Draft European AI Regulation

Opportunities and challenges

First Workshop on
Challenges and Adequacy Conditions for Logics in the New Age of Artificial Intelligence
ACLAI'22

Víctor **Rodríguez Doncel**

November 4th, 2022

Residencia Lucas Olazábal, Cercedilla, Madrid

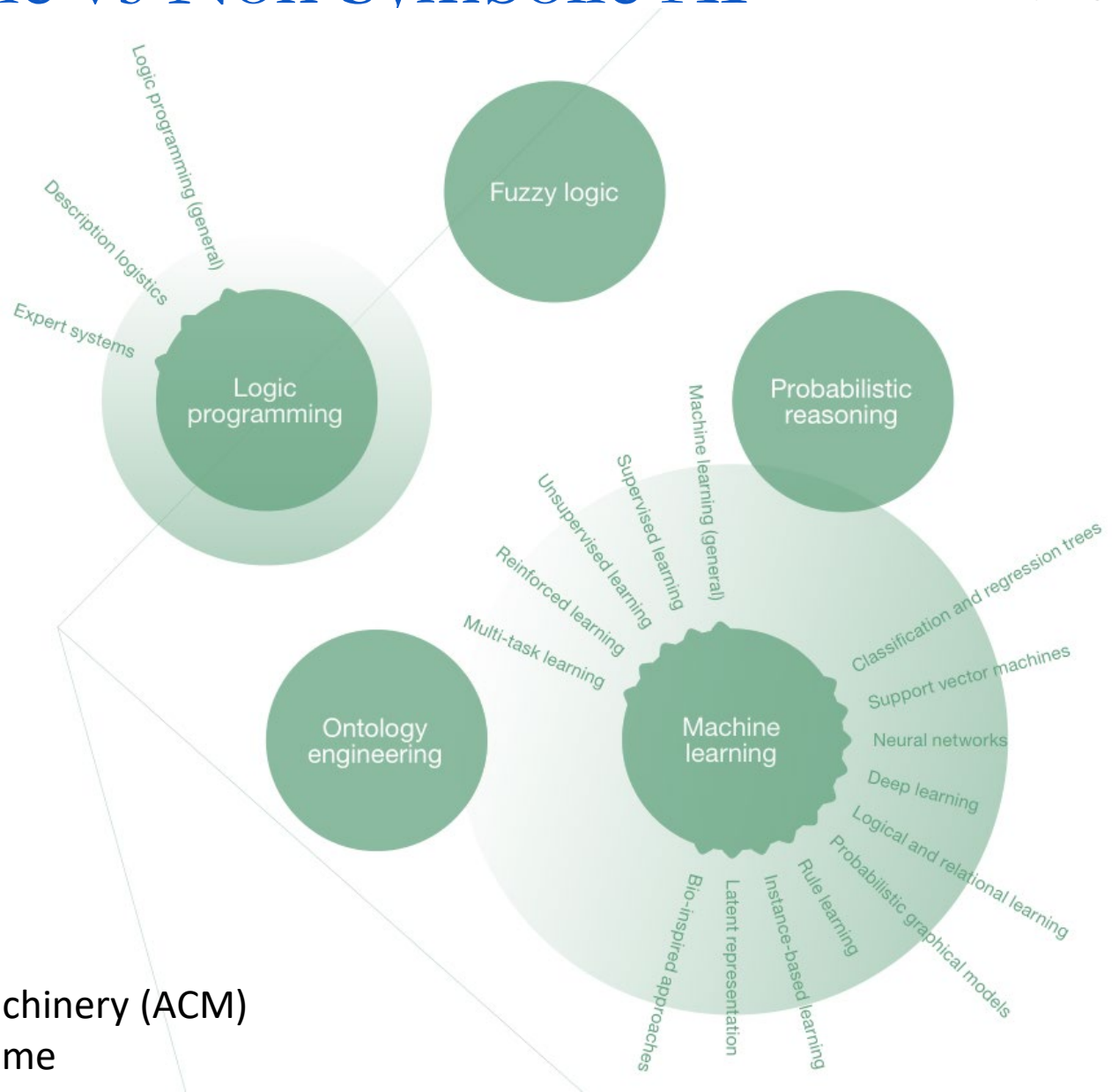
LIANDA - *Lógica e Inteligencia Artificial: Nuevos Desafíos y Adecuación*

Session objectives

- Describe the European regulation of AI
- Discuss challenges in relation to logic-based systems

How many of you knew that a Regulation on AI is being discussed?

Symbolic vs Non Symbolic AI

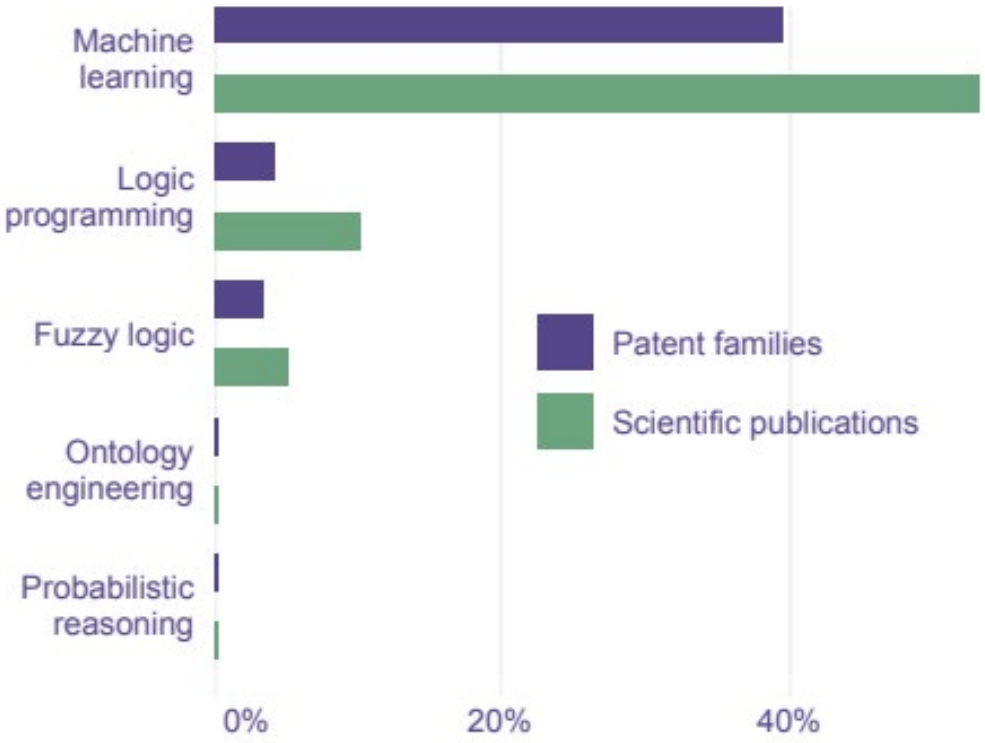


Association for Computing Machinery (ACM)
Computing **Classification** Scheme

Share of patents and academic publications

Figure 3.6. Patent families and scientific publications related to AI techniques as a share of the total for AI

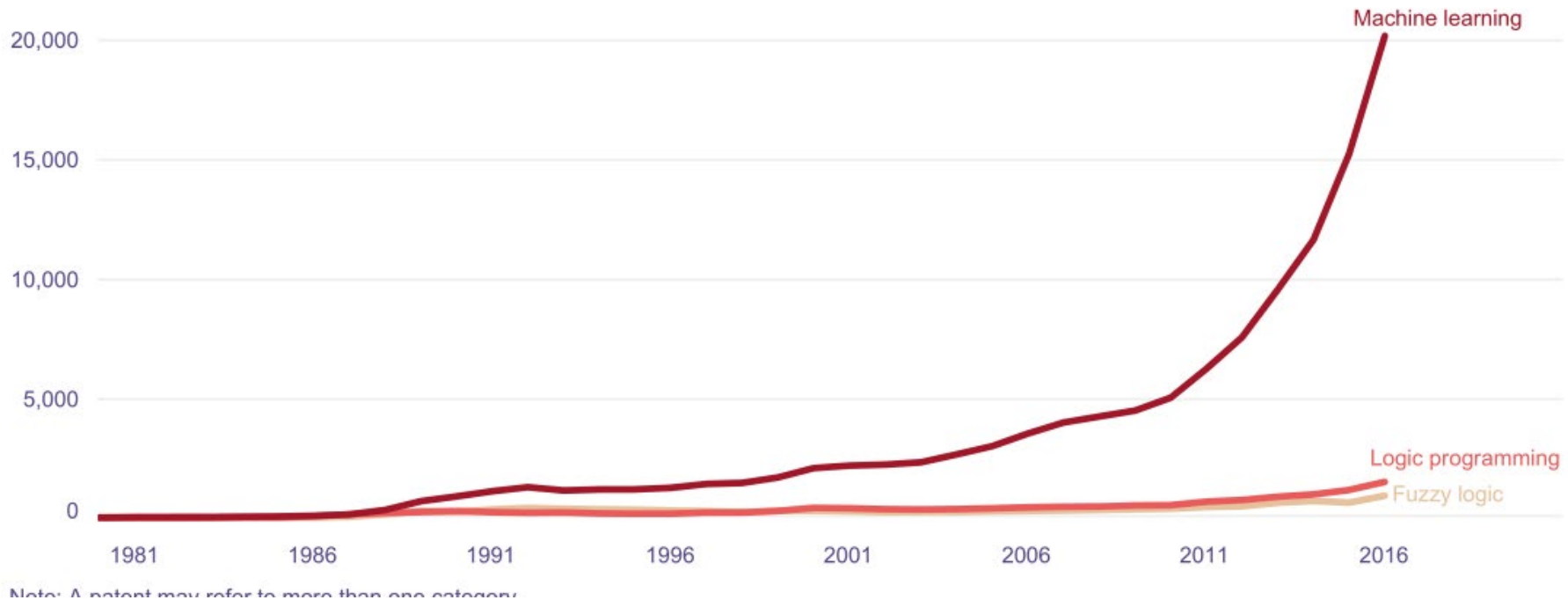
The share of scientific publications is generally higher than patent families for AI techniques



Trend

Figure 3.4. Patent families for top AI techniques by earliest priority year

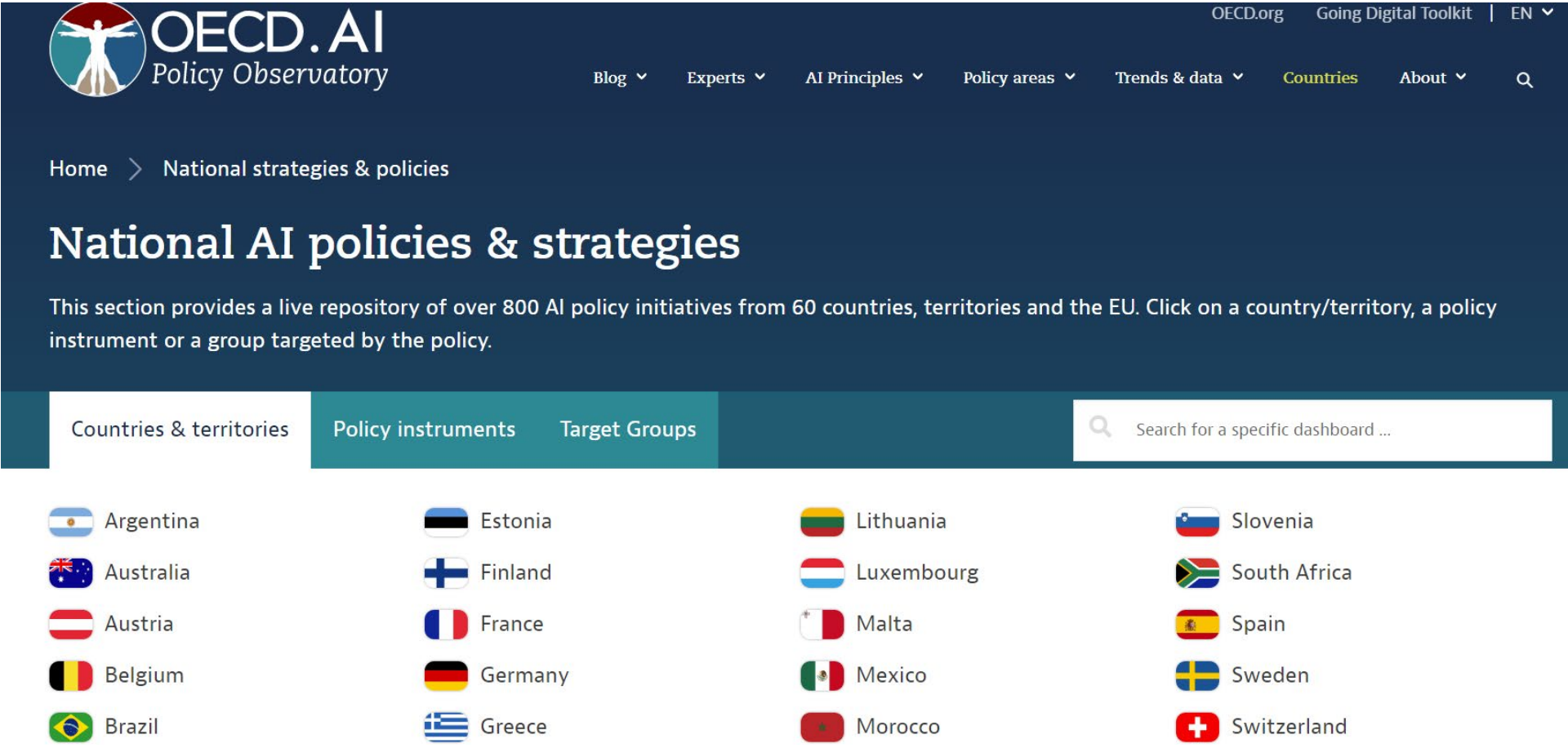
Machine learning grew by an average of 26 percent annually between 2011 and 2016



Source: WIPO technology trends 2019 – Artificial Intelligence

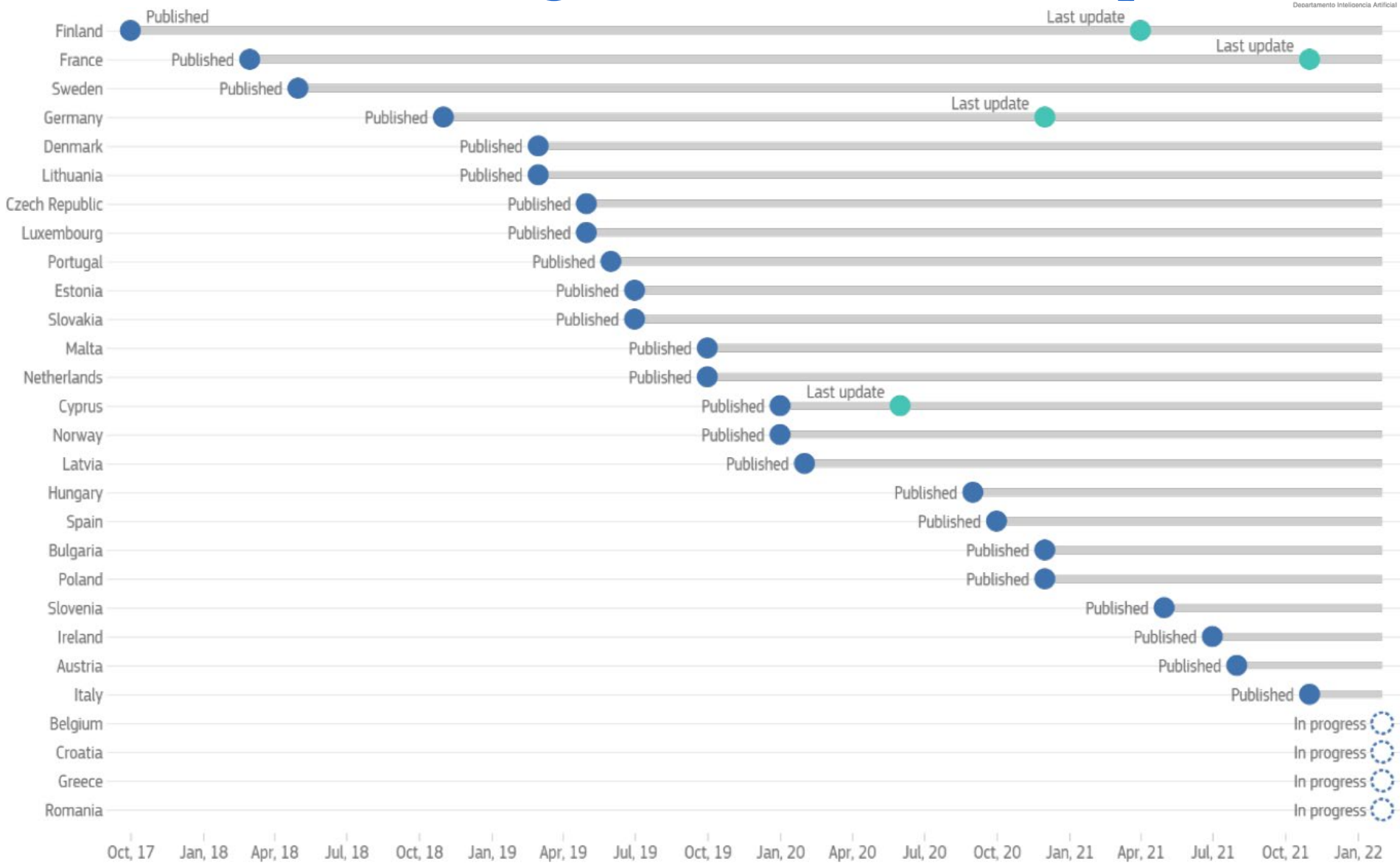
AI Strategies

National AI Strategies around the world



https://oecd.ai/en/dashboards/overview

National AI Strategies around the Europe



AI Watch. European landscape on the use of Artificial Intelligence by the Public Sector

National AI Strategy in Spain

2019

2020

https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf

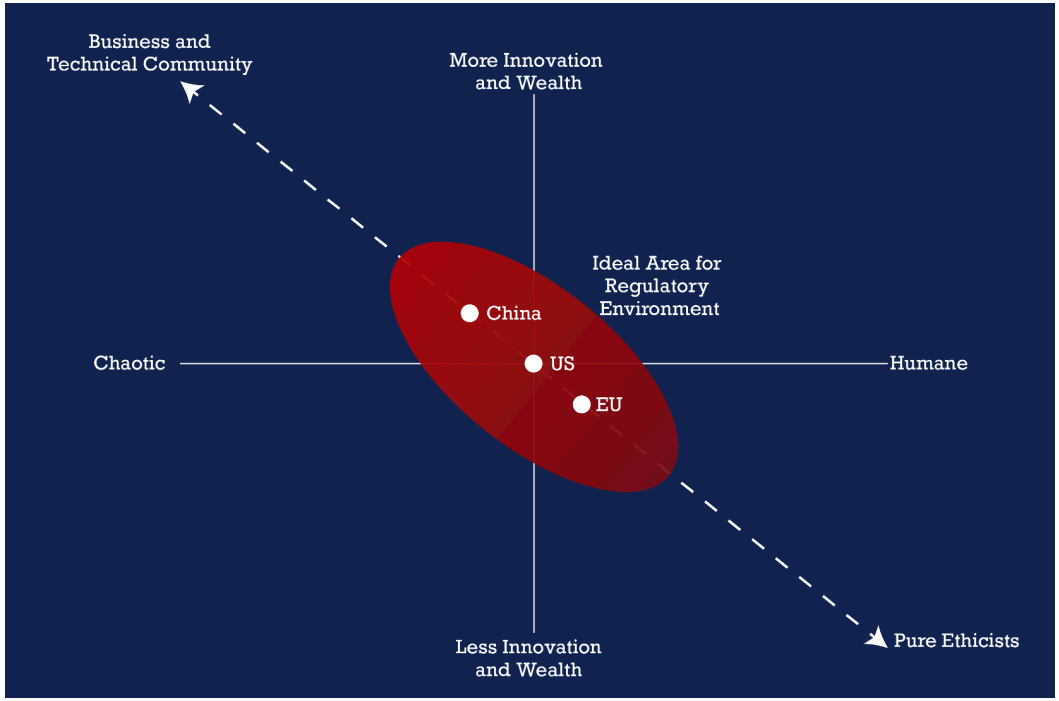
Spanish National Strategy for AI (ENIA)

- Presented in December 2020
  - Part of the Spanish Digital Agenda 2025 and of the Recovery Plan (*Plan de Recuperación, Transformación y Resiliencia de la economía española*)
- Objectives
  - Make Spain a reference for the transformation towards the Data Economy
  - Push Artificial Intelligence as an engine for innovation and economic, social, inclusive and sustainable growth
  - Prepare Spain for the socioeconomic transformation driven by AI
  - Strengthen competitiveness by R&D activities
- Main tools
  - «AI National Strategy»
  - Data Office and Chief Data Officer (CDO)
  - Advisory board for AI (Consejo Asesor de Inteligencia Artificial) - Our Professor Asunción Gómez in here, together with 18 other people!
  - European Data Spaces

European Strategy for AI

European Approach to AI (first movements)

- Expert groups
  - High-Level Expert Group on AI (since 2018)
  - HLEG Ethics guidelines for Trustworthy AI (April 2019)
  - HLEG on the Impact of Digital Transformation on EU Labour Markets
  - HLEG on Liability of New Technologies
- Whitepaper on AI – A European approach to excellence and trust (2020)



Self perception of Europe

O. Yalcin. Designing Explainable AI systems in a Quest for Sustainable AI: A Multidisciplinary Study on AI Explainability and Sustainable AI

European Approach to AI

Objectives:

- Boost the EU's technological and industrial capacity and AI uptake across the economy
- Prepare for socioeconomic changes
- Ensure an appropriate ethical and legal framework

Rules and actions:

- Fostering European Approach to Artificial Intelligence
- Coordinated Plan on AI (2018, 2021)
- Proposal for an AI Regulation

Trustworthy AI

Ethics guidelines for trustworthy AI

Framework for Trustworthy AI

INTRODUCTION

Trustworthy AI

- **Lawful AI** — (not dealt with in this document)
- **Ethical AI**
- **Robust AI**

CHAPTER I

Foundations of Trustworthy AI

Adhere to ethical principles based on fundamental rights

4 Ethical Principles

Acknowledge and address tensions between them

- Respect for human autonomy
- Prevention of harm
- Fairness
- Explicability

CHAPTER II

Realisation of Trustworthy AI

Implement the key requirements

7 Key Requirements

Evaluate and address these continuously throughout the AI system's life cycle via

- **Technical Methods**
- **Non-Technical Methods**

- Human agency and oversight
- Technical robustness and safety
- [illegible]nance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental wellbeing
- Accountability

CHAPTER III

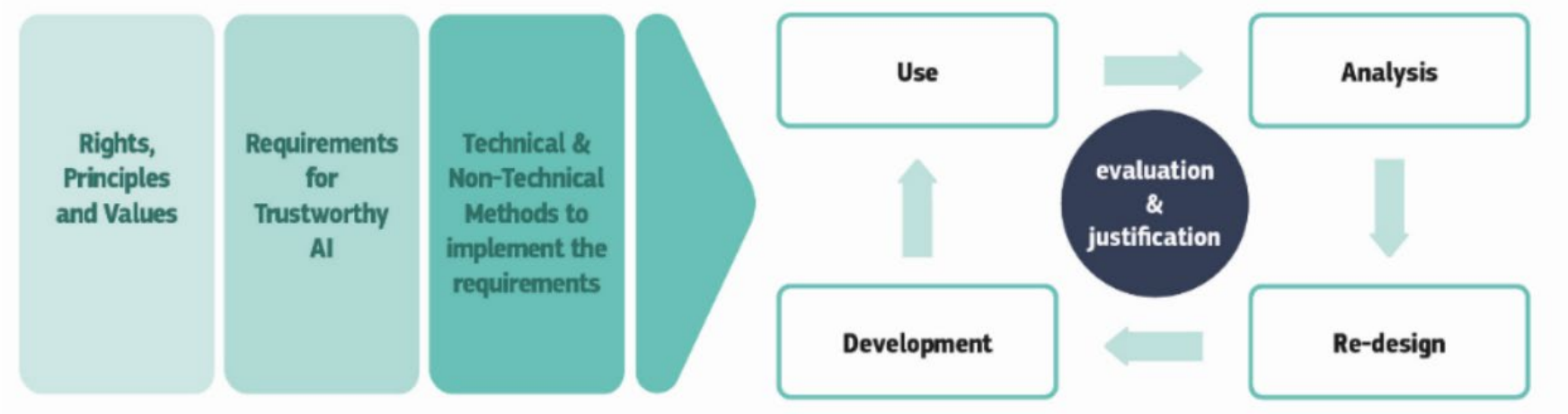
Assessment of Trustworthy AI

Operationalise the key requirements

Trustworthy AI Assessment List

Tailor this to the specific AI application

Realisation of trustworthy AI



Assessment list

https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60440

Some questions

Explainability

Q34 - Did you assess:

Q34.1 -- to what extent the decisions and hence the outcome made by the AI system can be understood?

Q34.2 -- to what degree the system's decision influences the organisation's decision-making processes?

Q34.3 -- why this particular system was deployed in this specific area?

Q34.4 -- what the system's business model is (for example, how does it create value for the organisation)?

Q35 - Did you ensure an explanation as to why the system took a certain choice resulting in a certain outcome that all users can understand?

Q36 - Did you design the AI system with interpretability in mind from the start?

Q36.1 -- Did you research and try to use the simplest and most interpretable model possible for the application in question?

Q36.2 -- Did you assess whether you can analyse your training and testing data? Can you change and update this over time?

Some questions

Traceability

Q30 - Did you establish measures that can ensure traceability? This could entail documenting the following methods:

Q31 - Methods used for designing and developing the algorithmic system:

Q31.1 -- Rule-based AI systems: the method of programming or how the model was built;

Q31.2 -- Learning-based AI systems; the method of training the algorithm, including which input data was gathered and selected, and how this occurred.

Q32- Methods used to test and validate the algorithmic system:

Q32.1 -- Rule-based AI systems; the scenarios or cases used in order to test and validate;

Q32.2 -- Learning-based model: information about the data used to test and validate.

Q33 - Outcomes of the algorithmic system:

Q33.1 -- The outcomes of or decisions taken by the algorithm, as well as potential other decisions that would result

Auditability

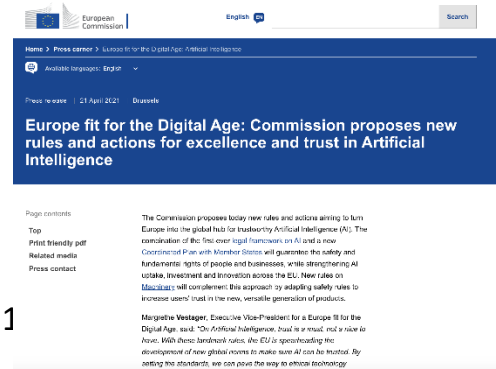
Q54 - Did you establish mechanisms that facilitate the system's auditability, such as ensuring traceability and logging of the AI system's processes and outcomes?

Q55 - Did you ensure, in applications affecting fundamental rights (including safety-critical applications) that the AI system can be audited independently?

Draft Regulation on AI

Draft AI Regulation

- April 2021: The European Commission released the proposal for a regulation *laying down harmonized rules on artificial intelligence and amending certain union legislative acts on*
- Regulation, not directive (same rules across EU)
- The proposal is the result of several years of preparatory work by the commission and its advisers, including the publication of a **"White Paper on Artificial Intelligence"**. Coordinated with Member States.
- The proposal is a key piece in the Commission's ambitious **European Strategy for Data.**



Press release: https://ec.europa.eu/commission/presscorner/detail/en/IP_21_1682 21 April 2021

AI Regulation Objectives

- Ensure that AI systems placed on the Union market and used are **safe** and respect existing law on **fundamental rights and Union values**
- Ensure **legal certainty** to facilitate investment and innovation in AI
- Enhance **governance** and effective **enforcement** of existing law on fundamental rights and safety requirements applicable to AI systems
- Facilitate the development of a **single market for lawful, safe and trustworthy AI applications** and prevent market fragmentation

Who is affected by the AI Regulation?

The regulation applies to:

(1) **providers** that place on the market or put into service AI systems irrespective of whether those providers are established in the European Union or in a third country;

(2) **users of AI systems in the EU**; and

(3) **providers and users of AI systems** that are located in a third country where the output produced by the system **is used in the EU**.

Is the regulation applicable to you?

(Read between lines: also aimed at Silicon Valley giants)

What is an AI system?

The term “AI system” is broadly defined as

- software that is developed with one or more of the **techniques and approaches** listed in Annex I; and
- can, for a given set of human-defined objectives, **generate outputs** such as
  - content,
  - predictions,
  - recommendations, or
  - decisions

  influencing environments they interact with.

☞ Why do you claim the system in your group work is an AI system, according to the output?

This is not for me…?

I am not doing Machine learning…

ANNEX I
ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES
referred to in Article 3, point 1

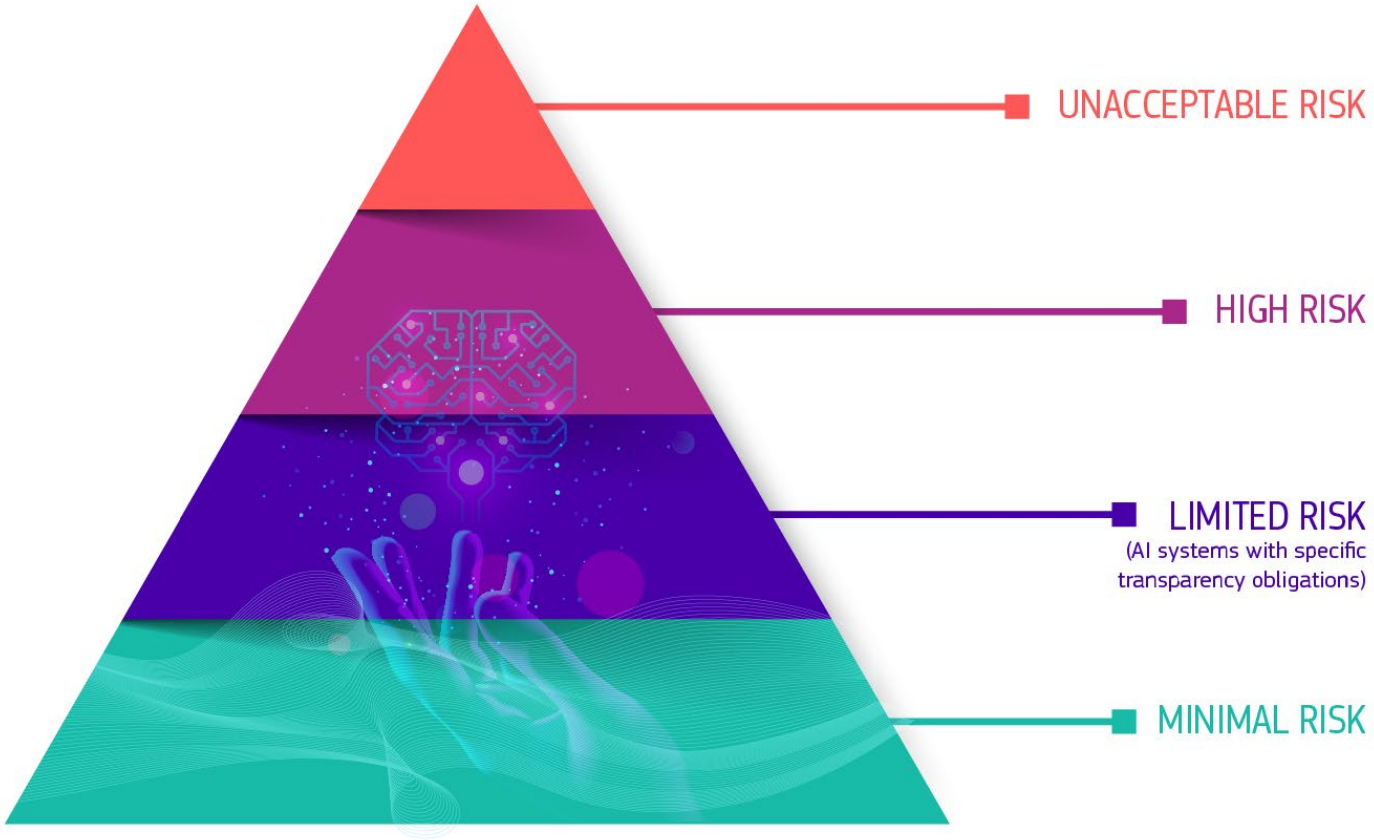
(a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning;

(b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems;

(c) Statistical approaches, Bayesian estimation, search and optimization methods.

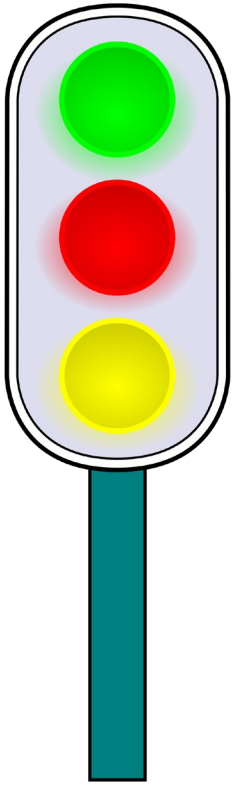
You knew you were doing AI. Now you know you are doing AI Systems to the eyes of the Law

Definition of AI risks

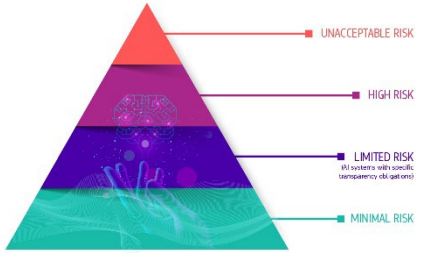


2021 Review: Fostering a European approach to Artificial Intelligence

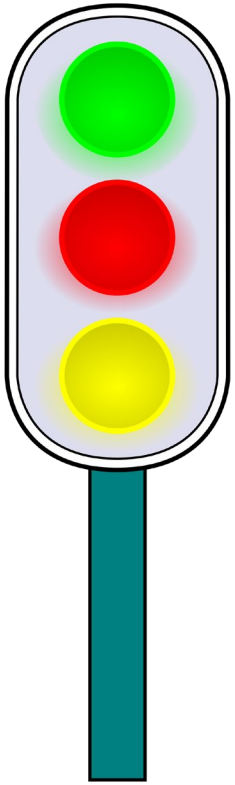
Risk assessment and risk levels (I)



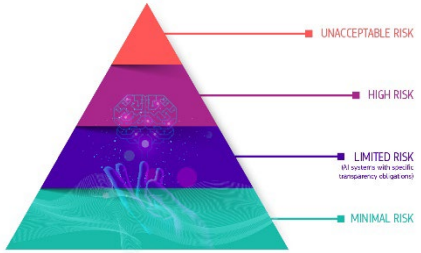
- Prohibited AI practices (**unacceptable risk**)
  - Violating fundamental rights
  - Manipulation of persons through subliminal techniques
  - Exploitation of vulnerable groups (children, persons with disabilities)
  - Psychological or physical harm
  - Social scoring
  - Real-time remote biometric identification systems in publicly accesible spaces for law enforcement (with limited exceptions)



Risk assessment and risk levels (II)



- High risk (based on the purpose of the AI system)
  - Full list in the following slides
- Legal requirements for them
  - Data governance
  - Documentation and record keeping
  - Transparency and provision of information to users. Users should be able to interpret the system output and use it appropriately
  - Human oversight
  - Robustness, accuracy and security



New rules for providers of high-risk AI systems

New rules for providers of high-risk AI systems

Step 1	Step 2	Step 3	Step 4	
				
A high-risk AI system is developed	It needs to undergo the conformity assessment and comply with AI requirements For some systems a notified body is involved	Registration of stand-alone AI systems in an EU database	A declaration of conformity needs to be signed and the AI system should bear the CE marking. The system can be placed on the market	*If substantial changes happen in the AI system's lifecycle, go back to Step 2*

More examples of high risk AI systems

- Biometric identification and categorisation of natural persons
  - ‘real-time’ and ‘post’ remote biometric identification of natural persons;
- Management and operation of critical infrastructure
  - safety components in the management and operation of road traffic and the supply of water, gas, heating and electricity.
- Education and vocational training:
  - determining access or assigning natural persons to educational and vocational training institutions;
  - assessing students and assessing participants in tests commonly required for admission to educational institutions
- Employment, workers management and access to self-employment:
  - recruitment or selection of natural persons, notably for advertising vacancies, screening or filtering applications, evaluating candidates in the course of interviews or tests;
  - making decisions on promotion and termination of work-related contractual relationships, for task allocation and for monitoring and evaluating performance and behavior of persons in such relationships.

More examples of high risk AI systems

- Access to and **enjoyment of essential private services and public services** and benefits
  - Used by public authorities or on behalf of public authorities to evaluate the eligibility of natural persons for public assistance benefits and services, as well as to grant, reduce, revoke, or reclaim such benefits and services;
  - Evaluate the creditworthiness of natural persons or establish their credit score, with the exception of AI systems put into service by small scale providers for their own use;
  - Dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid.

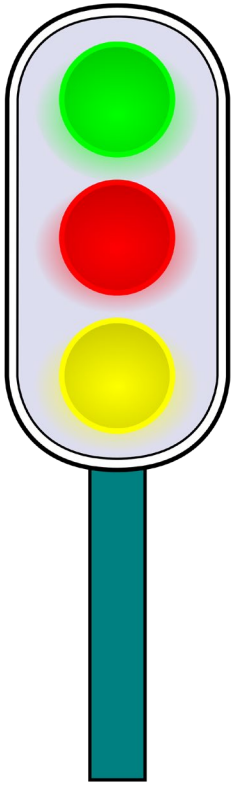
- **Law enforcement** (used by law enforcement authorities)
  - Making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences;
  - Polygraphs and similar tools or to detect the emotional state of a natural person;
  - Detect deep fakes
  - evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences;
  - predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups;
  - profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences;
  - crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data.

More examples of high-risk AI systems

- **Migration, asylum and border control management**
  - Polygraphs and similar tools or to detect the emotional state of a natural person;
  - assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State;
  - verification of the authenticity of travel documents and supporting documentation of natural persons and detect non-authentic documents by checking their security features;
  - examination of applications for asylum, visa and residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status.

- **Administration of justice and democratic processes:**
  - assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.

Risk assessment and risk levels (III)



- Low or minimal risk
  - Human interaction
  - Emotion detection
  - Biometric data
  - Deep fakes
- Legal requirements
  - Informed choices: Users need to be informed that they are interacting with a system, or that content is generated by a system.
  - Exceptions: law enforcement, freedom of expressions

Liability of AI products

Liable machines

liability = legal responsibility

Unavoidable!

Who pays for the broken monolyth?

New liability rules on products and AI to protect consumers and foster innovation

- Liability rules for circular economy business models (companies that modify products)
- Liability rules for products in the digital age (compensations by software updates, AI systems...)
- Non-EU manufacturers: the importer compensates
- Manufacturers must disclose evidence. Burden of proof of AI in the manufacturer side.
  - The new rules will, for instance, make it easier to obtain compensation if someone has been discriminated in a recruitment process involving AI technology

Published on September 2022

New liability rules on products and AI to protect consumers and foster innovation

- More on the Burden of Proof
  - Presumption of causality
  - Right of access to evidence
    - in cases in which high-risk AI is involved

Right of access to evidence

Published on September 2022

Opportunities and challenges for logicians

Opportunities

- You are doing AI Systems. Legislation affects you.
  - Explainable AI is necessary to **Trustworthy AI** (explicability, transparency, accountability)
  - If you do a High Risk AI System
    - Transparency and provision of information to users.
    - **Users** should be able to interpret the system output and use it appropriately
  - Consumers of AI: Right of access to evidence

Challenges

- Different users require different forms of explanation in different contexts

(maybe adapt to the context and cognitive level of the user?)

- System design often needs to balance competing demands (accuracy, privacy, interpretability)

("my machine detects cancer with 90% but cannot say why")

- Data quality and provenance is part of the explainability pipeline

(*truths have to be qualified*)

- Explainability alone cannot answer questions about accountability (users can contest, etc.)

(interactive, *non monotonic systems*)

https://ec.europa.eu/futurium/en/system/files/ged/ai-and-interpretability-policy-briefing_creative_commons.pdf