LINKSFOUNDATION.COM

FONDAZIONE links
PASSION FOR INNOVATION

Applied Data Science Project

Politecnico di Torino

L10 – AI Ethics

ellis
European Laboratory for Learning and Intelligent Systems

The European Artificial Intelligence Act (AIA)

EUROPEAN
COMMISSION

Brussels, 21.4.2021
COM(2021) 206 final

2021/0106 (COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

{SEC(2021) 167 final} - {SWD(2021) 84 final} - {SWD(2021) 85 final}

Regulatory framework (proposal): https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European AI approach: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

A future-proof AI

according the EC

European Commission is working to promote a modern AI, able to respond and comply with **people-defined goals** (**Human Centred**) and is **aware of the impacts** on physical and digital environments.

Some intrinsic features of AI systems are at the core of this goal:

OPACITY

- Of systems not providing users or affected parties any insight as to **how they came to produce the results**.
  AI algorithms can be opaque even when they are reliable

(Vaassen, B. 2022)

COMPLEXITY

- resulting from many different interacting aspects, parts, or functions, which cannot be easily described, analyzed, or predicted because of its intricacy and internal interactions. Over the “known unknows” (expected potential drawbacks), complexity implies the “**unknown unknowns”, challenges whose existence we are not aware of.**

(Pawson, Wong, and Owen 2011)

DATA DEPENDENCE

- Bias can be manifested in **multimodal data** through **sensitive features** and their **causal influences**, or through under/over-representation of certain groups.

AUTONOMY

- Of systems operating independently of human guidance, for a fixed goal (or “utility function”) with respect to which the appropriateness of actions will be evaluated.

(Ntouts, E et al. 2020)

The Artificial Intelligence Act (AIA)

The **first ever** legal framework on AI

The proposal expresses the **political commitment** to apply a **coordinated approach** to AI and **aware of the human and ethical implications**.

Four specific objectives:

- ensure that AI systems placed on the Union market and used are **safe** and respect existing **laws on fundamental rights** and **Union values**
- ensure **legal certainty to facilitate investment** and innovation in AI
- enhance governance 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**.

https://artificialintelligenceact.eu/the-act/

AI Act

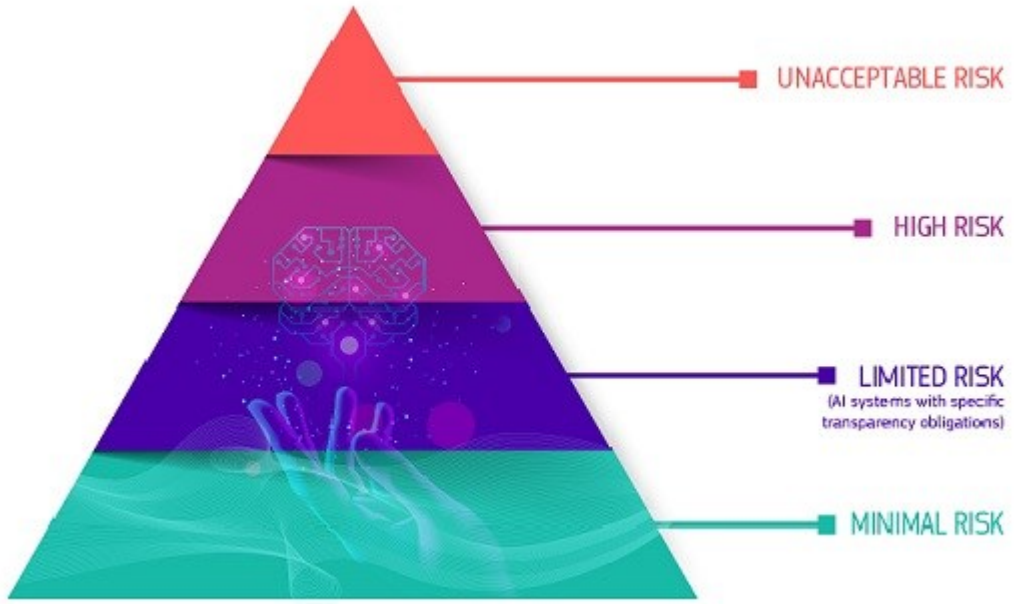
Key points:

- **Human** centric
- **Risk** based
- **EU Value** oriented

Considered risks:
Erroneous or distorted decisions suggested or supported by AI in critical areas, where there is a high risk to harm health, safety and human fundamental rights.

Risks: known, foreseeable, emerging, residual (associated with hazards)

Risks



Unacceptable applications

Systems violating human rights, persons assessment or classifications, subliminal techniques, real-time biometric identification systems)

High risk applications

- Systems for traffic and mobility control
- Water and electrical systems
- Education (evaluation of people)
- Work and access to work
- Access Systems to Public Services (credit score or creditworthiness)
- Law enforcement systems
- Migration and national borders
- Systems related to the administration of Justice

Real-time and post-hoc remote biometric identification systems

Limited risk applications

Administrative proceedings

Minimal risk applications

AI-enabled games or spam filters

Requirements and obligations

Mandatory requirements for high-risk AI systems:

- High quality data (adequate statistical properties)
- **Documentation** (also for the end users)
- **Transparency**
- **Human Oversight**
  - Interpretability
  - Minimize the risks of excessive trust (delegation)
- **Accuracy and robustness** (errors, defects, inconsistencies, unforeseen problems)

+ **cybersecurity** (addressing also the end users)

Obligations of providers

- Implement a **risk management system** that 1) identifies and analyses 2) estimates 3) manages known, foreseeable, emerging, residual risks
  - Guarantee an **effective communication** to the end-users
- Implementing a **quality management system**
  - **Data governance** including relevant, complete and free of errors training, validation and testing data sets (appropriate statistical properties)
  - **Bias monitoring**
- Provide **technical documentation**
- Automatically record the **logs**
- Carry out the **compliance assessment**
- Interface with **national authorities** and comply with legal obligations
- Should apply the **CE mark**

Forward looking

Two aspects remain central in this dicussion, but not thorougly answered yet

- Certification of AI products
- «Code of conduct» for the AI makers

FONDAZIONE links

PASSION FOR INNOVATION

Thank you for your attention.

Questions?

CONTACTS

Giuseppe Rizzo

Program Manager (LINKS Foundation) and
Adjunct Professor (Politecnico di Torino)

giuseppe.rizzo@polito.it

FONDAZIONE LINKS
Via Pier Carlo Boggio 61 | 10138 Torino
P. +39 011 22 76 150
LINKSFOUNDATION.COM