OECD.AI
Policy Observatory

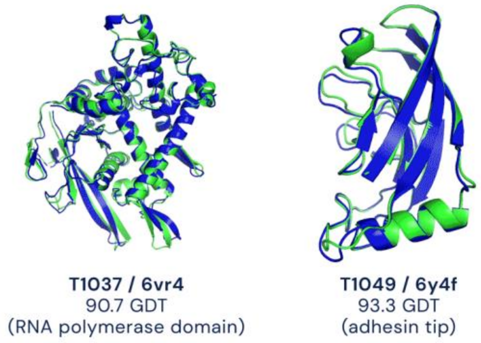
Enabling effective AI policies: Launch of the OECD Framework for Classifying AI Systems

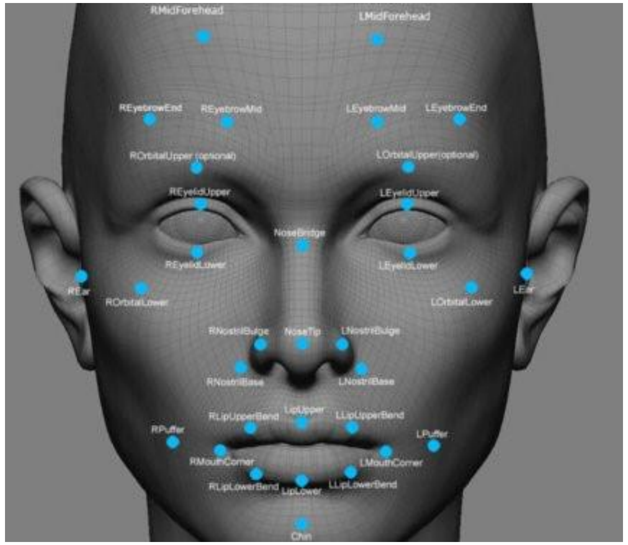
International Conference on AI in Work, Innovation, Productivity and Skills, *22 February 2022*

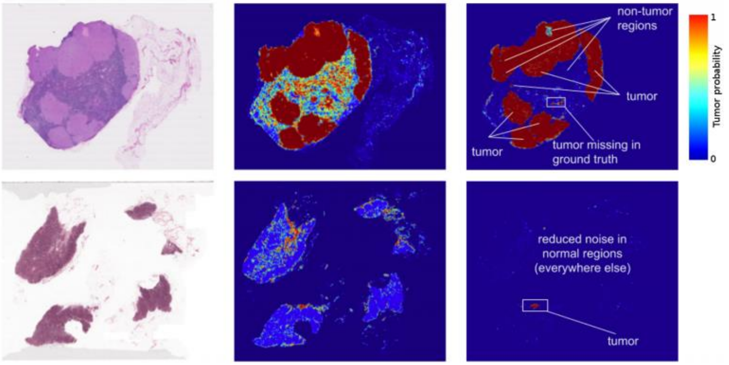
OECD

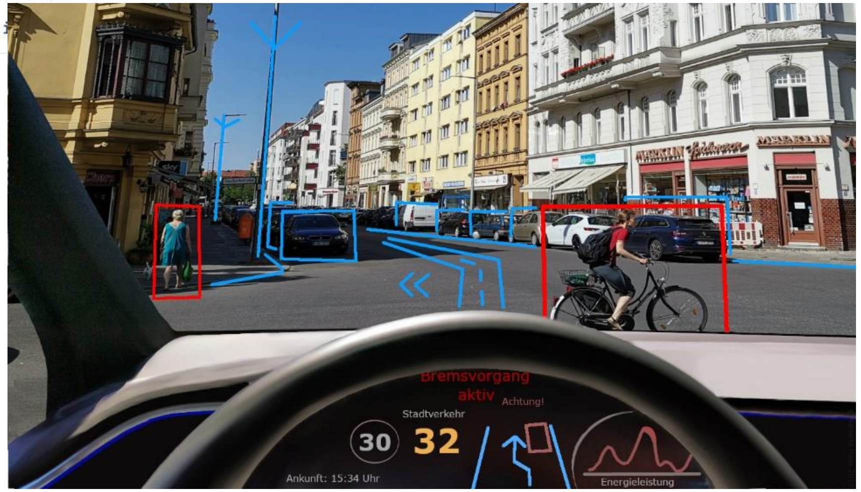
Why classify AI systems?

A variety of systems and policy implications






non-tumor regions
tumor
tumor
tumor missing in ground truth
Tumor probability
1
0
reduced noise in normal regions (everywhere else)
tumor



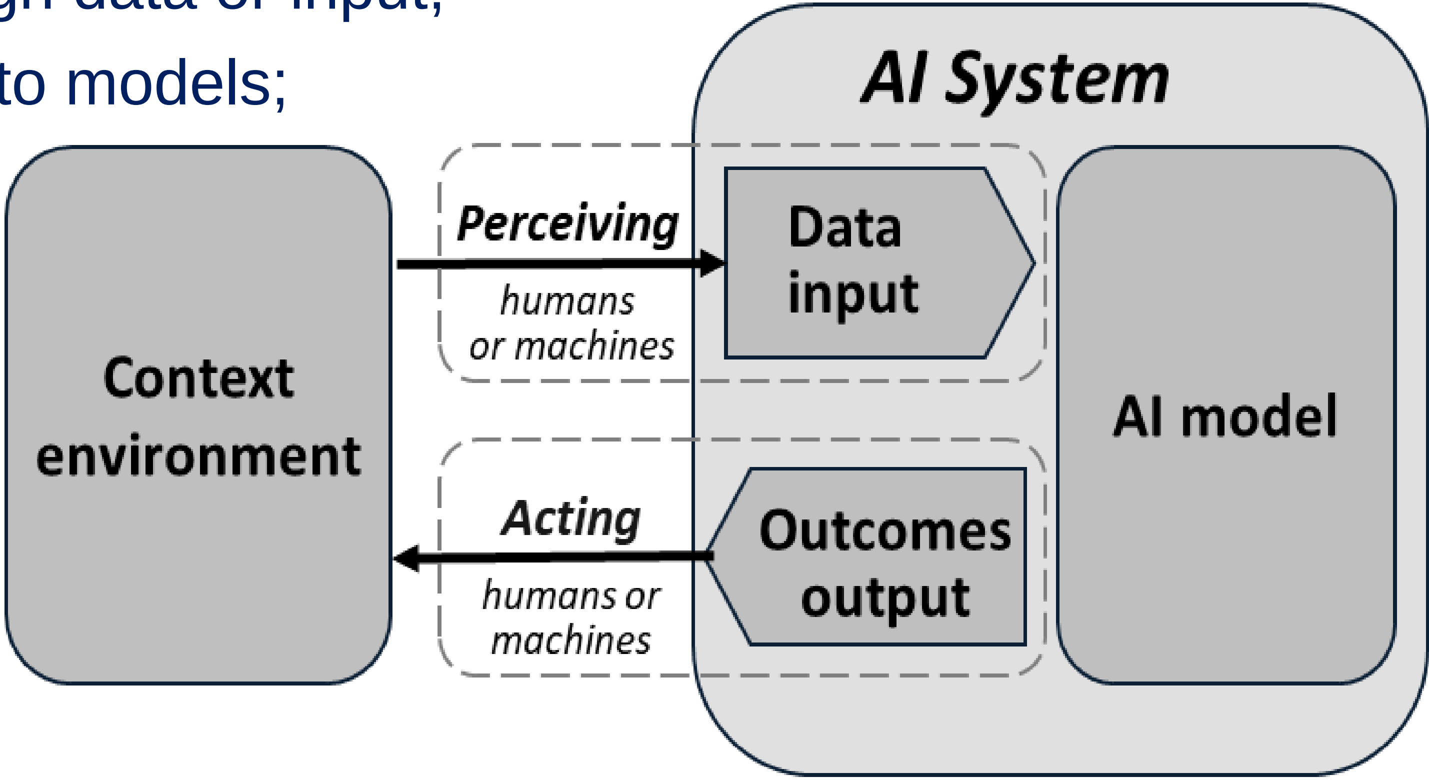
OECD AI System Definition (OECD, 2019)

“An AI system, is a machine-based system that is capable of influencing the environment by producing an output (recommendations, predictions or decisions) for a given set of objectives.

i) perceives environments through data or input;

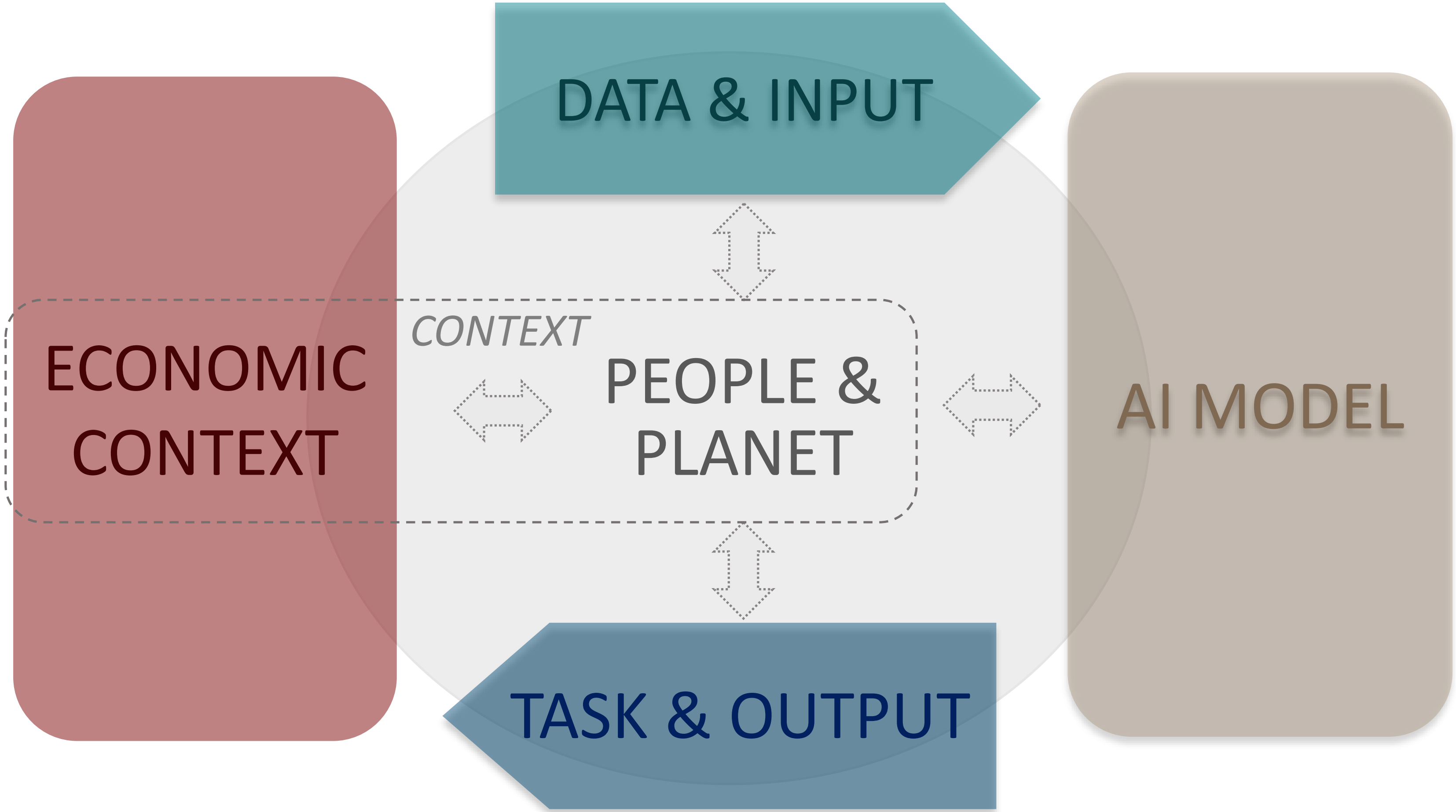
ii) abstracts these perceptions into models;

iii) uses the models to formulate options for outcomes.”



OECD Framework for Classifying AI systems: Key dimensions characterise AI systems' policy impact

OECD.AI Policy Observatory



Linking the classification & AI system lifecycle actors

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

Uses of the OECD AI Classification Framework

APPLICABILITY: Most relevant to classifying specific AI applications, rather than generic AI systems

GOAL: Provide a baseline framework to help support and advance :

1. a common understanding of AI, and metrics.
2. structure registries or inventories of AI systems.
3. sector-specific frameworks, e.g. in healthcare (NICE).
4. **risk assessment and incident reporting** (***next steps***).
5. risk management & work on accountability along the AI system lifecycle (***next steps***).

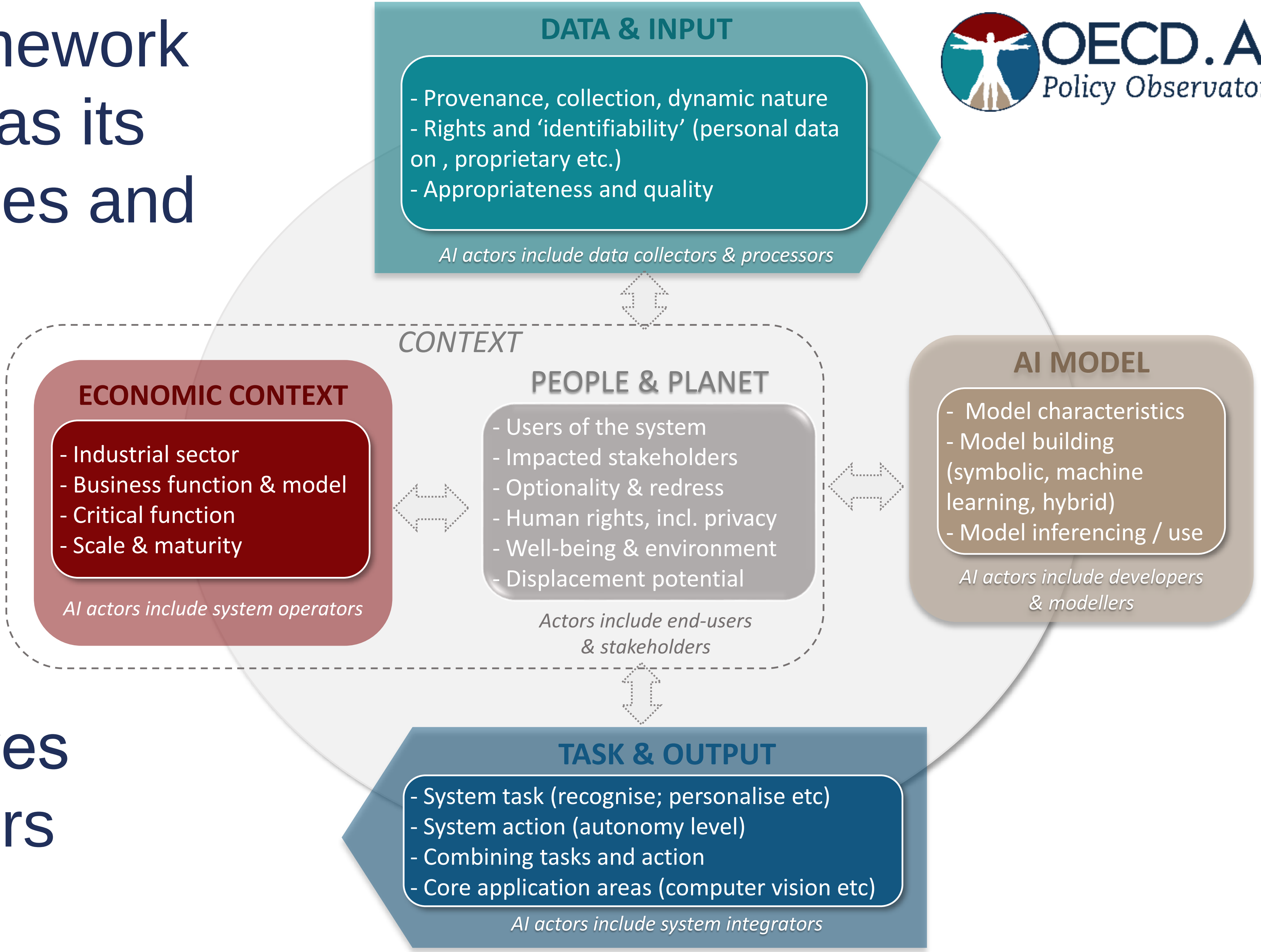
PROCESS:

- Consensus of group of 60+ experts
- Testing & public consultation May-June 2021:
  > 850 comments & survey responses => Adapted framework.

Sincere thanks for invaluable input to all who commented and tested the framework. 🙏

OECD.AI
Policy Observatory

Each AI framework dimension has its own properties and attributes…



DATA & INPUT

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

AI actors include data collectors & processors

CONTEXT

ECONOMIC CONTEXT

- Industrial sector
- Business function & model
- Critical function
- Scale & maturity

AI actors include system operators

PEOPLE & PLANET

- Users of the system
- Impacted stakeholders
- Optionality & redress
- Human rights, incl. privacy
- Well-being & environment
- Displacement potential

Actors include end-users & stakeholders

AI MODEL

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

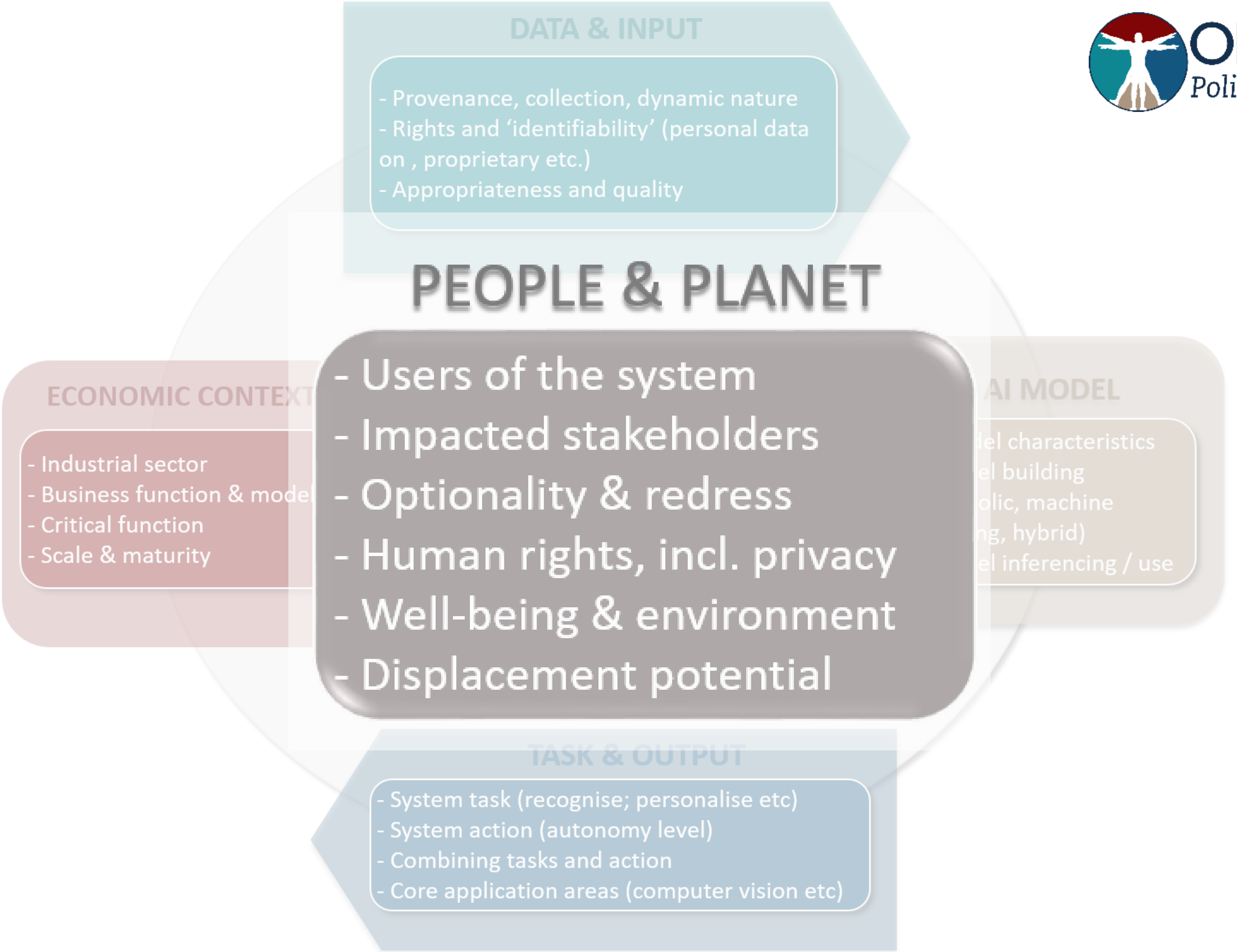
AI actors include developers & modellers

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

…and involves specific actors



DATA & INPUT

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

PEOPLE & PLANET

- Users of the system
- Impacted stakeholders
- Optionality & redress
- Human rights, incl. privacy
- Well-being & environment
- Displacement potential

ECONOMIC CONTEXT

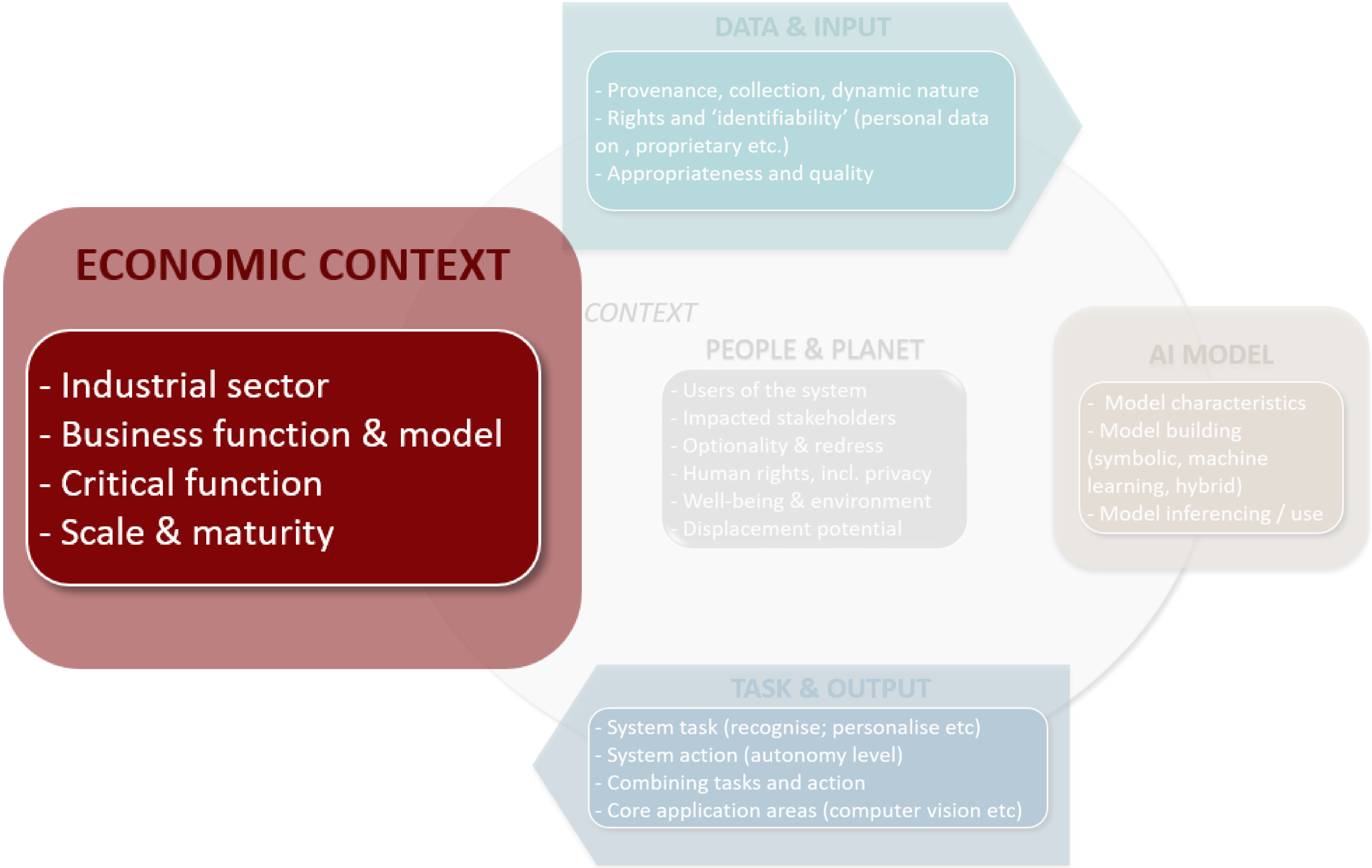
- Industrial sector
- Business function & model
- Critical function
- Scale & maturity

AI MODEL

del characteristics
el building
olic, machine
ng, hybrid)
el inferencing / use

TASK & OUTPUT

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



ECONOMIC CONTEXT

- Industrial sector
- Business function & model
- Critical function
- Scale & maturity

DATA & INPUT

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

CONTEXT

PEOPLE & PLANET

- Users of the system
- Impacted stakeholders
- Optionality & redress
- Human rights, incl. privacy
- Well-being & environment
- Displacement potential

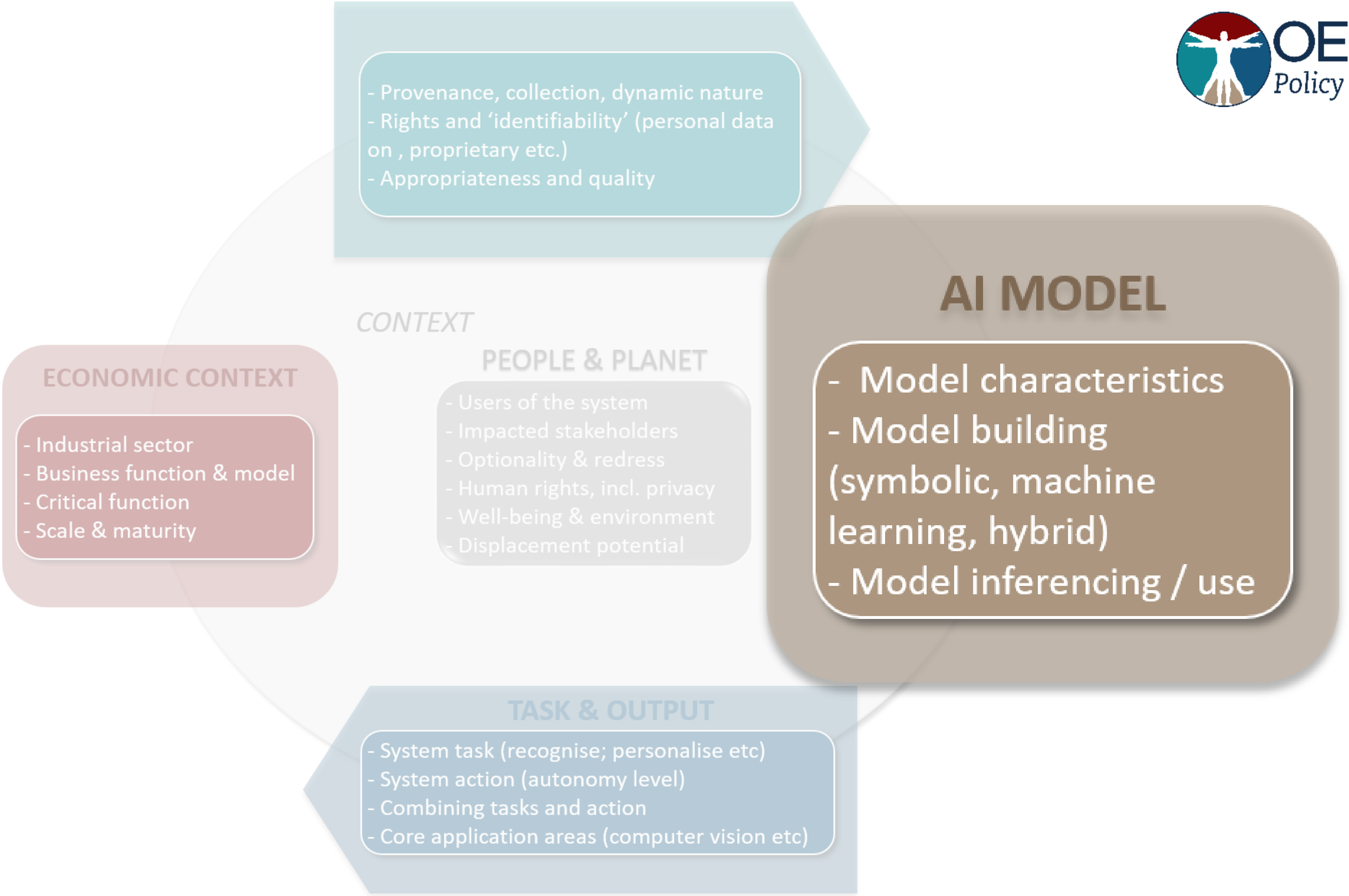
AI MODEL

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

TASK & OUTPUT

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

OECD.AI Policy Observatory



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

CONTEXT

ECONOMIC CONTEXT

- Industrial sector
- Business function & model
- Critical function
- Scale & maturity

PEOPLE & PLANET

- Users of the system
- Impacted stakeholders
- Optionality & redress
- Human rights, incl. privacy
- Well-being & environment
- Displacement potential

AI MODEL

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

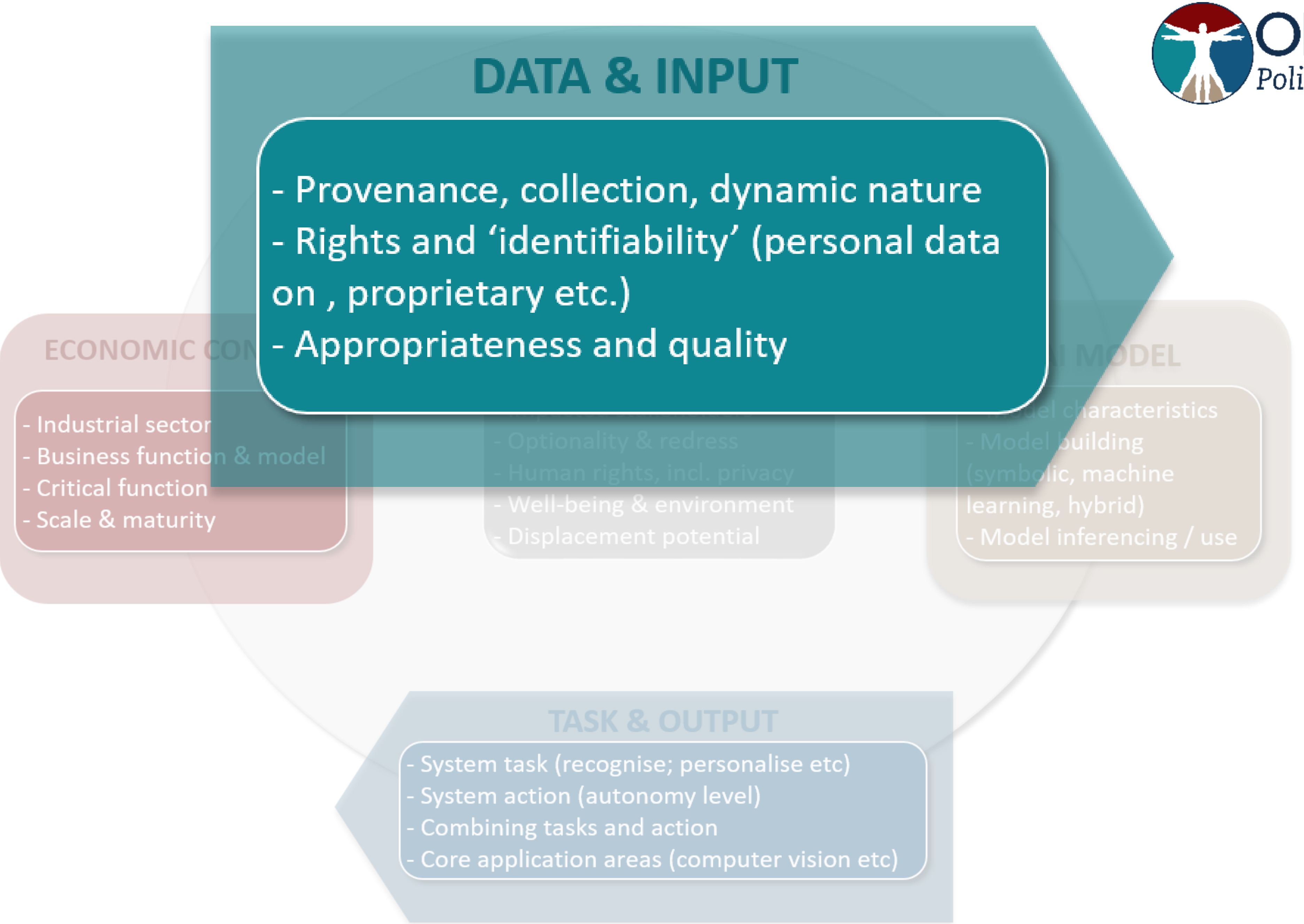
TASK & OUTPUT

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

OECD.AI Policy Observatory

DATA & INPUT

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



ECONOMIC CON

- Industrial sector
- Business function & model
- Critical function
- Scale & maturity

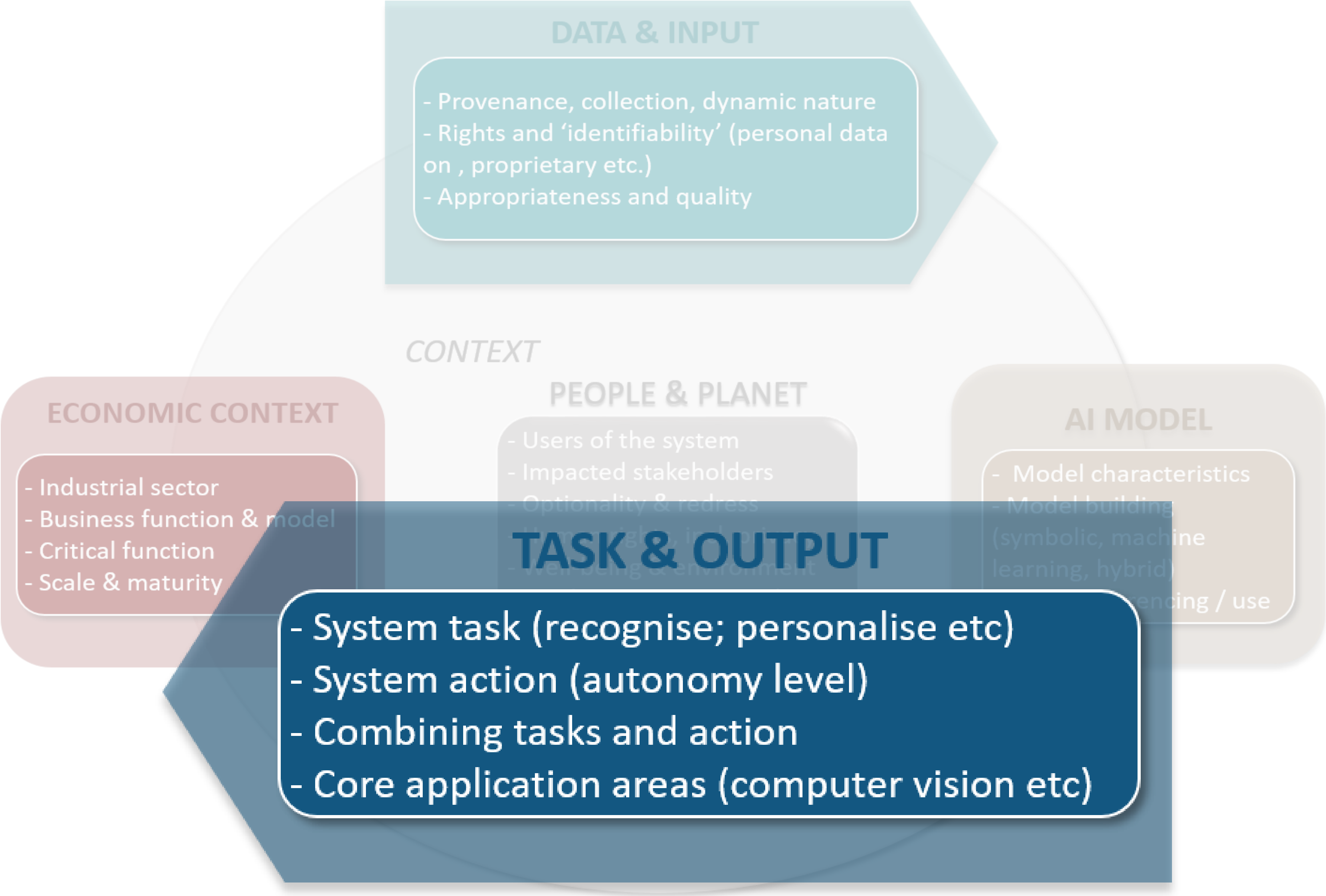
- Optionality & redress
- Human rights, incl. privacy
- Well-being & environment
- Displacement potential

I MODEL

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

TASK & OUTPUT

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



DATA & INPUT

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

CONTEXT

ECONOMIC CONTEXT

- Industrial sector
- Business function & model
- Critical function
- Scale & maturity

PEOPLE & PLANET

- Users of the system
- Impacted stakeholders
- Optionality & redress
- Human rights, inclusion
- Well-being & environment

AI MODEL

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

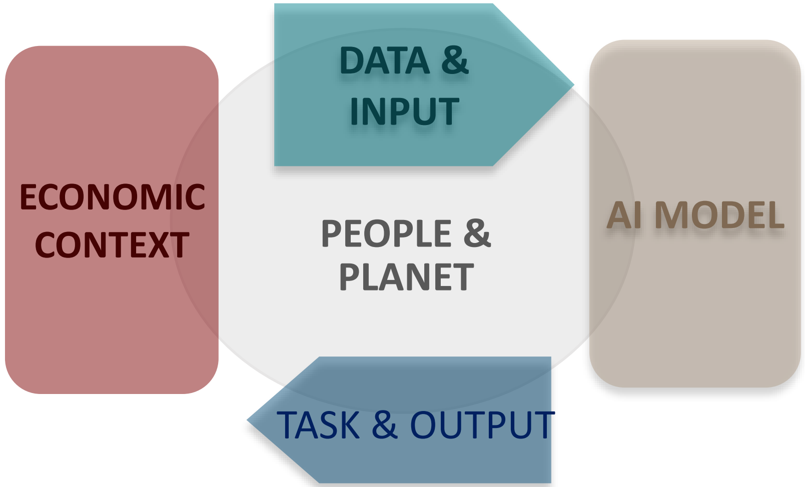
TASK & OUTPUT

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

Testing the framework with real AI systems

Key conclusions from survey responses :

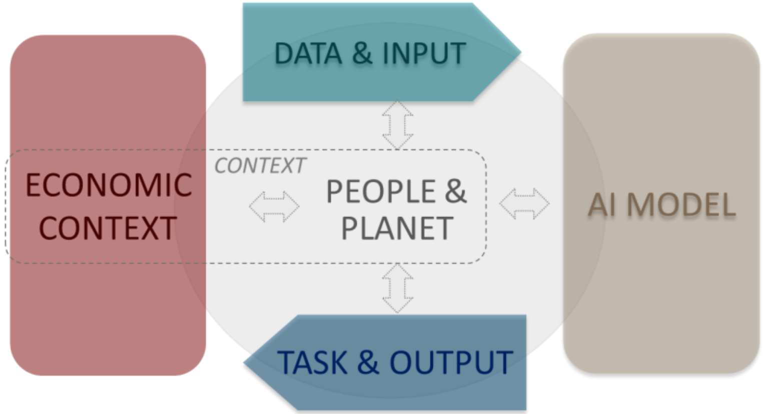
- The framework is best suited to **specific applications of AI systems** rather than to generic AI systems => the more specific the applications, the more consistent the survey responses.
- Respondents were better at classifying criteria in **People & Planet** and **Economic Context**. Classifying **Data & Input, AI Model, and Task & Output** often requires more technical information than is available publicly.



Example 1: Credit-scoring AI systems

OECD.AI Policy Observatory

Selected criteria:



- **System users –** Amateur (bank employee)
- **Optionality –** Cannot opt out
- **Human rights impact –** Yes
- **Sector of deployment** – **Financial system (e.g., banking, insurance)**
- **Critical function** – **Critical function/activity (availability of financial services, inclusion)**
- **Data collection –** Human (set of rules) and automated sources (e.g. profiles, loan payments)
- **Rights** – Mix of proprietary and public data
- **"Identifiability" –** often personally identifiable data
- **Model building** – e.g., statistical/hybrid model; learns from provided data, augmented by human knowledge
- **Model evolution** – Can evolve during operation
- **System task** – Forecasting: uses past & existing behavior to predict future outcomes
- **Level of action autonomy** – Medium (human on-the-loop)

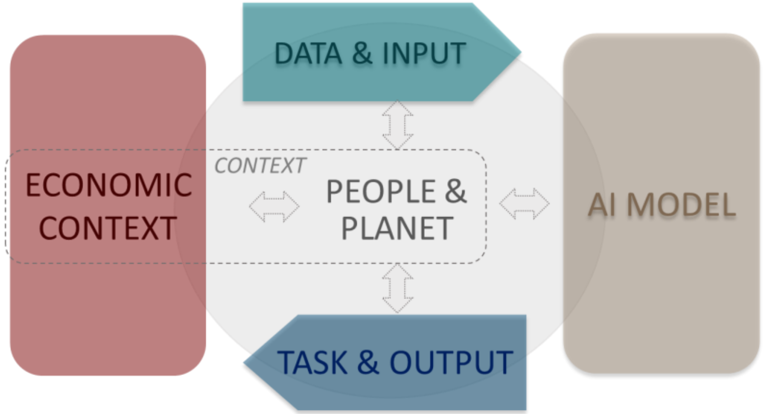
Example 2: GPT-3, text generation

OECD.AI Policy Observatory

Selected criteria:

Caveat: general purpose AI system, so nearly all responses depend on the specific application context! Medical advice, content filter, creative writing…

- **System users** – Primary users are amateur
- **Impacted stakeholders** – workers, consumers
- **Sector of deployment** – Information & communication
- **Critical function** – None
- **Data collection** – Human sources (text strings)
- **Rights** – Largely public data sources (some proprietary)
- **Model building** – Learn from provided data
- **Model evolution** – Evolution during operation
- **System task** – Goal-driven optimization, Reasoning with knowledge structures, interaction support, recognition, personalisation
- **Level of action autonomy** – Low autonomy [human action required e.g., to use generated text]

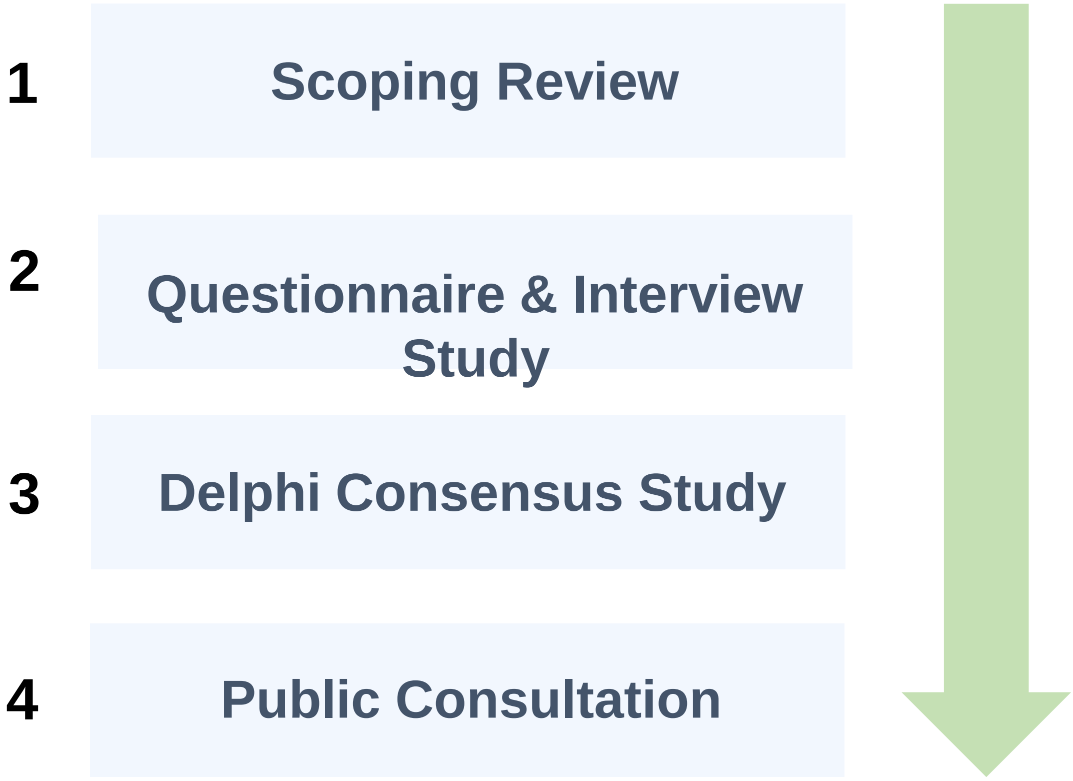


Imperial College London

The Alan Turing Institute

UNIVERSITY OF BIRMINGHAM

Using the framework to frame evidence standards for healthcare



1 Scoping Review

2 Questionnaire & Interview Study

3 Delphi Consensus Study

4 Public Consultation

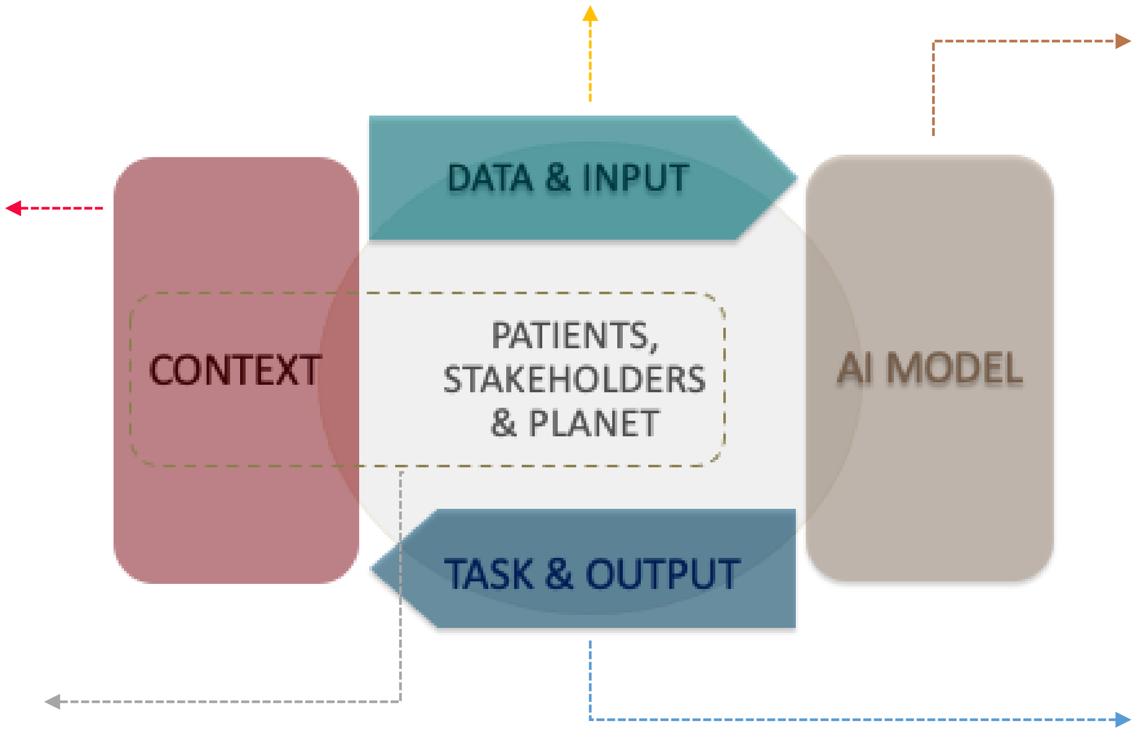
- OECD classification framework was independently ranked as most complete system from a shortlist of 21 candidates when mapped against the 9 core domains of HTA (EUnetHTA)
- Highest rated by a global multistakeholder panel of experts in both a questionnaire and interview study
- Framework now pivotal in developing evidence standards to underpin post-market evaluation in UK health sector

Using the framework for health technology assessment

The Alan Turing Institute

UNIVERSITY OF BIRMINGHAM

Input description



Clinical liability

Current use

Changes to data flows

Local deployment factors

Local performance

Model reporting

Training data reporting

Performance in different groups

Handling outlier data

Harm

Autonomy

End-user training

End-user well being

End-user/patient autonomy

Consent

Communication strategies

Output description

Post deployment change management plan

Benchmarking

Oversight committees

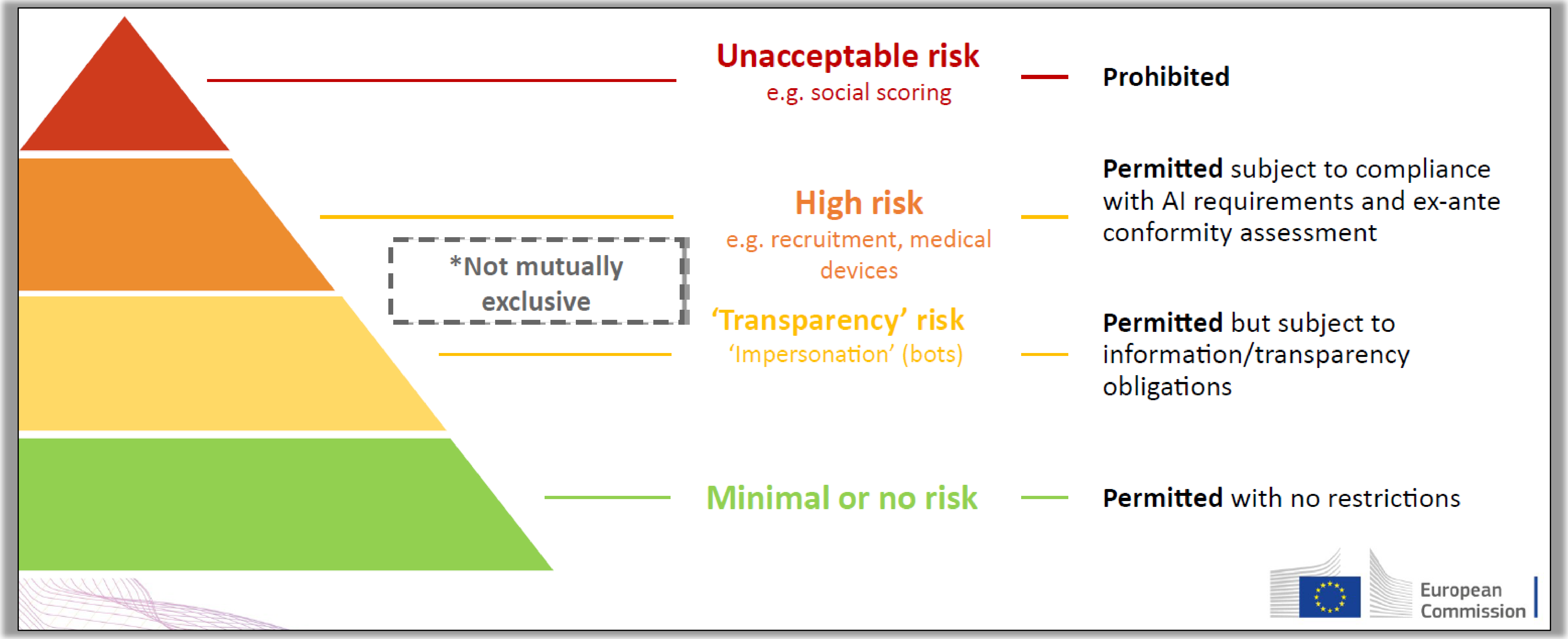
Next steps at the OECD:

OECD.AI Policy Observatory

- **Refine classification criteria**
  - Add more real-world AI systems and identify possible indicators
- **Develop a risk assessment framework** **to facilitate global interoperability**
  - leveraging the classification plus possible governance at the corporate, institution or AI systems level
  - Leveraging work in partner organisations, including EU, US, ISO
  - Leveraging risk assessment work in other parts of the OECD
  - Develop a common framework for reporting about AI incidents.
- **Support risk management**: Inform related work on mitigation, compliance and enforcement along the AI system lifecycle, and responsible business-impact assessment.

Reminder: Risk categorisation of **uses of AI** in the draft AI Act of the EU

OECD.AI Policy Observatory



Unacceptable risk
e.g. social scoring
— **Prohibited**

High risk
e.g. recruitment, medical devices
— **Permitted** subject to compliance with AI requirements and ex-ante conformity assessment

*Not mutually exclusive

'Transparency' risk
'Impersonation' (bots)
— **Permitted** but subject to information/transparency obligations

Minimal or no risk
— **Permitted** with no restrictions

European Commission

Details will be needed on European regulation/standardisation aspects …

- Rules:
  Mapping of uses of AI to these four categories (detailed rules, examples …)

- Governance:
  AI integration into / extension of frameworks for –
  - risk management
  - risk mitigation

- Tools:
  Assessment tools to operationalize mapping rules, due diligence, AI governance etc.

Formal EU standardisation requests expected H1/22

Harmonised standards created by ESOs, especially CEN-CENELEC JTC21

Mix of actors, including private sector, European Commission, NGOs, …

OECD.AI
Policy Observatory

For more information visit
www.oecd.ai/classification

email: ai@oecd.org

OECD