



The Role of Artificial Intelligence in Building Responsible Open Science Infrastructures

Natalie K Meyers, MA MLIS
On behalf of RDA AIDV-WG

9.27.2023

These Slides:



1

Artificial Intelligence & Data Visitation Working Group (AIDV-WG)



An update from recent work of the European Open Science Cloud-Future Project (EOSC) & Research Data Alliance (RDA) AIDV-WG

Co-chaired by Francis P. Crawley (BE), Perihan Elif Ekmekci (TR), Claudia Bauzer Medeiros (BR), Natalie Meyers (USA)



2

2



EOSC-Future/RDA Artificial Intelligence & Data Visitation
Working Group (AIDV-WG)

1. A survey related to current ethical, legal, policy, and societal concerns related to generative AI
2. Guidance on legal considerations for AI and DV
3. Guidance for Informed Consent in AI and DV
4. Guidance for ethics committees reviewing AI and DV
5. **AI Bill of Rights**

3

3

EOSC-Future/RDA Artificial Intelligence & Data Visitation Working Group (AIDV-WG)

WP5 AI Bill of Rights

Mission: Produce AI Bill of Rights communiques for EOSC Future and RDA that promote fundamental human rights and advance trust in AI and federated systems for Open Science.

Team: Eyiuche Ezigbo(NG), Shiny Martis B(FR) Natalie Meyers(US), Ronit Purian-Lukatch(IL), Yeyang Su (CN)

4

EOSC-Future/RDA Artificial Intelligence & Data Visitation Working Group (AIDV-WG)

We hold as important tenets of our work that:



Adoption of AI Bills of Rights/policies within member organizations can create mutual trust and benefits for all.

And that

Attention to AI Governance can improve the longevity and relevance of research environments and communities


5

WP5 AI Bill of Rights

Report to EOSC Future

AI Bill of Rights Recommendations & Guidelines for RDA Membership



zotero.org/groups/4922635/aidv-wg/library

Review of Emerging Jurisdictional and Special interest AI Rights & Protections

Jurisdictional

Disciplinary/Niche

National: US, Israel, Brazil

EU AI Act

Development Bank of Latin America

UN Global Agreement

Arts

Education

Rome Call

Inform constituents on Layering and Balancing of Rights & Protections

	Protected Party/Parties	Specific Right or Protection
consent	x	Lorem ipsum dolor sit
copyright	x	Vitae tortor condimentum
provenance		Mattis vulputate enim nulla
authorship	x	Consequat mauris nunc congue

Matrix of Rights & Protections

6

WP5 Review of Emerging Jurisdictional and Special Interest AI Rights & Protections

Shared Library



Tagged by:

- Country
- AIDV SubTeam
- Keyword

Categorized by:

- Jurisdictional Rights Declarations & Acts
- Disciplinary & Special Interest Rights & Protections
- Education
- Healthcare/Medicine
- Legal
- Privacy/Surveillance
- Generative AI/Publishing, Authorship, Copyright
- Related News Stories
- Risk
- UN SDGs

zotero.org/groups/4922635/aidv-wg/library



7

WP5 Regional, country and municipal level policies and foci

Shared Library



Tagged by:

- Country
- AIDV SubTeam
- Keyword

Examples from:

OECD, China, Africa, Nigeria, the Middle East and North Africa region, Israel, The European Union, the Council on Europe, Italy, the United Kingdom, the United States, California, New York City, Canada, Latin America, Australia, and Taiwan.

zotero.org/groups/4922635/aidv-wg/library



8

Australia: Artificial Intelligence Ethics Framework

- **Human, societal and environmental wellbeing:** AI systems should benefit individuals, society and the environment.
- **Human-centred values:** AI systems should respect human rights, diversity, and the autonomy of individuals.
- **Fairness:** AI systems should be inclusive and accessible, and should not involve or result in unfair discrimination against individuals, communities or groups.
- **Privacy protection and security:** AI systems should respect and uphold privacy rights and data protection, and ensure the security of data.
- **Reliability and safety:** AI systems should reliably operate in accordance with their intended purpose.
- **Transparency and explainability:** There should be transparency and responsible disclosure so people can understand when they are being significantly impacted by AI, and can find out when an AI system is engaging with them.
- **Contestability:** When an AI system significantly impacts a person, community, group or environment, there should be a timely process to allow people to challenge the use or outcomes of the AI system.
- **Accountability:** People responsible for the different phases of the AI system lifecycle should be identifiable and accountable for the outcomes of the AI systems, and human oversight of AI systems should be enabled.

Essential Elements of Responsible AI

1. AI & Acceptable Use Policies: Formalize the definition of AI in your community, document key roles and responsibilities related to AI, outline the acceptable and unacceptable uses of AI for your community.
2. AI Working Group (oversight committee): Form an oversight committee whose function it is to oversee AI systems, allocate resources, develop and maintain AI governance policies and procedures for your community.
3. AI Training and Education: Training and change management are integral to the successful operationalization of AI and AI policy. Offer learning opportunities on an Introduction to AI, on Ethical considerations for AI, as well as training on Development of AI strategy, Effective AI governance and Risk management, in addition to training on product/project management for how to run effective AI projects, and how to use technical tools in an AI landscape/cyberinfrastructure.
4. AI Impact Assessment
5. Ongoing Monitoring: Consider, will you have a Human in the loop, Human over the loop, Human out of the loop monitoring plan?
6. Transparency, Notice & Disclosure
7. Vendor Due Diligence
8. Test AI systems for Bias : Identify different types of biases: Systemic, Human, computational . Mitigating bias requires action at each stage of the AI lifecycle

Impact Assessment

Because it is essential to govern systems that can potentially cause certain harms and it is essential to assess the Impacts of AI Systems, we recommend the below Impact assessment elements as essential to prioritize as parts of a responsible AI strategy for any group: (organization, company, university, hospital, jurisdiction/community):

- Determine what will be the AI system impact assessment process
- Documentation of AI system impact assessments
- Assessing AI System impact on individuals and groups of individuals
- Assessing societal impacts of AI systems
- AI system impact assessment process

11

AI Alignment

AI Alignment is an AI safety research/process that aims to ensure AI systems achieve desired outcomes. If you're creating or implementing an AI or AI policy you need to be able to define the system or policy's desired outcomes, and articulate how you'll document the processes for monitoring, measuring and logging where you achieve and fall short of achieving desired outcomes/benefits of AI or AI policy. Any group creating or implementing an AI or AI policy should:

- **Consider auditable detection for bias.**
Bias detection exposes cases where decision logic is in violation of agreed upon ethics.
- **Consider auditable detection for privacy violations.**
Privacy leak detection exposes cases where decision logic is in violation of agreed upon ethics ensuring privacy.
- **Have a control and escalation process** to assign responsibility for investigation, a process for investigation, a process for notification and disclosure of violations, and a process for reconciliation/restitution to impacted individuals and groups to mitigate and correct consequences of non-aligned AI .

12

AI Governance can be approached from human rights, risk based, or safety based motivations or a combination of all three.

Most jurisdictions have considered regulating Artificial general intelligence (A(G)I) based on the risks such models may give rise to during their development, deployment and dissemination and, correspondingly, such governance strategies put safeguards along the way.

“Risk-based regulation—like regulatory excellence more generally—is not a merely technical enterprise. It requires not only technical competence, but also principled decision-making, transparency, careful attention to empirical evidence and on-the-ground implementation”.

An exclusively risk-based analysis and regulatory design approach, precludes questioning, at the first place, whether or not one should apply or be subject to use of A(G)I in certain application domains.

Recent development in A(G)I raises fundamental questions regarding the essence of being human and our relationship with AI and with one another in the future. At most, risk assessments inform regulators’ decisions; they do not provide a full basis for them . . . Risk assessment provides scientific or empirical answers about probabilities, hazards, and their distribution; it does not supply the policy principle or normative reason needed to make regulatory or risk management decisions about these hazards



Coglianese, C. (2019) *What Does Risk-Based Regulation Mean?* The Regulatory Review.



13

OECD: “Ethics Guidelines for Trustworthy AI”

The OECD guidelines for ethical AI(2019) – promote seven governance principles: (1) human agency and oversight, (2) technical robustness and safety, (3) privacy and data governance, (4) transparency, (5) diversity, nondiscrimination and fairness, (6) environmental and societal well-being, and (7) accountability.

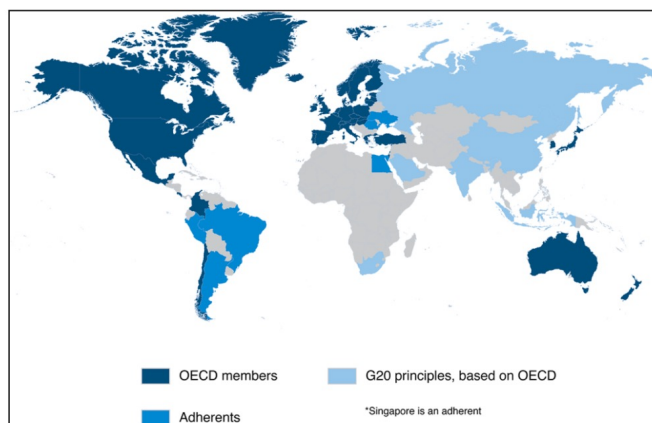


Figure 1 “Governments that have Committed to the AI Principles” © 2023 OECD. All rights reserved



14

14

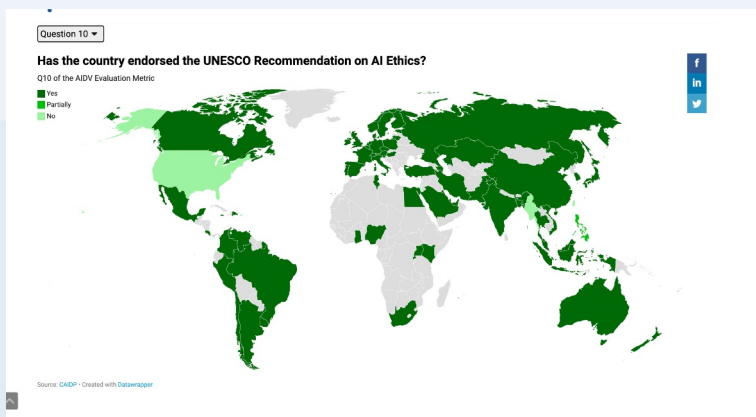
Global Partnership on Artificial Intelligence (GPAI)

GPAI (2020) brings together experts from science, industry, civil society, governments, international organisations and academia to foster international cooperation related to AI and convenes working groups on both Data Governance and Responsible AI.

The GPAI countries have put climate action and biodiversity preservation at the top of their agenda, positing that as a general-purpose technology, AI can be harnessed responsibly to accelerate positive environmental action. In 2021 the WG released an action-oriented set of recommendations entitled “**CLIMATE CHANGE AND AI: Recommendations for Government Action**” to guide policy makers developing AI related climate action strategies.

15

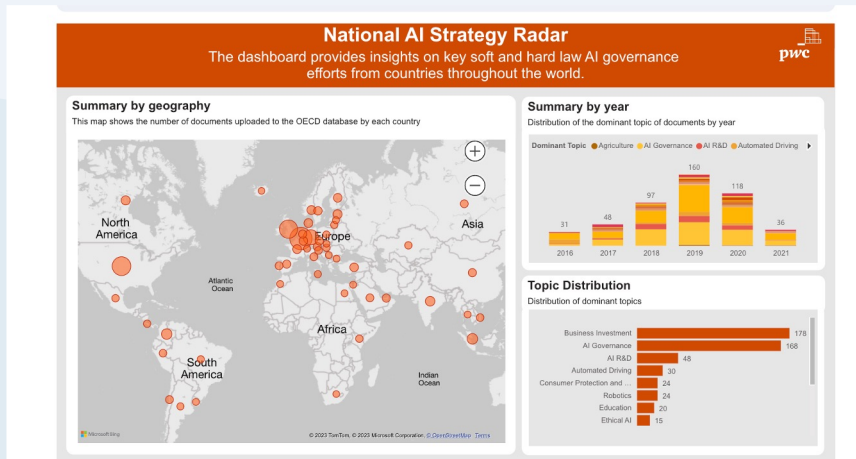
UNESCO Recommendations and More



The Artificial Intelligence and Democratic Values 2022 AI Index covers 75 countries. Relevant regional frameworks have been systematically taken into account in the country reports. The 2022 report is the result of the work of more than 200 AI policy experts in almost 60 countries (<https://www.caidp.org/reports/aidv-2022/aidv-maps/>)

16

National AI Strategy and More



The Future of Life AI National Strategy Radar dashboards
<https://futureoflife.org/resource/ai-policy/>

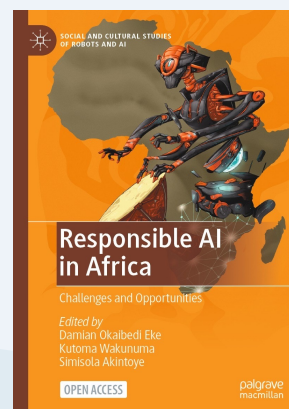
How to use:

- Clicking on one of the countries on the map will display the year and topic distribution of that country.
- Clicking on a topic in the bottom right frame will display the distribution of that topic on the map and the bar chart.

17

Africa

The African Union (AU) established a working group on Artificial intelligence with a mandate to study the creation of a common African stance on AI, the development of an Africa-wide capacity-building framework, and to establish an AI think tank to assess and recommend projects to collaborate on in line with the AU's Agenda 2063 and the United Nations ("UN") Sustainable Development Goals (SDGs).



Responsible AI in Africa: Challenges and Opportunities (2023) is an open access collection edited by Damian Okaibedi Eke, Kutoma Wakunuma, and Simisola Akintoye that offers a good overview of regional progress and challenges.

18

China

Data Security Law of People's Republic of China, and the Personal Information Protection Law of the People's Republic of China were released and put into effect in the year of 2021

A series of efforts have been made to steer ethical governance of AI development at the national level in China by entities such as the Expert Committee for Ethical Governance of the Next Generation AI and National Informational Security Standardisation Technical Committee. These efforts have led to the publication of *the Governance Principles for a New Generation of Artificial Intelligence: Develop Responsible Artificial Intelligence* in 2021.

Eight guiding principles are 1) harmony and friendliness, 2) fairness and justice, 3) inclusiveness and sharing, 4) respect for privacy, 5) security and controllability, 6) shared responsibility, 7) open cooperation and 8) agile governance.

Standards have also been developed in both key technical areas and in ethical terms, such as guidelines for prevention of ethical & security risks of artificial intelligence in 2021 and guidelines for the standardisation of artificial intelligence ethical governance in 2023.

19

EU AI Act (aka AIA)

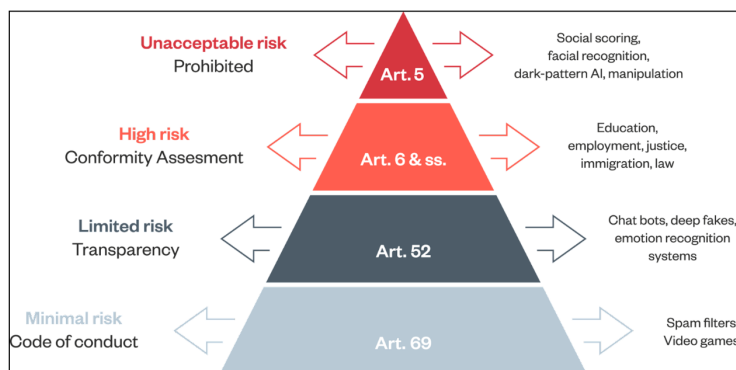


Figure 2 Pyramid of Risks.

Lilian Edwards. (2022). "The EU AI Act proposal". Ada Lovelace Institute. CC-BY-4.0

The draft AI act distinguishes between AI systems posing

- (i) unacceptable risk,
- (ii) high risk,
- (iii) limited risk,
- (iv) and low or minimal risk.

20

UK Algorithmic Transparency & Framework

The Cabinet Office's Central Digital and Data Office has developed an **algorithmic transparency standard** for government departments and public sector bodies with the Centre for Data Ethics and Innovation **making the UK one of the first countries in the world to develop a national algorithmic transparency standard.**

The standard has been piloted by some public sector organisations and will be further developed based on feedback.

The UK framework is underpinned by 5 principles to guide and inform the responsible development and use of AI in all sectors of the economy. These five principles will be issued on a non-statutory basis and implemented by existing regulators :

- Safety, security and robustness
- Appropriate transparency and explainability
- Fairness
- Accountability and governance
- Contestability and redress

21

US Initiatives & Regulation on Artificial Intelligence

- [National AI Research Resource \(NAIRR\)](#) Report: Implementation Plan for a National Artificial Intelligence Research Resource **Jan 24, 2023**
- [Blueprint for an AI Bill of Rights](#): Five principles that should guide the design, use, and deployment of automated systems to protect the American public in the age of artificial intelligence. **2022**
- [National AI Initiative Act](#) : To ensure continued U.S. leadership in AI research and development, lead the world in the development and use of trustworthy AI in the public and private sectors, and prepare the present and future U.S. workforce for the integration of AI systems across all sectors of the economy and society **became law on January 1, 2021**
- Food and Drug Administration (FDA): [Clinical Decisions Support Software](#)
- National Institutes of Standards and Technology:
 - [Four Principles of Explainable AI](#) (2021)
 - [Risk Management Framework](#) (2018)
- PROPOSED: [H.R.6580 - Algorithmic Accountability Act](#) / [S.3572](#) (2022)
- [ACR-215 23 Asilomar AI Principles](#): Guiding values for the development of artificial intelligence and of related public policy. Adopted in California (2018)



Notre Dame is also a signatory of the the Rome [Call for AI Ethics](#) (2022)

22

Latin America and Caribbean

The CAF-Development Bank of Latin America and UNESCO signed a letter of intent in 2022 to work together on the implementation of the **Recommendation on the Ethics of Artificial Intelligence (AI) in Latin America and the Caribbean**.

At the country level, **Colombia, Argentina, Brazil, Chile, and Uruguay** have released national AI strategies. The first countries to publish their strategies were Argentina, Colombia and Uruguay, followed by Brazil and Chile in 2021. Brazil, Chile and Uruguay have all incorporated responsible AI principles into their broader AI policies and are due to publish their own AI ethics policies.

Mexico white paper in June 2018 entitled: “**Towards An Ai Strategy In Mexico: Harnessing the AI Revolution**”, followed by IA2030Mx citizen coalition founded by nine institutions from all sectors addressing the need to have a national, multidisciplinary, multisectoral and collaborative exercise to develop an AI action plan which could serve government, academia, civil society and industry. This output was presented as the **Mexican National Artificial Intelligence Agenda** in 2020.

23

All of us can learn to ask the hard questions that will keep solutions working for, and not against, the development challenges we care about.

USAID report explores some of the ways in which ML/AI may fail or be ill-suited for deployment in developing-country contexts. *Awareness of these risks, and acknowledgement of our role in perpetuating or minimizing them, will help us work together to protect against harmful outcomes and ensure that AI and ML are contributing to a fair, equitable, and empowering future*




24



Thank you for attending this webinar on:

The Role of Artificial Intelligence in Building Responsible Open Science Infrastructures

Contact natalie.meyers@nd.edu |  [0000-0001-6441-6716](https://orcid.org/0000-0001-6441-6716)

These Slides: