



# OUTCOMES OF THE POLICY SESSION

---

## AISIS 2019

MEXICO CITY  
25 OCTOBER 2019



# CONCEPT OF THE POLICY SESSION WITHIN AISIS

Concerns pertain to governance, ethical and societal issues.

- **Transparency and explainability**
  - 'Black box' issue
- **Security**
  - Robustness and dependability of AI systems
- **Privacy**
  - not infringing inappropriately on the privacy of users or third parties
- **Broad questions of governance**
  - Ensuring that AI systems ultimately operate in ways that align with a broad range of societal preferences
- **Technology sometimes moves faster than policy**
  - Capabilities are needed such that policies and institutions anticipate and adjust to changing circumstances.



# TWO PANELS



## Panel 1: Artificial Intelligence strategies – preparing for the future



## Panel 2: New policy ideas – how can we harness AI for societal progress?



# TWO PANELS



- Introductory remarks by Min. J. Arrieta, Minister of the Foreign Service and A. Paic

## Panel 1: Artificial Intelligence strategies – preparing for the future

- Andreas Hartl, Head of Division on AI and Data Economy, German Ministry of Economy
- Ana Valcarel Orti, INRIA and Co-ordination of the French national plan of research in AI
- Cristina Martínez Pinto, AI for Good Lab Director at CMINDS (NGO)
- Estelle Parker, Australian Embassy in Mexico
- Ashley Casovan, Executive Director of [AI Global](#), formerly Director of Data Architecture and Innovation, Government of Canada

## Panel 2: New policy ideas – how can we harness AI for societal progress?

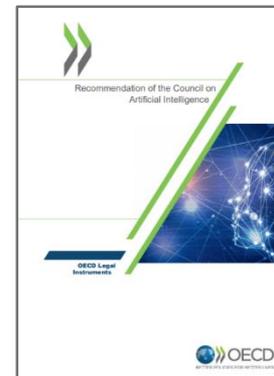
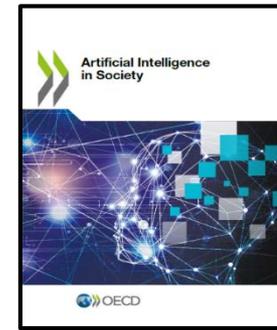
- Julien Chiaroni from General Secretariat for Investment (Office of the Prime Minister, France)
- Mirjana Stankovic, Vice-President, Emerging Technologies at Tambourine ventures
- Ashley Casovan, Executive Director of [AI Global](#),
- Cédric Bourrasset, Artificial Intelligence product manager at ATOS
- Prof. Andrea Bertolini, Dirpolis Institute and University of Pisa
- Andrea Escobedo Lastiri, IBM
- Christian Hernandez, Intelligent Computing Dept., Huawei,



# OECD WORK ON ARTIFICIAL INTELLIGENCE



- **G7 ICT Ministerial meeting in Japan**, Apr 2016
- **OECD events in 2016 and 2017**, incl. *“AI: Intelligent Machines, Smart Policies”*, Oct 2017
- **AI expert group at the OECD (AIGO)**  
Sep 2018-Feb 2019
  - scoped principles to foster trust in AI
  - multi-stakeholder and multi-disciplinary
- **OECD Recommendation on AI**, MCM, May 2019
  - OECD-wide consultations
  - adopted by 42 countries (36 Members + 6 others)
  - basis for **G20 AI Principles**, June 2019
- **Measurement and analytical work: Report on AI in Society**, June 2019



# OECD AI PRINCIPLES (PILLAR 1 OF OECD.AI)



- Adopted May 2019 by 42 governments: 36 OECD + 6 partners and the basis for [G20 AI Principles](#) June 2019.
- Goal: foster policy ecosystem for trustworthy AI that benefits people & planet

## 5 Values-based Principles

	Inclusive growth, sustainable development & well-being
	Human-centred values & fairness
	Transparency & explainability
	Robustness, security & safety
	Accountability

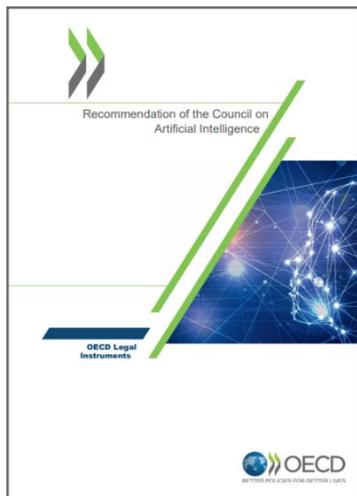
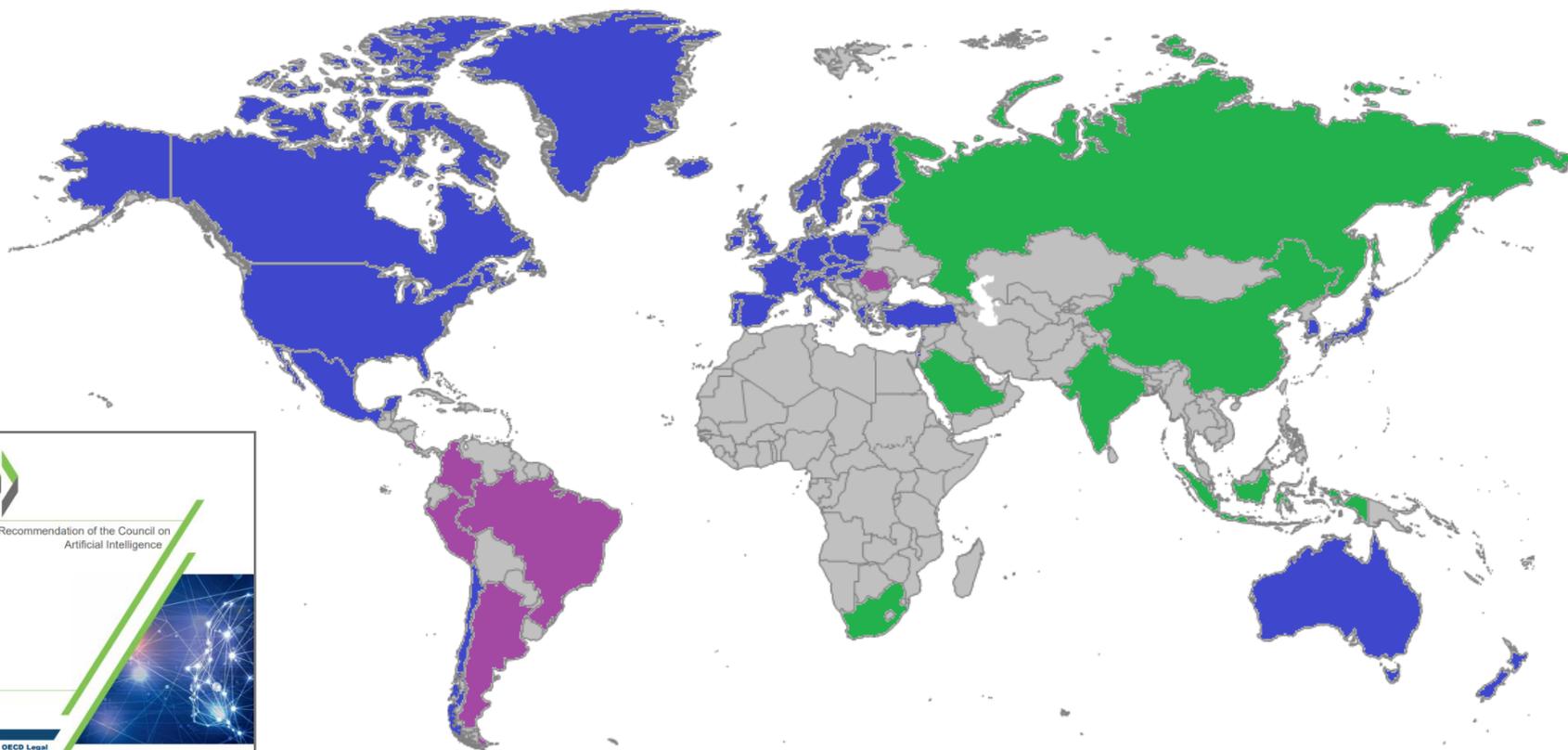
## 5 Recommendations for policy-makers

	Investing in AI research and development
	Fostering a digital ecosystem for AI
	Shaping an enabling policy environment for AI
	Building human capacity and preparing for labour market transformation
	International co-operation for trustworthy AI

# GOVERNMENTS THAT HAVE ADHERED TO THE OECD OR OECD-BASED G20 AI PRINCIPLES



 OECD members  Adherents  G20 principles, based on OECD



# NEXT: FROM PRINCIPLES TO IMPLEMENTATION

## OECD.AI, THE OECD'S AI POLICY OBSERVATORY



- An inclusive online hub for AI information, evidence and policy options

### 1 AI Principles

- » OECD AI Principles
- » Practical implementation guidance



### 2 Policy areas

- » Jobs, skills, health, transport and more



### 3 Trends & data

- » OECD metrics & methods
- » Live data from partners



### 4 Countries and Initiatives

- » National strategies policies
- » Stakeholder initiatives



Multi-disciplinary



Evidence-based



Global multi-stakeholder partnerships

# THE MEXICAN PERSPECTIVE ON AI



## AI needs to be inclusive - 2030 Agenda motto: 'leave no one behind/outside'

- AI to help strengthen democracy and reduce digital gap
- Digital Diplomacy 2.0
- Use of AI and Blockchain in humanitarian and emergency situations (prioritise rescue efforts)
- Underskilled jobs are challenged by AI but new jobs never thought of are emerging

## Global challenges brought by exponential technological change (UN res. 72/242 and 73/17)

- How to advance the 2030 Agenda with the help of new technologies
- Digital Co-operation report, co-chaired by Jack Ma and Melinda Gates

## Development of the National strategy for Artificial Intelligence in a 5-Helix approach

- National Development Plan priorities: rule of law, welfare, economic development, gender equality, inclusion, anti-corruption
- Participation: CONACYT research centers, civil society, business (Microsoft, Google), World Economic Forum
- Mexican diaspora involved: Solidarity with vulnerable migrants / cryptocurrency remittances
- Participation in drafting the OECD AI Principles



# AI STRATEGIES – FRANCE, GERMANY



## AI Made in Germany

- Global budget 3 billion Euros by 2025
- Priorities:
  - Global excellence in R&D
  - Application of AI
  - Ethical issues
- Fields of action
  - Knowledge transfer to SMEs through 26 Centres of Excellence
  - Funding measures: use of AI within existing instruments, or develop quicker new funding
  - International research cooperation (F-D, EU)
  - Systemic approach to software & processor
  - Boost stock of data
  - New professoral chairs for AI
  - Innovation competitions (10 M€)

## AI for humanity (France)

- 1.5 billion Euros for 5 years
- Priorities
  - Bet on French talent (700 M€): train and attract best global talents
  - Dissemination of AI to economy & society through PPP and data sharing
  - Establish an ethical framework
- Strong international component
  - OECD, G7, G20, EU
- National research programme for AI
  - 4 interdisciplinary institutes (3IA)
  - Research and teaching chairs & PhD pgm in AI
  - Priority calls for AI, including a FR-DE-JPN
  - Calls for research-industry common labs
  - Jean Zay supercomputer (17 petaFLOPS)

# AI POLICIES – AUSTRALIA, CANADA



## Australia's Tech Future

- AUD 30M to catalyse AI innovation
- Priorities:
  - People: skills and inclusion
  - Services: digital government
  - Digital assets: digital infrastructure and data
  - Enabling environment: cyber security and regulation
- Building trust in use of data
  - Consumer data right
  - Development of an ethics framework for use of data for AI and ML (150 submissions)
  - Streamline sharing of data (including across borders) while preserving privacy and confidentiality

## Directive on Automated Decision-Making (Canada)

- Objective
  - Reduce risks to Canadians and federal institutions,
  - More efficient, accurate, consistent, and interpretable decisions made by Automated Decision Systems.
- Expected results
  - Data-driven, responsible, procedurally fair decisions.
  - Impacts of algorithms are assessed and negative outcomes reduced
  - Data and information on the use of Automated Decision Systems are made available to the public, where appropriate.

# AI – THE PROMISE



What opportunities do you think will be created for you through using AI?



**38%**

think robots can do certain types of work **better** than their boss can

**64%**

would trust a robot **more** than their manager

**50%**

have asked a robot for advice **instead** of their boss

# AI – THE PROMISE (AUSTRALIA)



**Successful adoption of digital innovation can generate AUD315 billion in gross economic value** for Australia over the next decade [AlphaBeta (2018), *Digital Innovation: Australia's \$315B Opportunity*].

- In **manufacturing**, AI is driving productivity to improve R&D, shorten design time, reduce waste and improve production and maintenance.
- In the **space industry**, the infusion of AI technologies will help increase efficiency of space communication networks and help solve the challenges of long haul space travel.
- In **health**, the benefits of AI are being realised through diagnosis support, large scale data analysis and drug discovery.
- In **agriculture**, AI and automation are already being used to identify and eradicate weeds through visual recognition and more targeted use of herbicides.
- And in some **Australian beaches**, AI and drone technology has even been used to identify sharks as opposed to dolphins, so lifeguards have plenty of time to warn beachgoers to get out of the water.

# AI 'BLACK BOX' – WHAT CAN GO WRONG?



## Bias Amplification – issues of fairness and discrimination

- Risk assessment for criminal prosecution – bias against coloured inmates.
- Women less likely to get Google's ads for jobs >200k\$

## Accountability

- Who develops and profits from AI, and who suffers the consequences?

## Reproducibility issues

- Source code not enough; training data needed; workflow may not be repeatable

## Data-privacy paradox

- Bias can be reduced by training on more data, but this may cause risk of privacy breaches

## Cultural issues

- Many training data sets are created in Silicon Valley and reflect its norms and values
  - E.g. Asian people labeled as 'blinking'

## Towards a 'surveillance capitalism'?

- Facial recognition threatens privacy, can also be used to infer emotions, moods, guilt...

# EXPLAINABILITY OF AI



## Is explainability relevant?

- Yes, if it is bringing an important decision (e.g. attribution of welfare in the UK, or release from jail on parole, credit scoring)
- No, in other cases: Knowing how songs are shuffled and proposed in the app

## What does it depend on?

- Source code
- Training data
- Reproducibility is an issue

## What can be done?

- Simplify algorithms whenever possible. Whether CNN or decision trees, the more complex the less explainable
- Deploy “ethics by design” approach: human rights-centric, incorporating substantive, procedural, and remedial rights
- Disallow use of AI in cases when explainability is important and insufficient?
- Allow users to contest the decisions of AI – contestability
- Standards are needed

# DATA – THE CRUCIAL RESOURCE



**AI is only as useful as the quality and volume of data it is trained on.**

**Many AI-intensive start-ups master the technology but are constrained by data access.**

**Ways to enhance data availability such that AI-intensive firms can work on a broad set of problems relevant to the public interest:**

- Adequate data ownership, data access and data control laws
- Adequate information and cybersecurity laws
- Adequate IP laws
- Overhaul of the current data localization laws that restrict the flow of their citizens' data beyond their borders (so-called digital protectionism).
- Promoting the concept of open data in the context of provision of public services. (data should be public good – it is non-rivalrous but it is excludable)
- Data-specific skills are needed – data management skills for scientists, data standards skills for data stewards, and data literacy for users
- Adequate business models for sharing
- Personal incentives (recognition, data citation for researchers, payment for personal data)
- Infrastructure (repositories, fiber)

# DATA-SPECIFIC CHALLENGES FOR AI START-UPS IN DEVELOPING COUNTRIES



- **Low “datafication” of the economy** (government records and archives may not be digitized)
- **Limited data talent pool.** SMEs compete for scarce skilled labor against larger firms, local and foreign.
- **Restrictive data policies** (localization, poorly developed privacy & consumer protection laws)
- **A higher unit price for data**
- **Limited access to financial resources**, difficult to invest in new technologies
- **Limited awareness** of the opportunities offered by digitized business and operations.
- **Infrastructure constraints** connectivity issues, particularly outside major urban centers
- **Lack of trust in AI due to cyber risks.**
- **Lack of data governance frameworks in many countries, or a lack of awareness of them or an understanding of how to comply.** Complex regulations and policies pertaining to data protection, data privacy, and cross-border flows of data
- **Lack of appropriate (open) standards and fear of vendor lock-in, often due to proprietary solutions, can be strong barriers to adoption.** SMEs, which often lack the negotiating power and know-how about advanced ICTs such as cloud computing, data analytics, and the IoT.

# DOES INDUSTRY SELF-REGULATION WORK?



## Google acquisition of Deep Mind (2014)

- Google agreed to set up an external review board that would ensure the lab's research would not be used in military applications or otherwise unethical projects. But five years later, it is still unclear whether this board even exists.

**Google, Microsoft, Facebook** and other companies have created the **Partnership on A.I.** to guide the practices of the entire industry. But these operations are largely **toothless**.

The most significant changes have been driven by **employee protests**, like the one at Google, and pointed research from academics and other independent experts. After Amazon employees protested **the sale of facial recognition services to police departments** and various academic studies highlighted the bias that plagues these services, **Amazon and Microsoft called for government regulation in this area.**

In July 2019, 8 French large industrial companies signed the **Manifesto for AI for Industry**, which proposes an ecosystem approach, co-ordinating the development of public-private partnership for AI coherent with the vision of the 'AI for Humanity' strategy

# AN NGO APPROACH (AI GLOBAL) RESPONSIBLE AI CERTIFICATION MARK



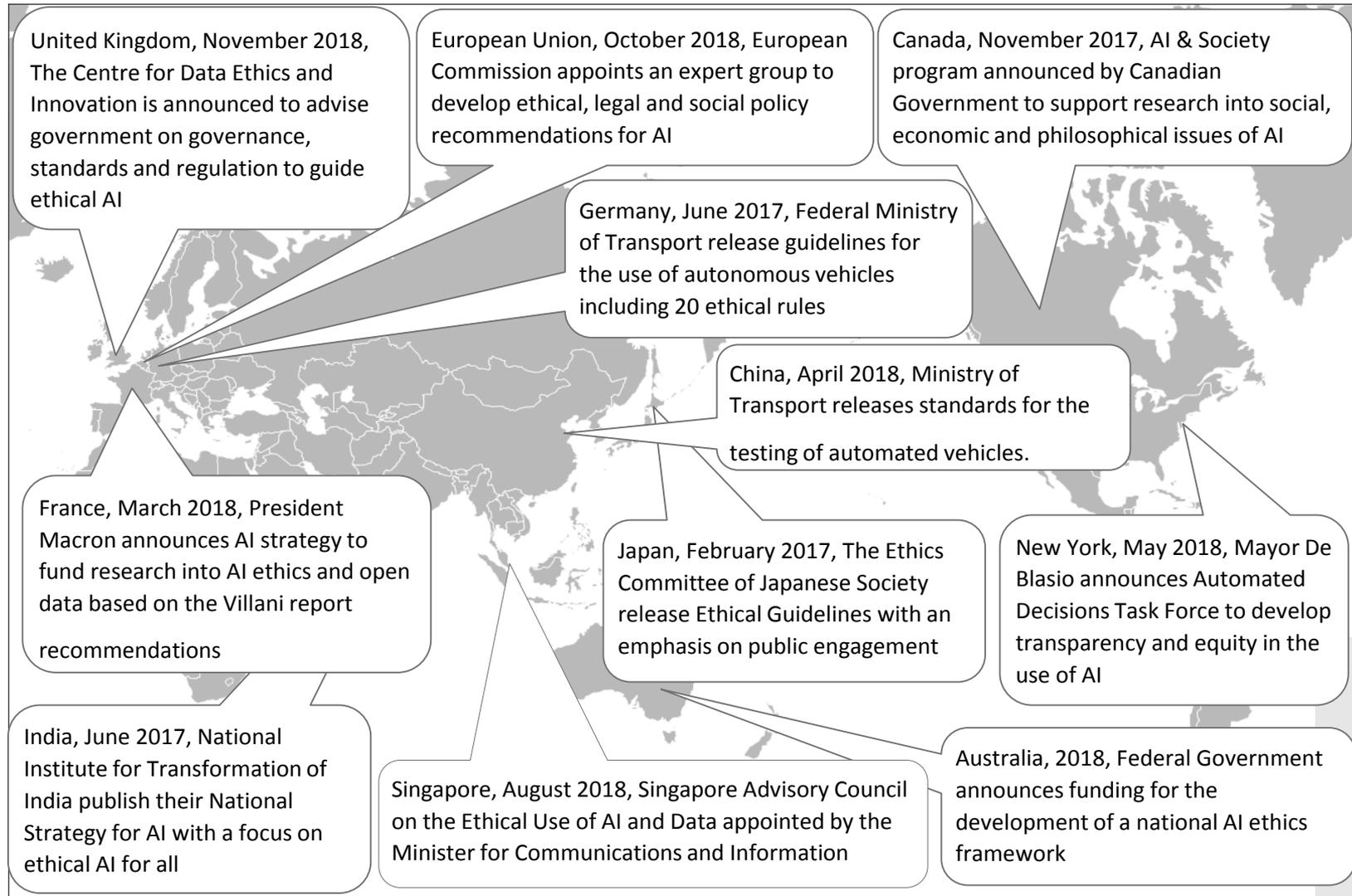
## Objective

- Build AI systems in a responsible way -- preventing data bias, increasing fairness, and avoiding the potential of rogue algorithms
- Inform the consumer about safe applications (not selling your data, not proferring fake news)

## The Responsible AI check

- A certification mark will provide the public with an easy to recognize symbol that will provide them with trust and assurance in the product that they are using.
- *What the Check is:*
  - A certification mark akin to a trademark, bringing together best practices and ways to evaluate existing standards and models in a comprehensive way
  - This work can help to inform future certification and standards development
  - A clear mark to put on products for companies to differentiate themselves from their competitors.
  - An easy to recognize symbol for consumers to help make decisions.
- *What the Check isn't:*
  - A tool to question ethics (what is good and bad);
  - A certification for whole companies
  - A method to define new technical standards

# THE RACE FOR ETHICAL FRAMEWORKS



# AN EXAMPLE ETHICAL FRAMEWORK



## Core principles for AI

**1. Generates net-benefits.** The AI system must generate benefits for people that are greater than the costs.

**3. Regulatory and legal compliance.** The AI system must comply with all relevant international, Australian Local, State/Territory and Federal government obligations, regulations and laws.

**5. Fairness.** The development or use of the AI system must not result in unfair discrimination against individuals, communities or groups. This requires particular attention to ensure the “training data” is free from bias or characteristics which may cause the algorithm to behave unfairly.

**7. Contestability.** When an algorithm impacts a person there must be an efficient process to allow that person to challenge the use or output of the algorithm.

**2. Do no harm.** Civilian AI systems must not be designed to harm or deceive people and should be implemented in ways that minimise any negative outcomes.

**4. Privacy protection.** Any system, including AI systems, must ensure people’s private data is protected and kept confidential plus prevent data breaches which could cause reputational, psychological, financial, professional or other types of harm.

**6. Transparency & Explainability.** People must be informed when an algorithm is being used that impacts them and they should be provided with information about what information the algorithm uses to make decisions.

**8. Accountability.** People and organisations responsible for the creation and implementation of AI algorithms should be identifiable and accountable for the impacts of that algorithm, even if the impacts are unintended.

# REGULATION TAKEAWAY (A. BERTOLINI)



- **Bottom up approach - need to determine**
  - (i) what regulation is already applicable,
  - (ii) what the consequences are of their application and if problems emerge as a consequence thereof,
  - (iii) how it could be improved, by modifying legislation (when needed).
- In this case, **regulation empowers technological development** whose purpose should always be that of improving the human condition (**anthropocentric perspective**).
- **Regulation differs from ethics:**
  - The legal ordering is grounded on fundamental values that need to be respected. Typically recalled by constitutions.
  - Ethical theories that might be deemed conflicting with such principles could not inform policy decisions. We might say that regulation pre-empts ethics.
- **Ethical values vary**, even more so their enactment through regulation.
  - International debates are useful and should be encouraged
  - **Solutions** are expected to be found and adopted **at regional (eg.: European) level**.
- **Diversity in regulation** and a multiplicity of alternative models is not to be feared (if not excessive), it always existed and benefitted – through comparative analysis – the advancement of all legal systems.

THANK YOU



Contact

[Alan.Paic@oecd.org](mailto:Alan.Paic@oecd.org)