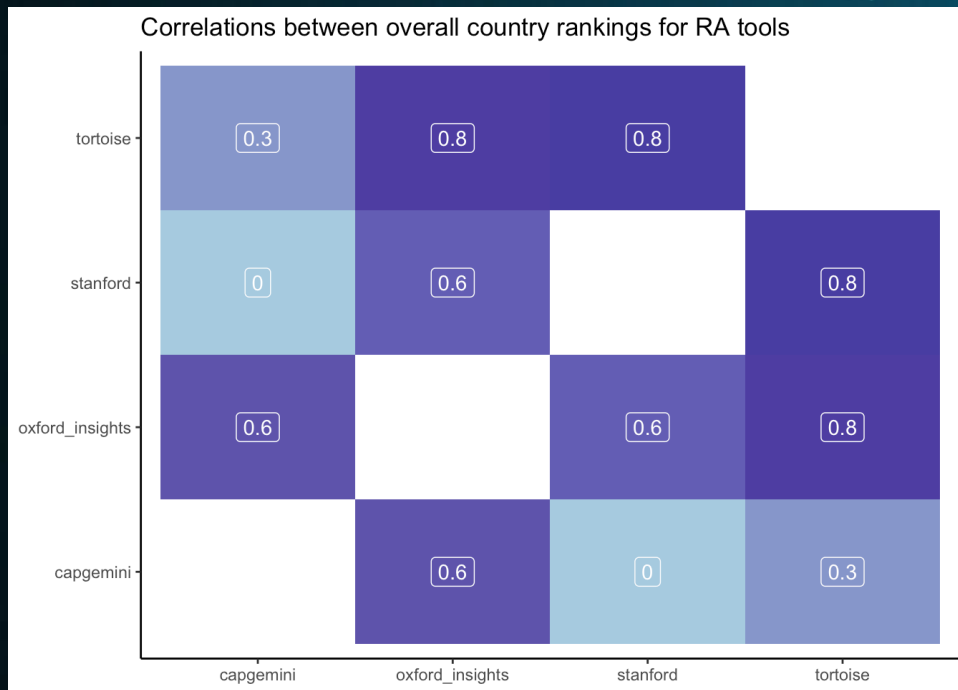


Gaps in Existing Tools — Types of Data Sources

- Data collection is a very difficult task, which is why **public data sources** play a major role in existing tools
- UNESCO's advantage is the ability to communicate with countries **directly** and collect **primary data**



Gaps in Existing Tools — Correlation Between the Tools



	Capgemini	Oxford Insights	Stanford	Tortoise
1	Netherlands	United States	United States	United States
2	Denmark	United Kingdom	China	China
3	Sweden	Finland	United Kingdom	United Kingdom
4	United Kingdom	Netherlands	Canada	Canada
5	Finland	Sweden	South Korea	South Korea

- There is a **lack of correlation** in terms of ranking and score **between tools**
- This also holds for **specific dimensions**
- Ranking/ scoring the countries is **subjective**: it changes a lot based on the selected indicators, their amount, and the weight given to each indicator.

UNESCO's Readiness Assessment Methodology

The RAM includes 5 key dimensions, each containing various indicators and sub-indicators:

1. Legal
2. Social/ Cultural
3. Scientific/ Education
4. Economic
5. Technological/ Infrastructural

As per the legal dimension, what are the areas of law that would require adaptation?



UNESCO's Readiness Methodology – the Legal Dimension

This dimension will include the questions on the following:

- *AI policy and regulation*
- *Data protection and privacy laws*
- *Data sharing and accessibility*
- *Procurement laws and policies*
- *Freedom of information acts/ access to knowledge acts*
- *Due process and accountability*
- *Online safety and integrity of speech*
- *Public sector capacity*

UNESCO's Readiness Methodology – the Legal Dimension

Example: Data protection and privacy laws

2.2.2 Data protection and privacy laws

QUANTITATIVE 2.2.2.1 Score of the country on the Cybersecurity Index³

QUALITATIVE 2.2.2.2 Does your country have a data protection law?⁴ If not, is your country in the process of enacting such regulation? Please elaborate and provide the name and link to any relevant document(s).

Evaluation:

2.2.2.2.1 Has the efficacy of the data protection law been assessed? Please elaborate and provide the link to any relevant document(s).

2.2.2.2.2 Does the data protection law give users control over their data and allow them to delete it?

2.2.2.2.3 Does the data protection law mention a notice and consent framework and specify in which cases it applies?

2.2.2.2.4 Does the data protection law include transparency requirements on data usage?

2.2.2.2.5 Does the data protection law include requirements on data minimization?

2.2.2.2.6 Does the data protection law highlight cases in which data protection or privacy impact assessment is required?

2.2.2.2.7 Does the data protection law include specific rules for sensitive information (e.g., health data)?

2.2.2.2.8 Does the data protection law include enforcement mechanisms and compensation schemes in case of violation?

2.2.2.2.9 Are there different standards of data protection applied for data collected by public vs. private entities?

2.2.2.2.10 Is privacy and/or respect for private and family life protected under the data protection law or another law? Please provide the link to any relevant document(s).

2.2.2.2.11 Do you have a data protection entity or data protection officer in your country? If yes, what is their mandate?



UNESCO's Readiness Methodology – the Social & Cultural Dimension

This dimension will include the questions on the following:

- *Diversity, inclusion and equality*
- *Public engagement and trust*
- *Environmental and sustainability policies*
- *Health and social well-being*
- *Culture*

UNESCO's Readiness Methodology – the Scientific/ Educational Dimension

This dimension will include the questions on the following:

Research and Innovation

- *R&D expenditure*
- *Research output*
- *Ethical AI research*
- *AI talent*
- *Innovation output*

Education

- *Education strategy*
- *Education infrastructure*
- *Curriculum content*
- *Educational attainment*
- *Public Access to AI education*

UNESCO's Readiness Methodology – the Economic Dimension

This dimension will include the questions on the following:

- *Labor markets*
- *Intermediate consumption*
- *Investments and output*

Example: Labor Markets

5.2.1 Labour markets	
QUANTITATIVE	5.2.1.1 Share of job vacancies posted requiring AI-related skills (online job vacancies ideally) ³⁹
	5.2.1.2 Share of current employees working as data scientists (exact SOC code to be refined)
	5.2.1.3 Relative AI skill penetration ⁴⁰
	5.2.1.4 AI talent concentration ⁴¹
QUALITATIVE	5.2.1.5 Does your country have a strategy to respond to AI impact on the labour market? This includes issues such as re-skilling of workers affected by automation, upskilling of workers to take advantage of opportunities presented by AI, and considering the soft skills advantages and complementarity of human skills relative to AI systems. If not, is such a strategy in the process of being adopted? Please elaborate and provide the name and link to any relevant document(s).
	Evaluation:
5.2.1.5.1	Has the efficacy of this strategy been assessed? Please provide the link to any relevant document(s).

UNESCO's Readiness Methodology – the Technological/ Infrastructural Dimension

This dimension will include the questions on the following:

- *Infrastructure and connectivity*
- *Applied standards*
- *Computing capabilities*
- *Statistical performance*

Example: Computing capabilities

6.2.3 Computing capabilities

QUANTITATIVE	6.2.3.1	Number of data centres in the country per capita ⁵⁷
	6.2.3.2	Distance to the closest data centre ⁵⁸ in km
	6.2.3.3	Colocation data centres ⁵⁹ per million population
QUALITATIVE	6.2.3.4	Does your country have a policy for AI-driven cloud computing? If not, is such a policy in the process of being adopted? Please elaborate and provide the name and link to any relevant document(s).
	Evaluation:	
	6.2.3.4.1	Has the efficacy of such policy been assessed before? Please provide the link to any relevant document(s).

Which stakeholders could benefit from the AI Readiness Assessment tool?

- Governments as a whole and their different agencies
- Civil society
- The public
- Academia
- The private sector
- Developers and engineers

1) How each one of those stakeholders can benefit from the tool?

2) What is the contribution of data scientists to the debate on regulating AI?

Impact Assessment

- Impact assessment is “a component of the policy and programming cycle in public management” and it can play two roles:
- Ex ante impact assessment- prospective analysis of what the impact of an intervention might be, equivalent to business planning
- Ex post impact assessment- part of the evaluation and management of policy cycle, to what extent the intervention corrects the problem it was intending to address
- Impact assessment has a future looking aspect, it is “a description of the cascade of cause and effect leading from an intervention to its desired effect
- Chain of causation theory from intervention to impact, including the changes induced along the chain

Existing forms of impact assessment

- Regulatory impact assessment
- Environmental impact assessment
- Human rights impact assessment
- Privacy impact assessment
- Data protection impact assessment
- Ethical impact assessment

Purpose of EIA

Ethical impact assessment:

Help Member States and other stakeholders to identify and assess benefits, concerns and risks of AI systems, as well as risk prevention, mitigation and monitoring measures.

For project teams (procurement)	<ul style="list-style-type: none">• Identify potential impacts of their projects• Ensure they are aware of and are implementing ethical safeguards throughout the project lifecycle.
For impacted communities	<ul style="list-style-type: none">• Provide impacted communities with an active voice in the development of AI systems• Mitigate any adverse or discriminatory impacts on individuals and groups.• Clearly set out how impacted groups can access remedies
For policymakers and Member States	<ul style="list-style-type: none">• Provide policymakers and Member States with the information they need to assess whether an AI system upholds the recommendation• Clarify requirements for ethical AI procurement in the public sector.

Ethical Impact Assessment – Background

The secretariat surveyed 26 tools that aim to measure the impact of AI:

7 tools were developed by governments

E.g. the Algorithmic Impact Assessment of the government of Canada and the Model AI Governance Framework developed by the Personal Data Protection Commission of Singapore

2 tools were developed by academic institutions

E.g. Aequitas of the Center for Data Science and Public Policy at the University of Chicago

2 tools were developed by intergovernmental organizations

E.g. ALTAI the Assessment List on Trustworthy Artificial Intelligence of the European Commission

8 tools were developed by non-profit organizations

E.g. the EDIA by the Information Accountability Foundation, and the Impact Assessment Tool for Automated Decision Making Systems in the Public Sector by AlgorithmWatch

1 suite of standards was developed by a standardization body

By the IEEE (IEEE P7000J series of standards projects)

6 tools were developed by private companies

E.g. the AI Fairness Checklist of Microsoft, the Model Cards of Google, and AI Fairness 360 of IBM



Gaps in Existing Tools

Collated from a survey by UNESCO of 26 existing tools

Scope:

- Many tools are **too general**
- Impact on the **environment** and **gender equality** rarely covered

Project lifecycle:

- Most tools do not concentrate on the earlier stages of the AI life cycle. Question of “**whether automation is the right solution**” is **rarely addressed**
- Most tools are aimed at **project teams**: no clear room for **stakeholder involvement**

Outputs:

- The final output of the tools is **often not published** to the public or interested parties
- Conducting impact assessment for AI is mainly **voluntary**
- Outputs from the tools are **rarely actionable enough**

Scoping Questions

PROJECT DESCRIPTION & DECISION TO DESIGN

A. PROJECT DESCRIPTION	
1. Description of System	<ul style="list-style-type: none">• Initial description of AI system and project status• Description of the features of the AI system
2. Description of Legitimate Aim:	<ul style="list-style-type: none">• Description of project goal• Description of project context, whether it is part of an existing project
B. DECISION TO DESIGN	
1. Prohibited Use Screening:	<ul style="list-style-type: none">• Social scoring• Mass surveillance• Possibility of outcomes which are irreversible/difficult to reverse• Possibility of life & death decisions
2. Establishing Proportionality	<ul style="list-style-type: none">• Unique development or adaptation• Consideration of non-algorithmic options and different AI methods• Scope of project• Efficacy of project (<i>ex-post</i>)

Scoping Questions

C. MULTI-STAKEHOLDER GOVERNANCE

C. MULTISTAKEHOLDER GOVERNANCE

1. Project Governance

- Decision-making authority attribution
- Project team: roles & responsibilities
- Positionality reflection

2. Multistakeholder Engagement

- Stakeholder engagement plan, including:
 - Impacted stakeholders
 - Stakeholder involvement at different stages
 - Objectives
 - Resources available
 - Modes of engagement
 - Activities for identifying potential impacts and mitigation plans

Implementing the Principles: UNESCO APPROACH

Project teams will have to answer 4 sets of questions for each principle from the UNESCO Recommendation.

Procedural
Questions

Identifying Positive
Outcomes

Identifying Negative
Impacts

Mitigation Strategies

- Safety and Security
- Fairness, Non-discrimination
- Diversity, Inclusiveness and Gender
- Sustainability
- Privacy and Data Protection
- Human Oversight and Determination
- Transparency and Explainability; Accountability and Responsibility
- Awareness and Literacy

Implementing the Principles:

Examples of Procedural Questions for Privacy and Data Protection

- *What types of personal data does the AI system have access to?*
- *Where are the data that the system processes is coming from?*
- *Is the data minimization principle being applied?*
- *Is sensitive data being subjected to different processing standards?*
- *Does the system actively link between different data bases?*

Implementing the Principles:

Mapping Out Positive Impacts

What are the prospected <u>positive impacts</u> of the system on <u>[PRINCIPLE]</u>?	<u>Who</u> benefits from this outcome? Please describe.	Please rank the <u>significance level</u> of this outcome.
<i>E.g. [sustainability]</i> The AI system will help to detect certain diseases in crops in early stages.	<i>E.g. Primary: Farmers of crops Secondary: Communities reliant on these food sources</i>	<i>Based on the impact's:</i> <ul style="list-style-type: none">• <i>Magnitude</i><ul style="list-style-type: none">• <i>Scale</i>• <i>Scope</i>• <i>Likelihood</i>

Implementing the Principles:

Mapping Out Negative Impacts

What are the prospected <u>negative impacts</u> of the system on [PRINCIPLE]?	Please rank the <u>severity level</u> of this impact.	<u>Mitigation Plan</u>: To what extent do the procedural safeguards described above mitigate this impact? What <u>additional mitigation and redressal strategies</u> will you need to implement to combat this potential harm?
<i>E.g. [sustainability] The AI system will require approximately XX kWh and has a projected carbon footprint of XX g CO2e</i>	<i>Based on the impact's:</i> <ul style="list-style-type: none">• Severity<ul style="list-style-type: none">• Scale• Scope• Remediability• Likelihood	<i>Project teams to indicate</i> <ol style="list-style-type: none">1. Extreme cases that would warrant immediate stopping of the AI system.2. Cases that would warrant investigation and redressals, and the time it would take for redressal.

Questions

- What are the advantages of disadvantages of an impact assessment tool that is policy oriented?
- How this tool could complement the technical tools?
- Do you think developers would benefit from going through such a policy based exercise?



unesco

Social and Human Sciences

Thank You!

Doaa ABU ELYOUNES

dabuelyounes@sjd.law.harvard.edu

