OECD.AI Policy Observatory

THE OECD FRAMEWORK FOR CLASSIFYING AI SYSTEMS

The Framework is currently being tested through a public consultation *(20 May – 30 June 2021).*
To participate, please visit https://www.oecd.ai/classification.

AI is diffusing apace across all sectors of the economy, with many different types of systems raising very different policy opportunities and challenges. The OECD Framework for Classifying AI Systems helps policymakers and others to differentiate between AI systems and assess their potential impact on public policy in areas covered by the OECD AI Principles. The Framework is based on the work of the OECD.AI Network of Experts and the OECD Secretariat.

The Framework classifies AI systems along four dimensions (see Figure 1):

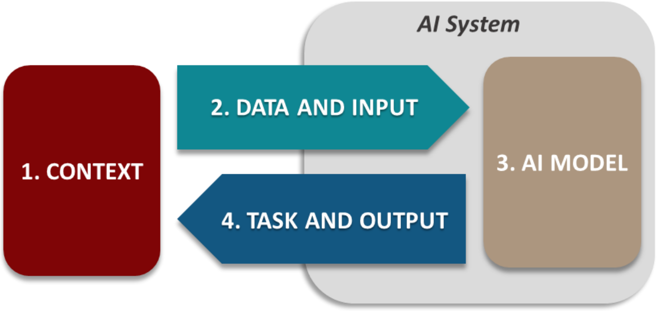
a. The **context** or socio-economic environment in which the system is deployed and used.

b. The **data and input** used by the AI system to build a representation of the environment.

c. The **AI model** is a computational representation of real world processes, objects, ideas, people and/or interactions that include assumptions about reality.

d. The **task and output** refer to the tasks the system performs (*e.g.*, personalisation or recognition) and its outputs, as well as the resulting action that influences the context.

Figure 1. The four dimensions of the OECD Framework for Classifying AI systems



Source: OECD, 2021 (forthcoming)

Each of the four dimensions has distinct properties and attributes – *i.e.* sub-dimensions – (detailed in Table 1). that have potential policy implications for the OECD AI Principles (Figure 2).

Figure 2. The OECD AI Principles

Values-based principles for all AI actors*	Recommendations to policy makers
Principle 1.1. Inclusive & sustainable growth	*Principle 2.1.* Investment in R&D
Principle 1.2. Human rights, privacy, fairness	*Principle 2.2.* Data, infrastructure, compute
Principle 1.3. Transparency, explainability	*Principle 2.3.* Enabling policy and regulation
Principle 1.4. Robustness, security, safety	*Principle 2.4.* Jobs, automation, skills
Principle 1.5. Accountability	*Principle 2.5.* International cooperation

Source: OECD AI Principles, 2019, oecd.ai/ai-principles
** AI actors play an active role in the AI system lifecycle, including organisations and individuals that deploy or operate AI.*

The proposed classification aims to balance simplicity and user-friendliness with useful information. Within each of the four dimensions, core criteria are proposed: these correspond to sub dimensions that are essential to policy and for which information is most readily available. Other criteria that may be more difficult to assess are proposed as optional criteria.

The Framework is being used as the basis for an illustrative risk assessment tool that identifies AI systems that are not low risk based on the criteria in each of the four dimensions.

Table 1. Descriptive summary of the four dimensions of the Framework and sub-dimensions

	Core criteria	*Description*	Optional criteria	*Description*
Context	*Which industry is the system deployed in?*	The industrial sector in which the system is deployed (*e.g.* finance, agriculture).	*Does the system perform a critical function?*	The system performs function or activity of which the disruption would affect essential services.
	For what business function?	The functional areas in which the AI system is employed (*e.g.* sales, customer service, HR).	*How mature is the system?*	The technical maturity of the system can be assessed using Technology Readiness Levels (TRLs).
	How widely is it deployed?	Relates to the number of people affected (*e.g.* pilot project, narrow, broad, widespread deployment).	*What business model does it serve?*	The system can be used for-profit, for non-profit or public services.
	Who are the system's users?	Users can range in competency from AI expert(s) to amateur end-user(s).	*What impact does the system have on well-being?*	The system can impact human well-being factors (*e.g.* job quality, social interactions, civic engagement).
	Who does the system impact?	The system can impact stakeholders such as consumers, workers, government agencies and others.		
	What degree of choice do users have?	Users can be able to correct or /and opt out of the system's output, or not.		
	What impact does the system have on human rights?	System outputs can impact human rights and democratic values (*e.g.* human dignity, privacy, fair trial, safety).		
Data and input	*How are the data and input collected?*	Data and input can be collected by humans or by automated sensors or both.	*What is the data's scale?*	Scale depends on data's dynamic nature (*e.g.*10s of petabytes per s if real-time; 100s of gigabytes if static).
	Where do the data and input come from?	Data and input can come from experts, be provided, and observed, synthetic or derived by the system.	*What is the data's format?*	Data and metadata can come in standardised or non-standardised format.
	How dynamic are the data?	Data can be static, dynamic and updated from time-to-time or real-time.	*What is data's level of appropriateness and quality?*	Includes fitness for purpose, adequate sample size, representativeness, completeness, level of noise.
	Are the data structured or not?	Data can be structured, semi-structured, unstructured or complex structured.	*If they are personal, are the data identifiable?*	If the data are private, they can be more or less identifiable (*e.g.* anonymised; pseudonymised data).
	Are rights attached to the data?	Data can be proprietary (privately held), public (no IP rights) or private (related to individuals).		
AI model	*What type of model does the system use?*	Model is symbolic (i.e. uses human-generated rules), statistical (i.e. uses data) or hybrid.	*Does the model evolve?*	Machine learning models can continue to evolve and acquire abilities from interacting with data, or not.
	How is the model built or trained?	Model building or "training" occurs via human-encoded knowledge or machine-learned knowledge.	*How does the model learn?*	Machine learning models can be trained centrally or on local servers or 'edge' devices.
			How does the system use the model?	The model can be used in a deterministic or probabilistic manner.
Task and output	*What task does the system perform?*	The system performs one or more tasks i.e. functions (*e.g.* recognition; event detection; forecasting).	*Does the system combine several tasks and actions?*	The system can combine different tasks (*e.g.* with content generation or autonomous systems).
	How autonomous are the system's actions?	Refers to the autonomy level of the system or degree to which it can act without human involvement.	*Can it displace human labour?*	Refers to the ability of a system to automate tasks that are or were being executed by humans.
	Which core application area does the system belong to?	Core areas include human language technologies, computer vision and robotics.		