OECD.Al work promoting interoperability of Al risk management frameworks

IGF Policy Network on AI Meeting #4



Celine Caira Economist/Policy Analyst 18 July 2023



The OECD AI Principles (2019)

5 values-based principles for trustworthy, human-centric Al



 \bigcirc

Benefit People & Planet

Human rights, values & fairness



Robust, secure & safe

Accountable

5 principles for national policies, for AI ecosystems to benefit societies



 $\sum_{i=1}^{n}$

Al research & development



Policy & regulatory environment

Jobs & skills, labour transitions

International cooperation & measurement

https://oecd.ai/ai-principles

Jurisdictions are moving from PRINCIPLES to PRACTICE through various policies (laws, regulations, standards) and programs.



How can we ensure these initiatives are INTEROPERABLE with one another?



What does interoperability mean?

Some examples..

"...the ability of various privacy regimes, or legal frameworks, to work together to facilitate transborder data flows while ensuring the consistent protection of these data"

- OECD "Mapping commonalities in regulatory approaches to cross-border data transfers", 2021

"...a pragmatic arrangement to promote policy coherence in the context of a shifting regulatory environment and multiple privacy frameworks and data regulations (e.g. data localisation requirements)"

- OECD Going Digital Toolkit Policy Note, 2021



OECD work promoting interoperability in Al governance





Going back to basics: OECD AI System Definition (OECD, 2019)



- i) perceives environments through data or input;
- ii) abstracts these perceptions into models;
- iii) uses the models to formulate options for outcomes."











Why classify Al systems? A variety of systems and policy implications





OECD Framework for Classifying AI systems

Each AI framework dimension has its own properties and

attributes...

CONTEXT **PEOPLE & ECONOMIC CONTEXT** - Industrial sector Business function & model Critical function

- Scale & maturity

Al actors include system operators

...and involves specific actors

DATA & INPUT

Provenance, collection, dynamic nature Rights and 'identifiability' (personal data on, proprietary etc.) - Appropriateness and quality

Al actors include data collectors & processors

PLANET Impacted stakeholders Optionality & redress Well-being & environment **Displacement potential**

Actors include end-users & stakeholders

TASK & OUTPUT

 System task (recognise; personalise etc) System action (autonomy level) Combining tasks and action Core application areas (computer vision etc)

AI actors include system integrators

AI MODEL

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

Al actors include developers & modellers



Linking the classification & AI system lifecycle actors

<u>Framework</u> dimensions	People & Planet	Economic Context		Data & Input	Al Model		Task & Output
		Ś					
<u>Actors</u> include	End-users & stakeholders	System operators		Data collectors & processors	Developers & modellers		System integrators
	\bigcirc						
<u>Lifecycle</u> <u>stage</u>	Use or are impact by	Plan & design	Operate & monitor	Collect & process data	Build & use	Build & validate	Deploy



Goal: enable international interoperability in AI risk management by identifying common guideposts

>MAPPING of existing and developing core standards, frameworks and guidelines for AI risk management to the top-level interoperability framework.

- STOCKTAKING of commonalities and differences in concepts and terminology between initiatives and conduct a gap analysis.
- TRANSLATING analysis into good practice to inform development of due diligence guidelines for responsible business conduct in AI.
- ANALYSING the alignment of certification schemes with OECD responsible business conduct (RBC) and AI standards.
- DEVELOPING an interactive online tool to help compare frameworks and navigate existing methods, tools and good practices for identifying, assessing, treating and governing AI risks.



High-level mapping of select risk management frameworks to the Interoperability Framework

OECD			GOVERN		DECINE	400500	TDEAT			
FRAMEWORK	Monitor & review	Communicate	Consult	Document	Embed	DEFINE	A33E33	IKEAI		
OECD DDG	TRACK	COMMUNICATE		EMBED		IDENTIFY & ASSESS		CEASE, PREVENT & MITIGATE	REMEDIATION	
ISO 31000	MONITORING & REVIEW	COMMUNIC CONSULT	ATION & ATION	RECORDING & REPORTING	LEADERSHIP & COMMITMENT	SCOPE, CONTEX, & CRITERIA	SCOPE, CONTEX, & RISK ASSESSMENT CRITERIA		RISK TREATMENT	
NIST AI RMF			GOVERN	MAP MEASURE				MANAGE		
EU AIA	Post-market monitoring system and regular systematic updating	Communication of residual risks, accuracy, conformity, serious incidents	N/A	Documentation, record keeping, traceability	Quality management system	Identify, analyse ar foreseeable r	nd evaluate known and risks, test system	Eliminate, reduce, mitigate and control any risks		
HUDEIRA	Iterative requirements	N/A	Stakeholder Engagement Process (SEP)	N/A	N/A	Context-Based Risk Analysis (COBRA)	Human Rights, Democracy and the Rule of Law Impact Assessment (HUDERIA)	Impact Mitigation Plan (IMP)		
IEEE 7000-21	N/A	Transparency management process	Ethical values elicitation and prioritisation	N/A	N/A	Concept of operations and context exploration	Ethical values elicitation and prioritisation	Ethical requirements definition and ethical risk-based design		
ISO/IEC Guide 51	Validation & documentation	N/A	N/A	Validation & documentation	N/A	Identify user, intended use and reasonably foreseeable misuse / Hazard identification	Estimation / Evaluation of risk	Risk reduction		

High-level mapping of select risk management frameworks to the Interoperability Framework

- Key risk management frameworks generally aligned with the 4 top-level steps of the Interoperability Framework (Define | Assess | Treat | Govern).
- Most of the differences between frameworks relate to the GOVERN function.



OECD	GOVERN					DEFINE	ASSESS	TRFAT	
FRAMEWORK	Monitor & review	Communicate	Consult	Document	Embed	DEFINE	ACCECC		
OECD DDG	TRACK	COMMUNICATE		EMBED		IDENTIFY & ASSESS		CEASE, PREVENT & MITIGATE	REMEDIATION
ISO 31000	MONITORING & REVIEW		ATION & ATION	RECORDING & REPORTING	LEADERSHIP & COMMITMENT	SCOPE, CONTEX, & RISK ASSESSMENT CRITERIA		RISK TREATMENT	
NIST AI RMF	GOVERN					MAP	MEASURE	MANAGE	
EU AIA	Post-market monitoring system and regular systematic updating	Communication of residual risks, accuracy, conformity, serious incidents	N/A	Documentation, record keeping, traceability	Quality management system	Identify, analyse and evaluate known and foreseeable risks, test system		Eliminate, reduce, mitigate and control any risks	
HUDEIRA	Iterative requirements	N/A	Stakeholder Engagement Process (SEP)	N/A	N/A	Context-Based Risk Analysis (COBRA)	Human Rights, Democracy and the Rule of Law Impact Assessment (HUDERIA)	Impact Mitigation Plan (IMP)	
IEEE 7000-21	N/A	Transparency management process	Ethical values elicitation and prioritisation	N/A	N/A	Concept of operations and context exploration	Ethical values elicitation and prioritisation	Ethical requirements definition and ethical risk-based design	
ISO/IEC Guide 51	Validation & documentation	N/A	N/A	Validation & documentation	N/A	Identify user, intended use and reasonably foreseeable misuse / Hazard identification	Estimation / Evaluation of risk	Risk reduction	

Further reading











THANK YOU

celine.caira@oecd.org

FOLLOW US: @OECD-AI