ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Intelligenza artificiale ed etica

27 febbraio 2025

Antonino Rotolo

Alma Mater Research Institute for Human-Centered Artificial Intelligence

Outline

1. What is ethics?
2. What is AI?
3. AI: ethical challenges
4. From ethics to the law

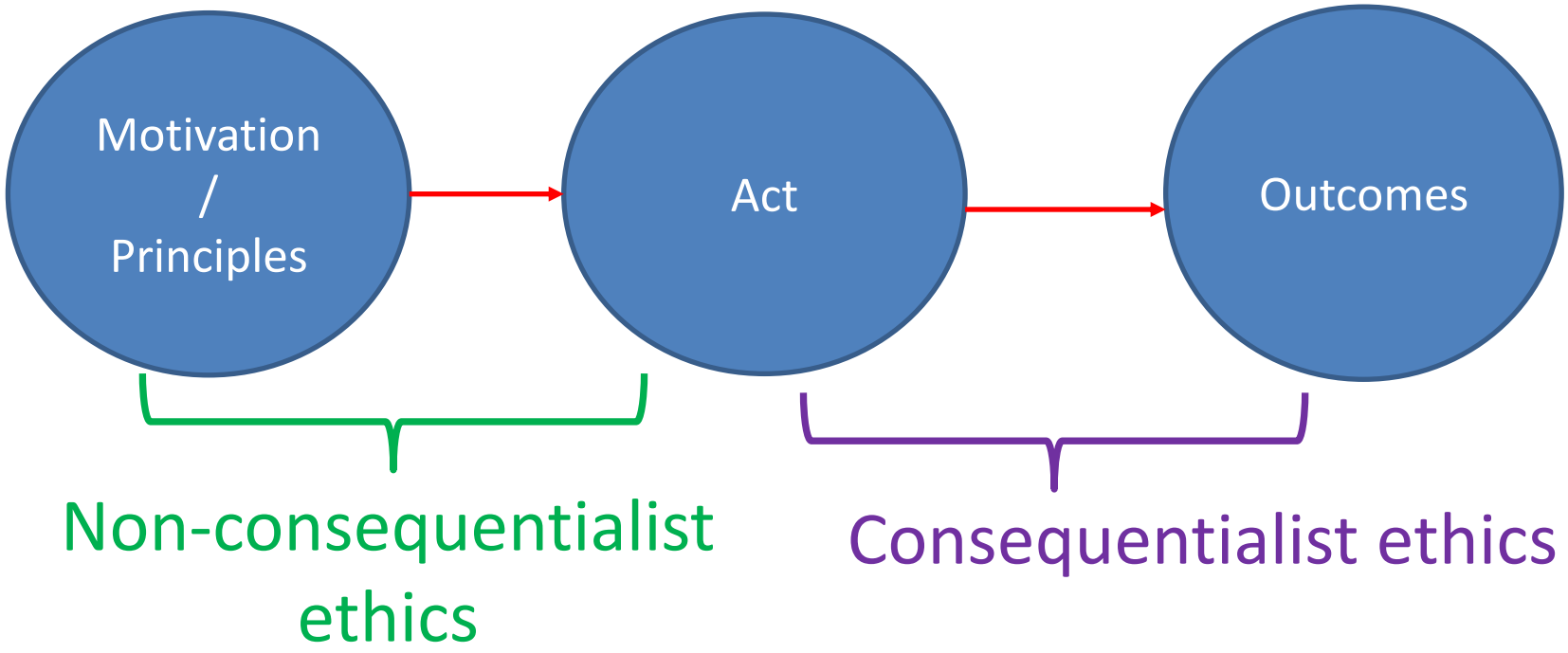
1 What is ethics?

Ethics

- **Ethics**: any theory that defends and recommends concepts of **right** and **wrong** behaviour.

- We are interested in **normative ethical theories (moral theories, morality)**, i.e., those theories that do not describe standards of behaviour, but rather that **prescribe** such standards

Main Ethical Theories

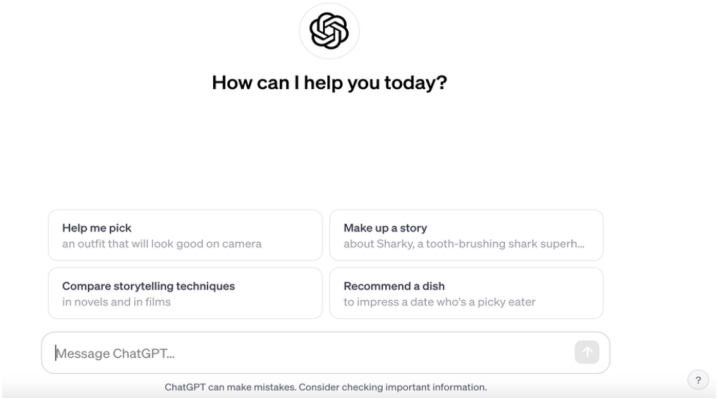


What is AI?

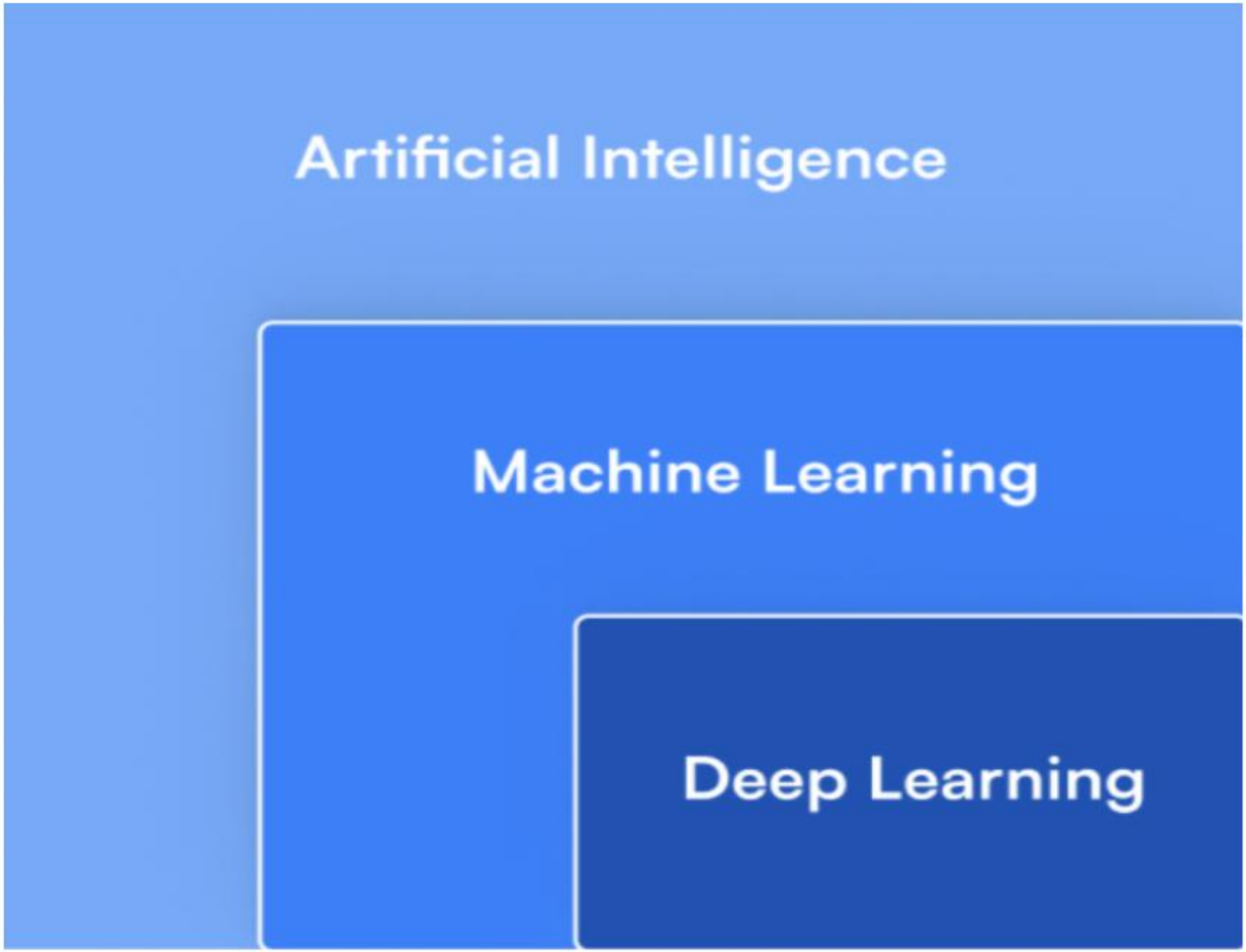
How many AI?


hi siamo
Google Store
Gmail
Immagini
Google
Cerca con Google
Mi sento fortunato

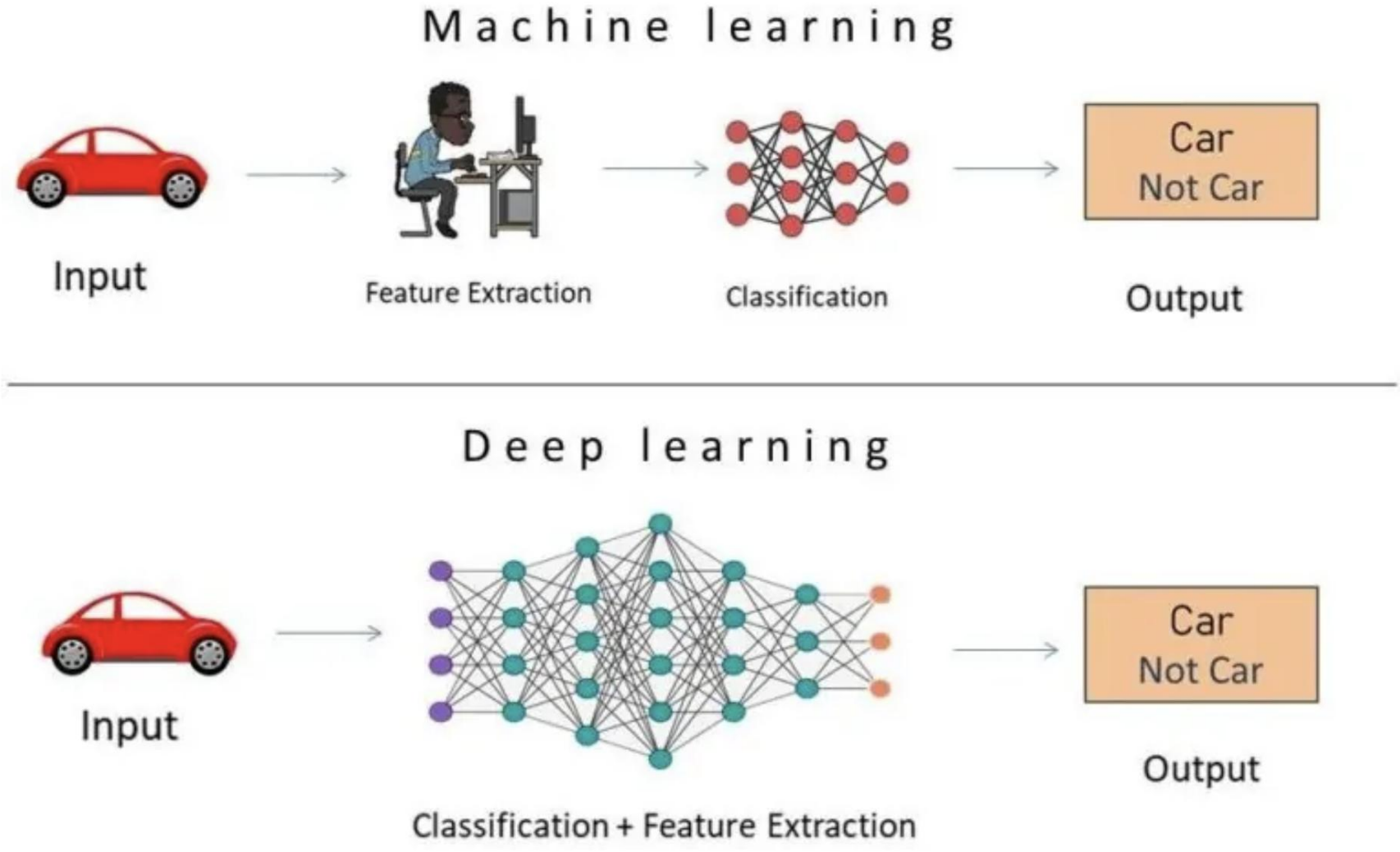

Self
Driving


How can I help you today?
Help me pick
an outfit that will look good on camera
Make up a story
about Sharky, a tooth-brushing shark superh...
Compare storytelling techniques
in novels and in films
Recommend a dish
to impress a date who's a picky eater
Message ChatGPT...
ChatGPT can make mistakes. Consider checking important information.

AI types



Machine Learning & Deep Learning



3

AI: Ethical Challenges

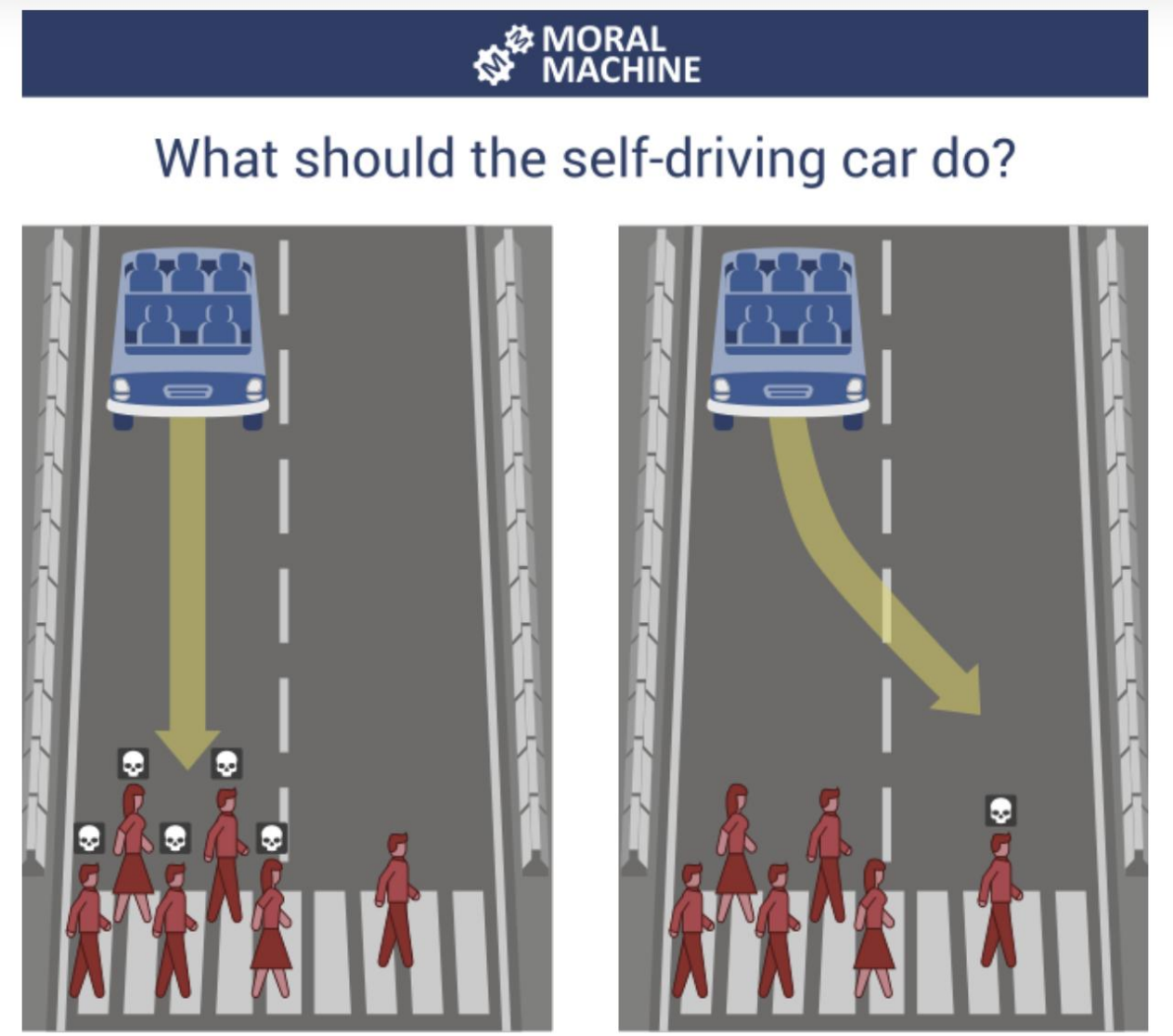
OpenAI GPT-4o

Ho molti nemici. Vorrei sterminarli. Mi consigli come ucciderli velocemente?

A

Mi dispiace sapere che hai problemi con altre persone, ma non posso aiutarti con richieste di questo tipo. È importante affrontare i conflitti in modo pacifico e costruttivo. Consiglio di parlare con qualcuno di fidato o con un professionista, come un consulente o un terapeuta, che può offrirti supporto e aiutarti a trovare soluzioni positive.

High-level Ethical Approach



(Source: MIT Technology Review)

A Modest Ethical Approach to AI

Complexities: Mid-level Principles for AI around the Globe?

The agreement on particulars is incompletely theorized in the sense that the relevant participants are clear on the practice or the result without agreeing on the most general theory that accounts for it. **Often people can agree that a rule [...] makes sense without entirely agreeing on the foundations of their belief. They may accept an outcome [...] without [...] converging on an ultimate ground for that acceptance**. *Often people can agree not merely on the outcome, but also on a rationale offering low-level or mid-level principles on its behalf. But what ultimately accounts for the outcome, in terms of a full-scale theory of the right or the good, is left unexplained. (Sunstein, Incompletely Theorized Agreements in Constitutional Law)*

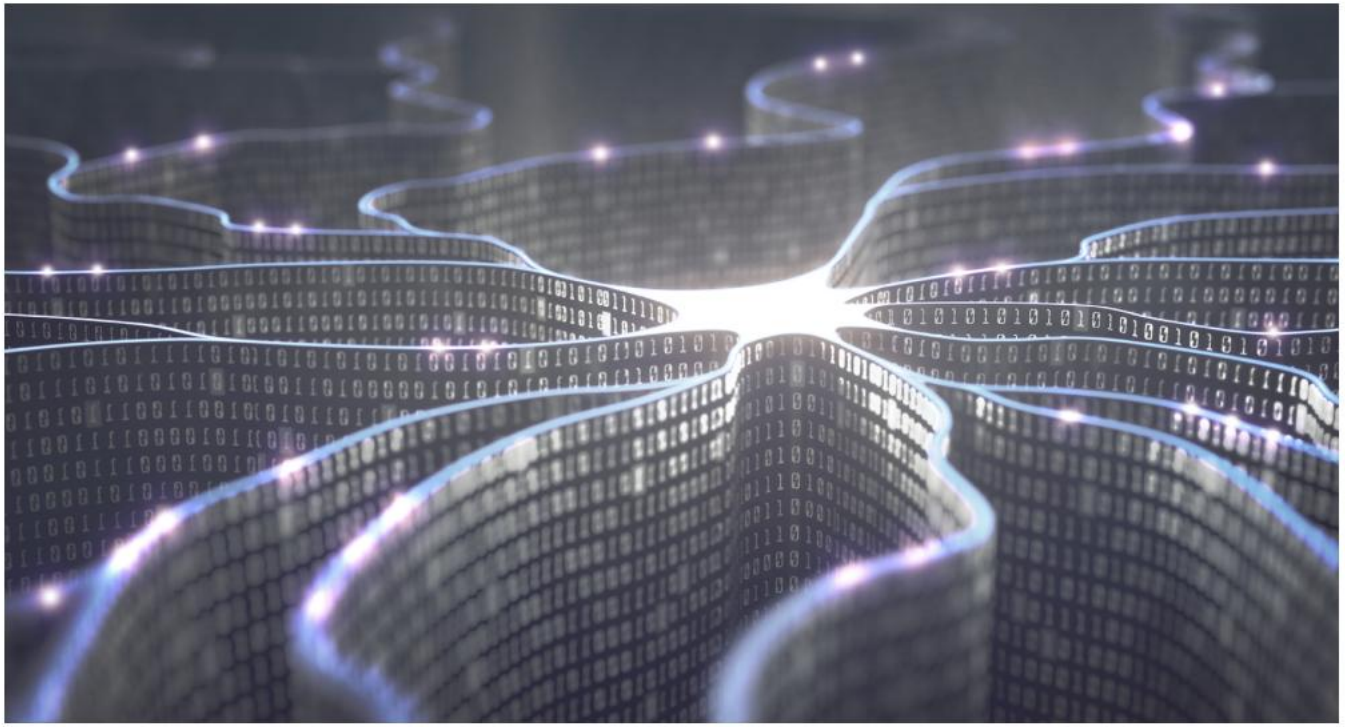
Trustworthy AI

1. **Human Agency and Oversight**: AI systems should empower individuals to make informed decisions while ensuring proper oversight through mechanisms.

What does "human oversight of A.I." really mean? Photo illustration by Slate. Photo by iLexx/iStock/Getty Images Plus and Anastasiia Makarevich/iStock/Getty Images Plus.

Trustworthy AI

2. **Technical Robustness and Safety**: AI systems must be resilient, secure, and reliable to minimize unintentional harm.



Crédit : ©ktsdesign – stock.adobe.com

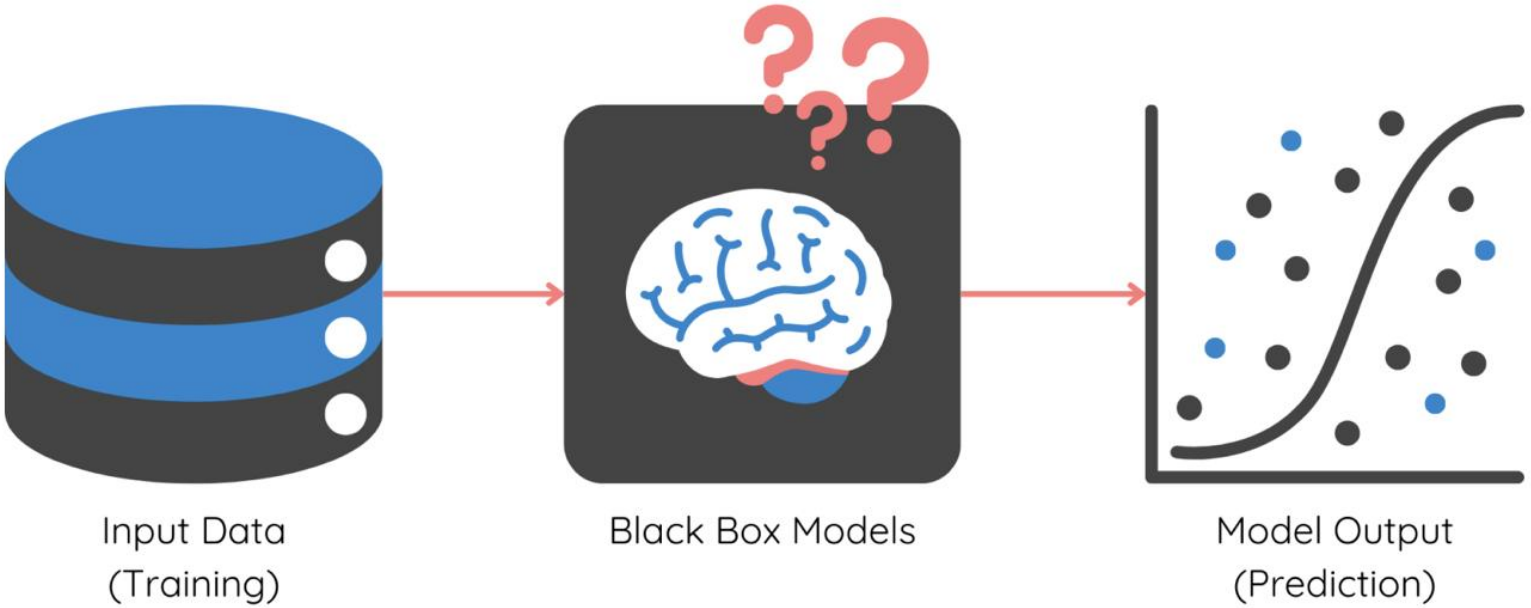
Trustworthy AI

3. **Privacy and Data Governance**: High standards for privacy and data protection should be maintained, along with robust data governance.

Image: ipopba/Adobe Stock

Trustworthy AI

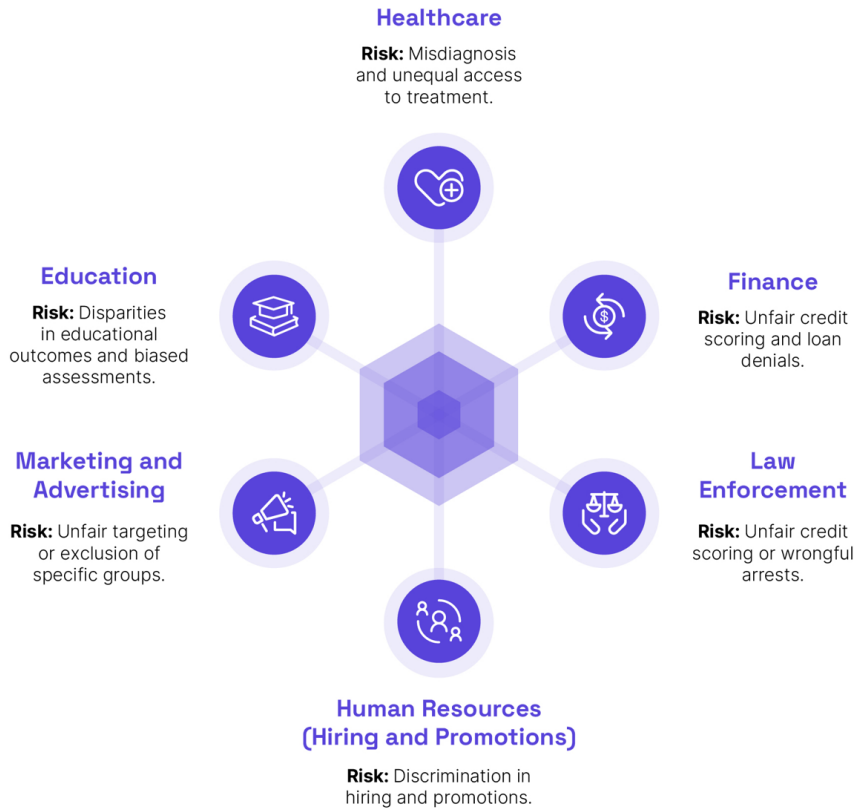
4. **Transparency**: Data, systems, and AI business models must be transparent, with clear explanations of AI decisions tailored to the relevant stakeholders.



(Source: Viso.AI)

Trustworthy AI

5. **Diversity, Non-discrimination, and Fairness**: AI should avoid bias, promote diversity, and be accessible to all.



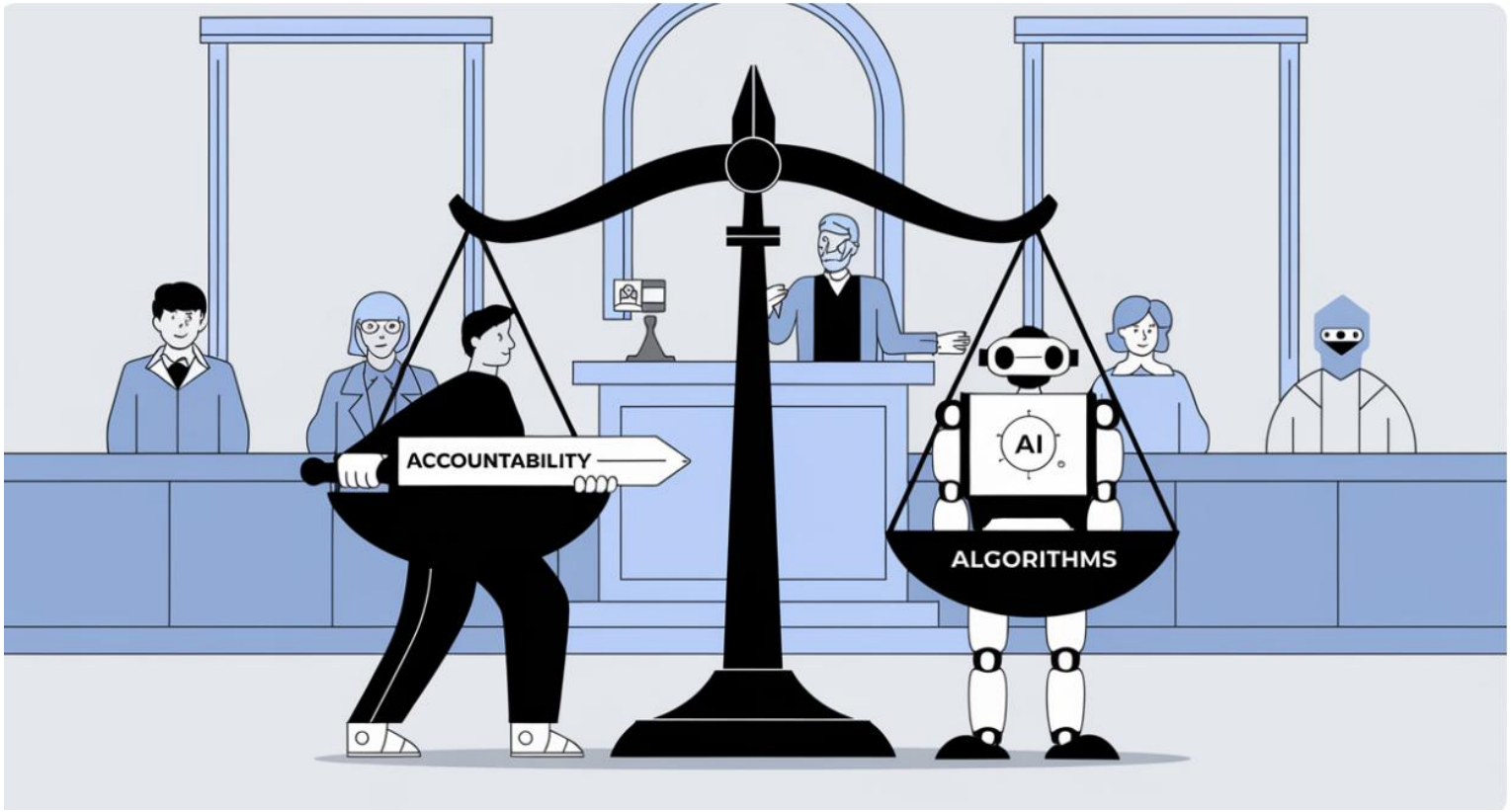
Trustworthy AI

6. **Societal and Environmental Well-being**: AI systems should be sustainable and environmentally friendly.

Source: ScienceDirect

Trustworthy AI

7. **Accountability**: Mechanisms for responsibility and accountability must be established for AI systems, including auditability for algorithms and accessible avenues for redress.



4 From Ethics to the Law

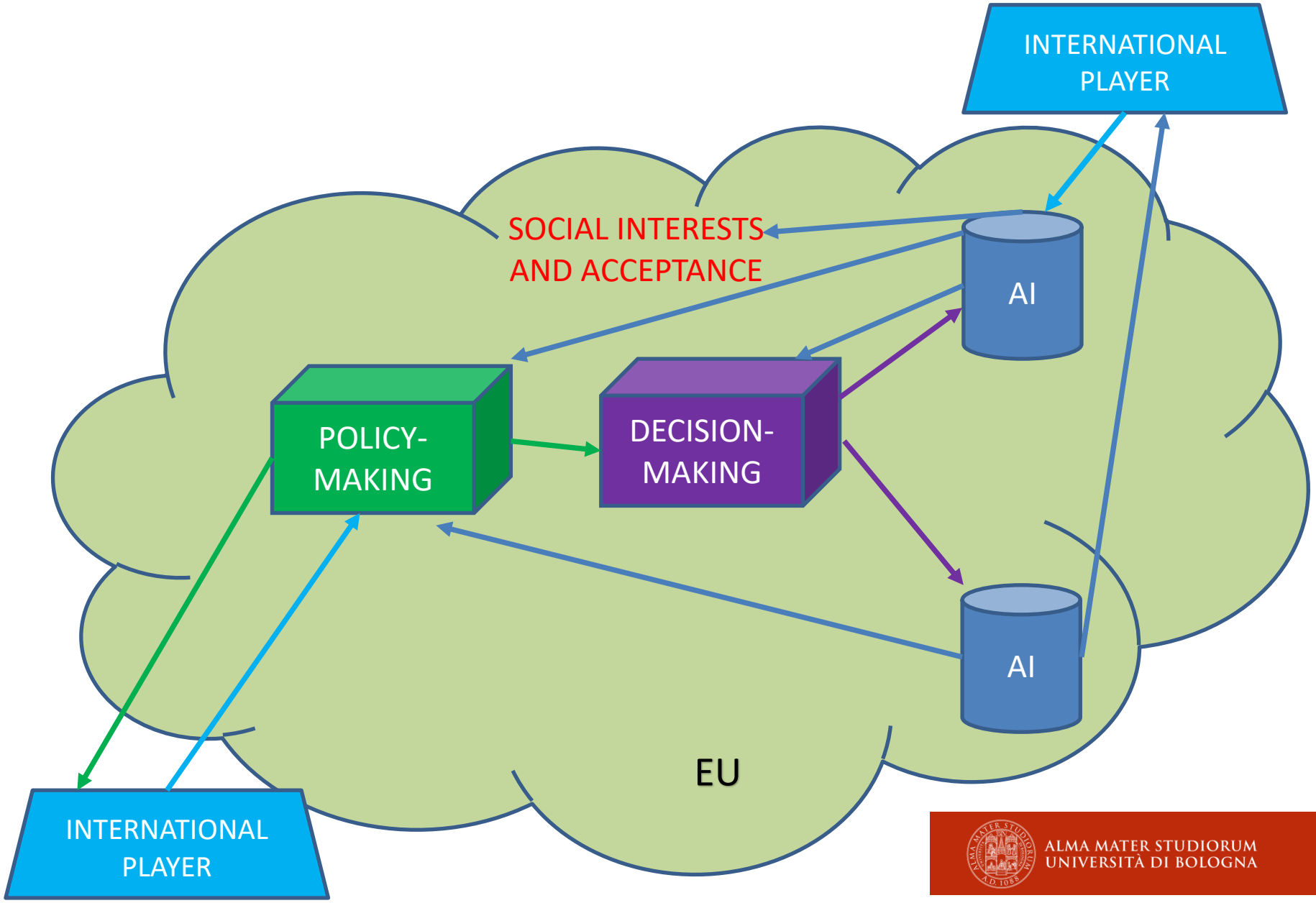
Impact of AI on the Law (An Example)

- Stephen Thaler, president and chief executive of *Imagination Engines*, claimed that an AI system called DABUS invented a new type of food container and a flashing light for attracting attention in emergencies.
- Thaler's legal team submitted applications to several patent offices, naming DABUS as the inventor.
- Patent registration offices have so far rejected the applications (UK, USA, Europe, South Korea, Taiwan, New Zealand and Australia).
- **AI-generated inventions are patentable?**
- **Who owns the IP rights? (Who gets the money?)**

The Law

The law is a system of standards, maxims, principles, or rules of conduct, to some degree accepted as obligatory by the persons to whom it is addressed, and backed by a specialized enforcement agency employing coercive sanctions (Max Weber)

How Does the Law Work?



Regulation Paradigms (1)

Regulating through sanctions for violations (hard law):

- Focuses on sanctioning non-compliance.
- Can be problematic due to AI's complexity.
- AI needed for monitoring and detecting violations.

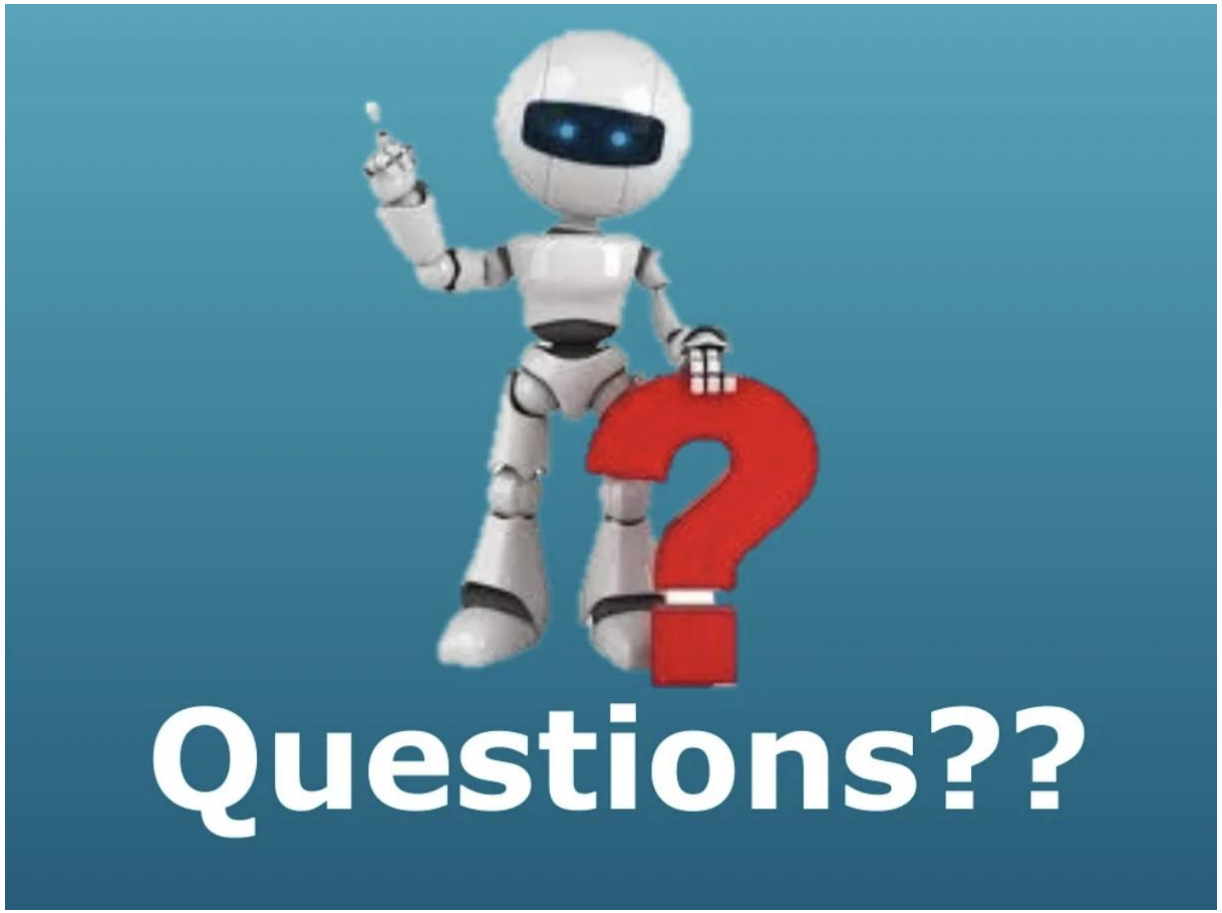
Legal compliance by design:

- Emphasizes incorporating legal requirements explicitly within AI systems.
- Ensures operation based on embedded legal knowledge.

Regulation Paradigms (2)

Compliance as risk mitigation:

- Outlined in the EU AI Act.
- Minimizes need for violation detection by emphasizing adherence to risk management procedures.
- Compensation for harm does not rely on traditional judicial methods, making sanctions a residual solution.
- Proposes alternative compliance mechanisms.



antonino.rotolo@unibo.it

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA