Liability regimes in the age of AI: a use-case driven analysis of the burden of proof

David Fernández-Llorca, Vicky Charisi, Ronan Hamon,

Ignacio Sánchez, Emilia Gómez

Joint Research Centre, European Commission

collaboration with DG JUST

European Commission

Motivation

- Artificial intelligence techniques have certain intrinsic properties linked to risks to safety and fundamental rights.
- Risk assessment & mitigation in the development stage.
- Harm occurring: victims should seek compensation.
- These same AI properties make difficult to prove causation.

Goal

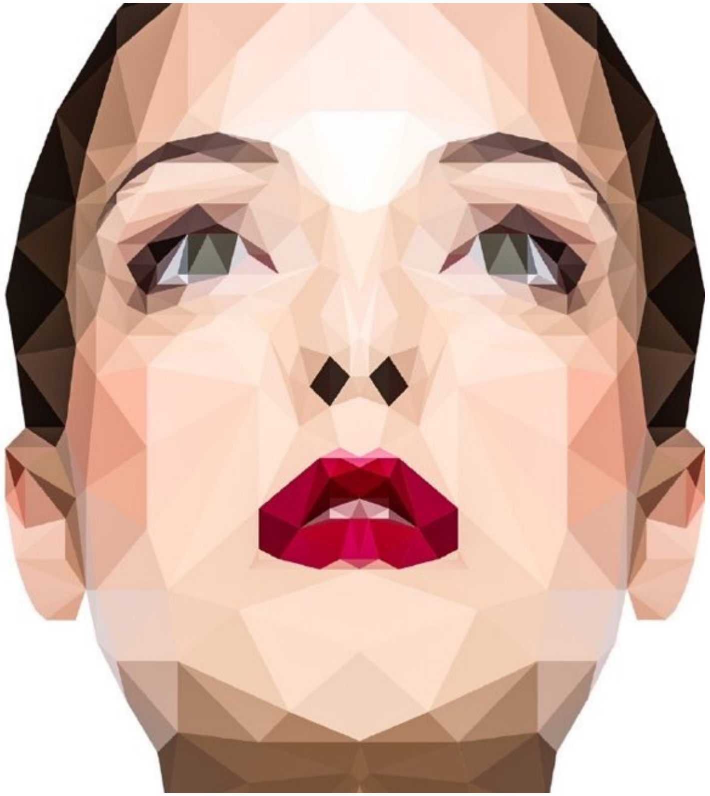
- Methodology to identify and describe a series of case studies on **harms produced by AI systems.**

- Study the **technical difficulties** in proving causation, i.e. *burden of proof* and the need to alleviate this burden for victims.

- Focus: systems able to produce physical & property damage, recent technological developments, potentially available in the short term, risks to third parties.

Human Behaviour and Machine Intelligence

1. advances the scientific understanding of **machine and human intelligence**,
2. studies the impact of algorithmic systems on **people and society**,
3. defines methodologies for **trustworthy** artificial intelligence,
4. provides scientific contributions to related European **policies**.



https://ai-watch.ec.europa.eu/humaint_en
#humaint

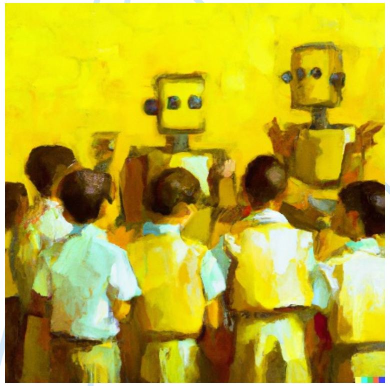
Current topics

https://ai-watch.ec.europa.eu/humaint_en
#humaint

facial processing

recommender systems

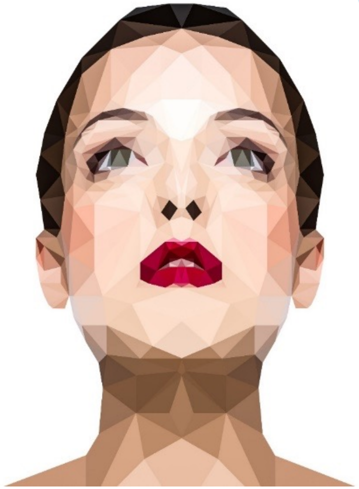
children



autonomous driving

AI in education, science
Policy design

AI Liability & Product Liability Directive
Proposal



AI Act
Negotiation

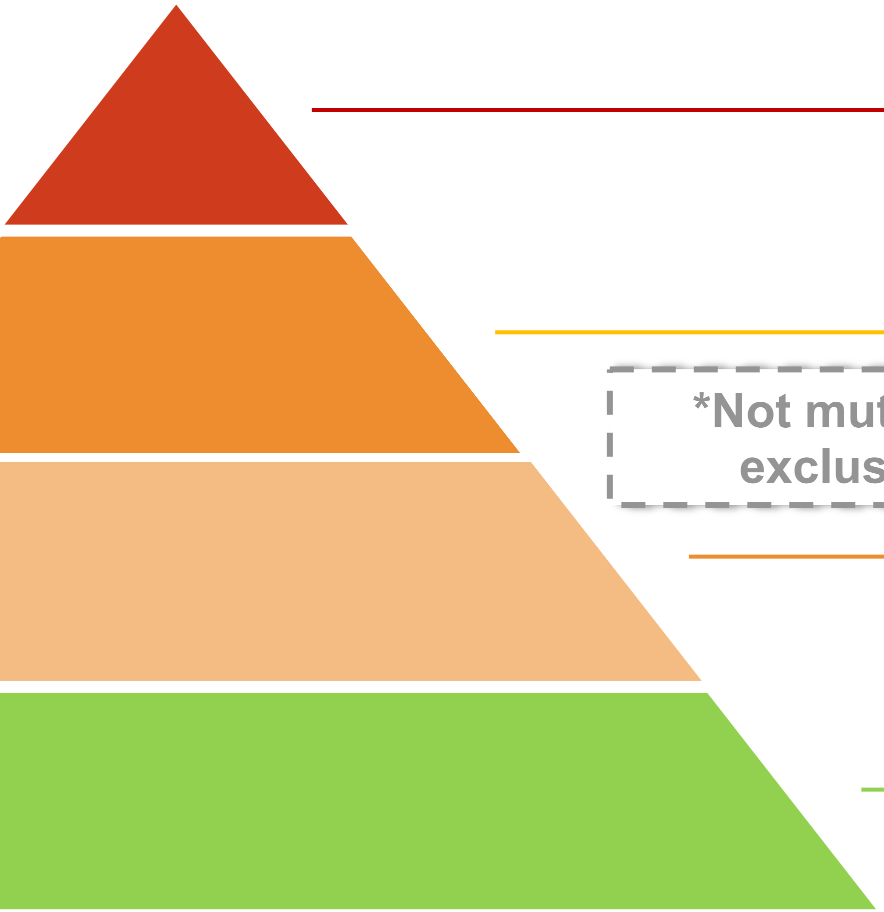
DSA Implementation
(European Centre for Algorithmic Transparency)

Images created with DALLE@openAI

European Commission

Artificial Intelligence (AI) Act

Scope: software products with AI.



*Not mutually exclusive

Unacceptable risk	**Prohibited**
High risk	**Permitted** subject to compliance with AI requirements and ex-ante conformity assessment
'Transparency' risk	**Permitted** but subject to information/transparency obligations
Minimal or no risk	**Permitted** with no restrictions

https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

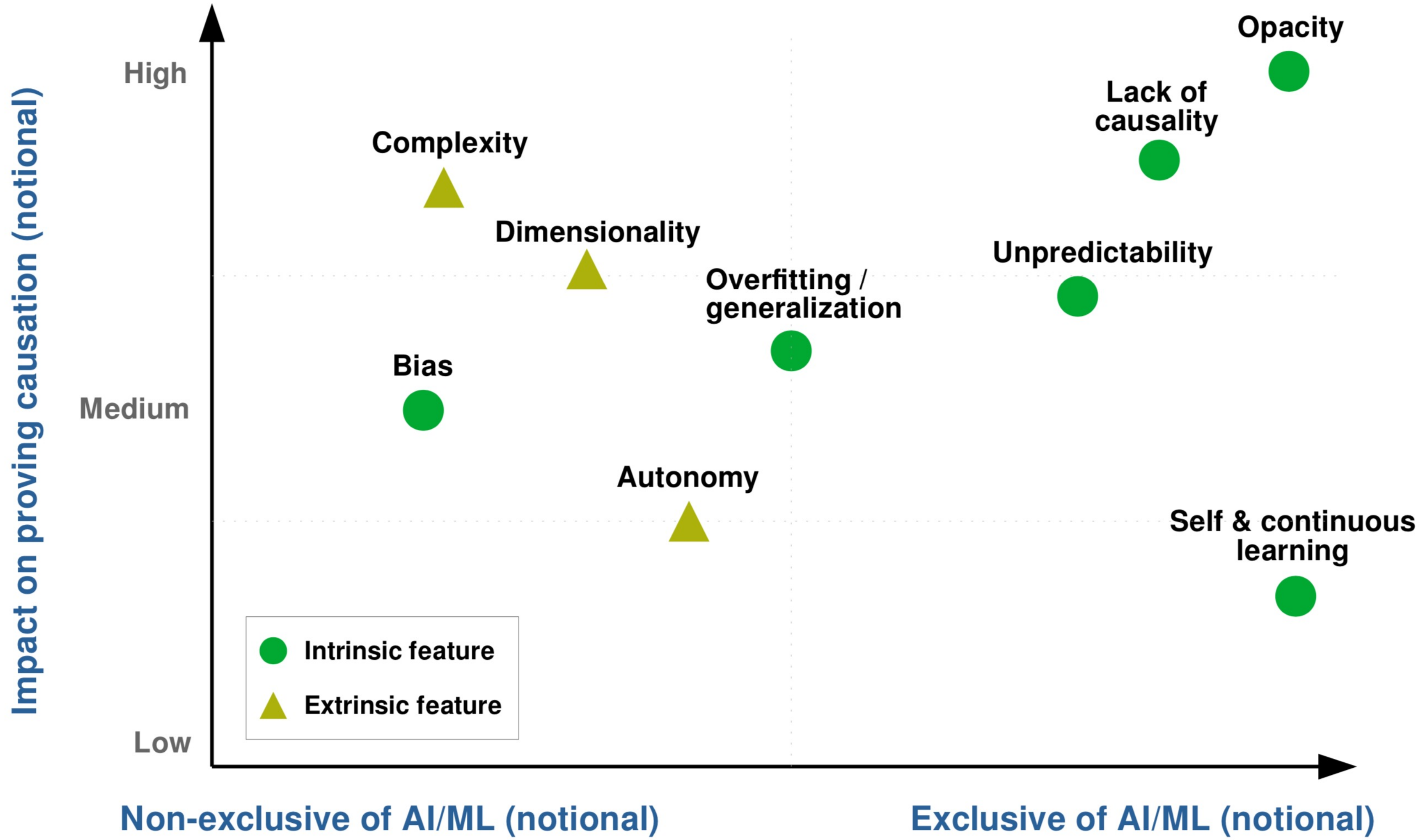
Obtaining compensation for product-induced damages

- Legal frameworks (EC, 2019)
- Fault-based liability: injured parties have to prove that the defendant caused the damage intentionally or negligently.
    - Identify the ***standard of care*** the defendant should have fulfilled.
    - Prove it was not fulfilled.
    - Negligent design, manufacturing, maintenance, marketing, operation or use.
- Strict-based liability, risk based: injured parties only need to prove that a risk materialised.
- Product-based liability: victims can claim for a defect present at the time the product was placed into market. ***Standard of safety***. Defective design, manufacturing, …

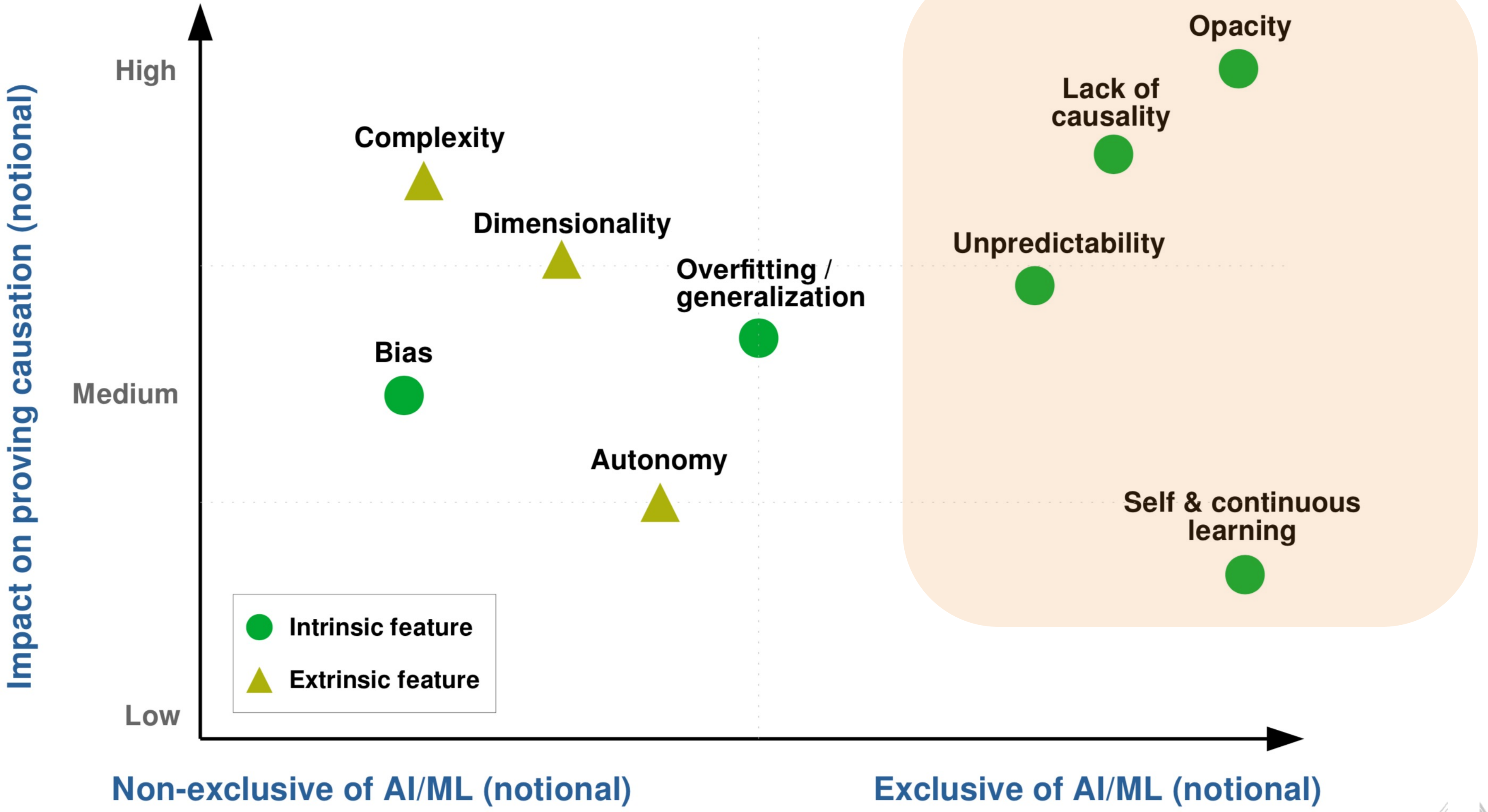
Relevant literature

- Lack of legal personality of AI systems (Gerka, Grigiene & Sirbikyte 2015)
- Person operating an AI tool as responsible (Sullivan & Schweikart, 2019)
- Challenges when AI becomes autonomous (Shook, Smith & Antonio, 2018)
- Harms attributable to existing persons or organizations (Abott & Sarch, 2019)
- Standard of care (*fault-based*) → standard of safety (*strict liability*), complexity of the value chain.

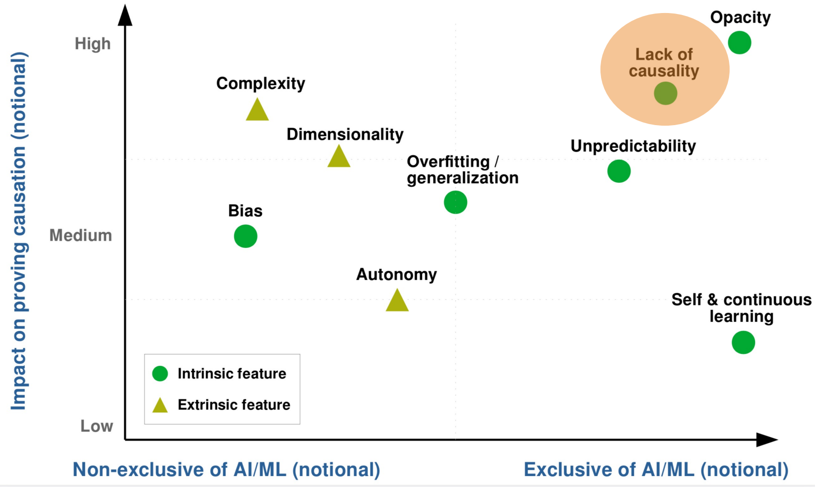
Characteristics of AI systems



Characteristics of AI systems



1. Lack of causality



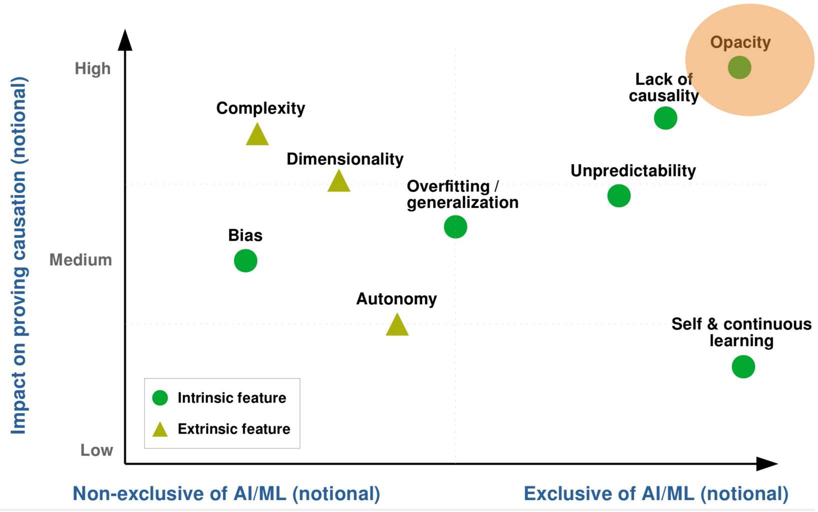
- Difference between statistical associations vs causation.

- *Independent and Identically Distributed* (i.i.d) assumption in machine learning (Schölkopf et al., 2021) → poor performance of models when different statistical distributions in real-world operation vs training, e.g. adversarial attacks.

- Despite research advancements, learning causal relationships still challenging (Schölkopf et al., 2021)

European Commission

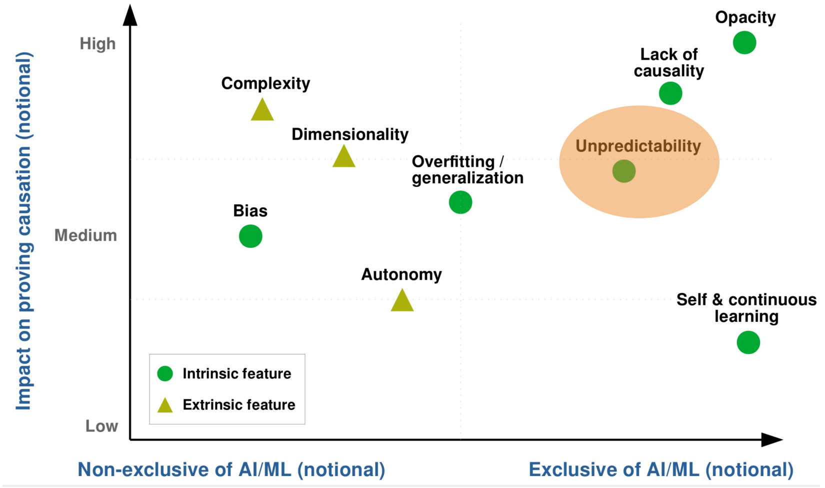
2. Opacity



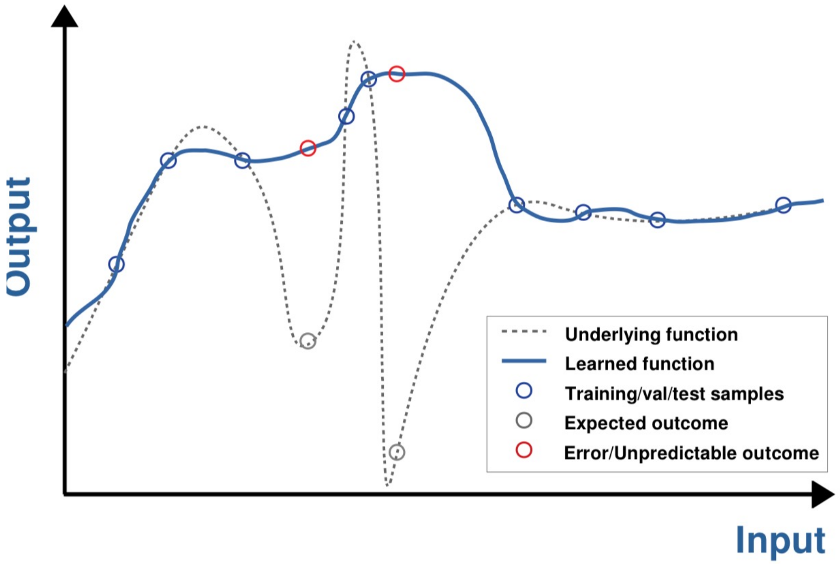
Obscurity of meaning, resistance to interpretation.

- Black-box character of the decision making process with ML and inability to provide human scale reasoning from complex models (Burrell, 2016).
- Transparency requirements (AI Act) alleviate the burden of proving causality.
- Attempts to explain black-box ML models might not be sufficient to demonstrate causality (Rudin, 2019).

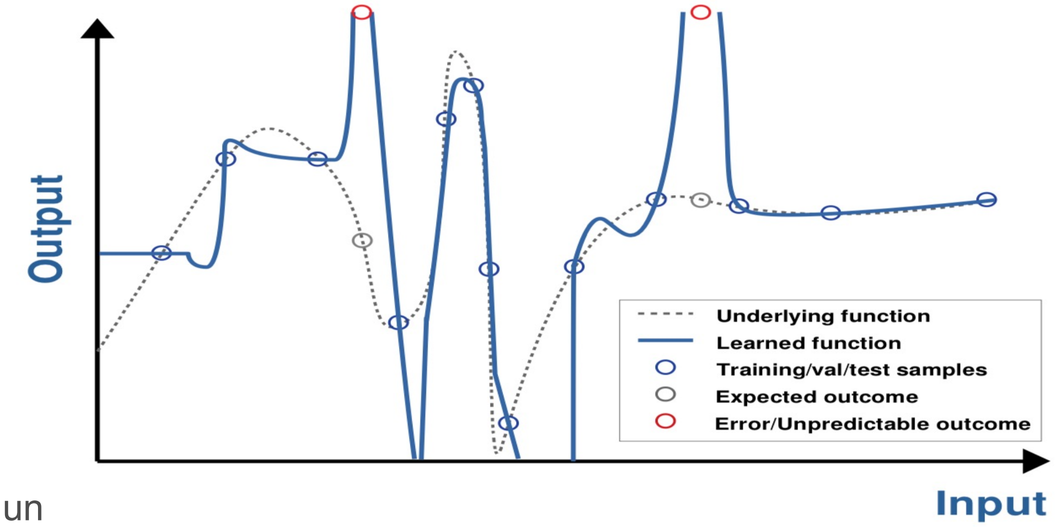
3. Unpredictability



Dataset not sufficiently. Solutions in poorly represented regions generate unpredictable results.



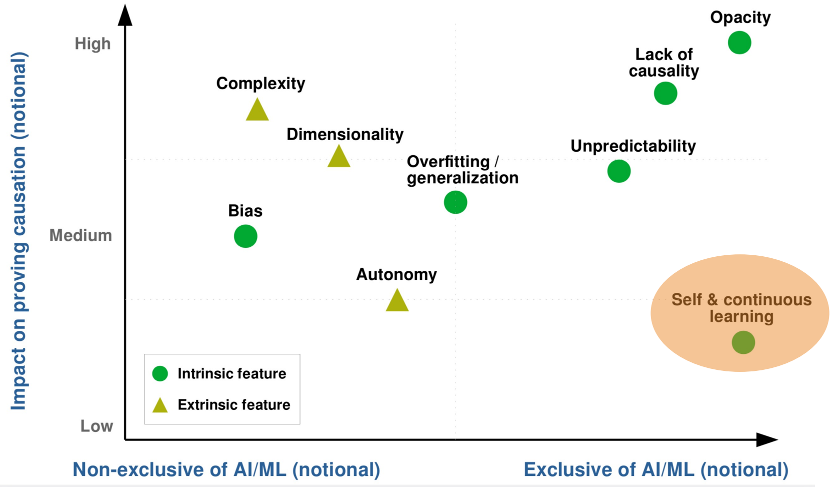
Overfitting, even if the input space is well represented. The outcome for samples not used in the training is unpredictable.



Recurrent models: output depends on input and state. Source of un predictability.

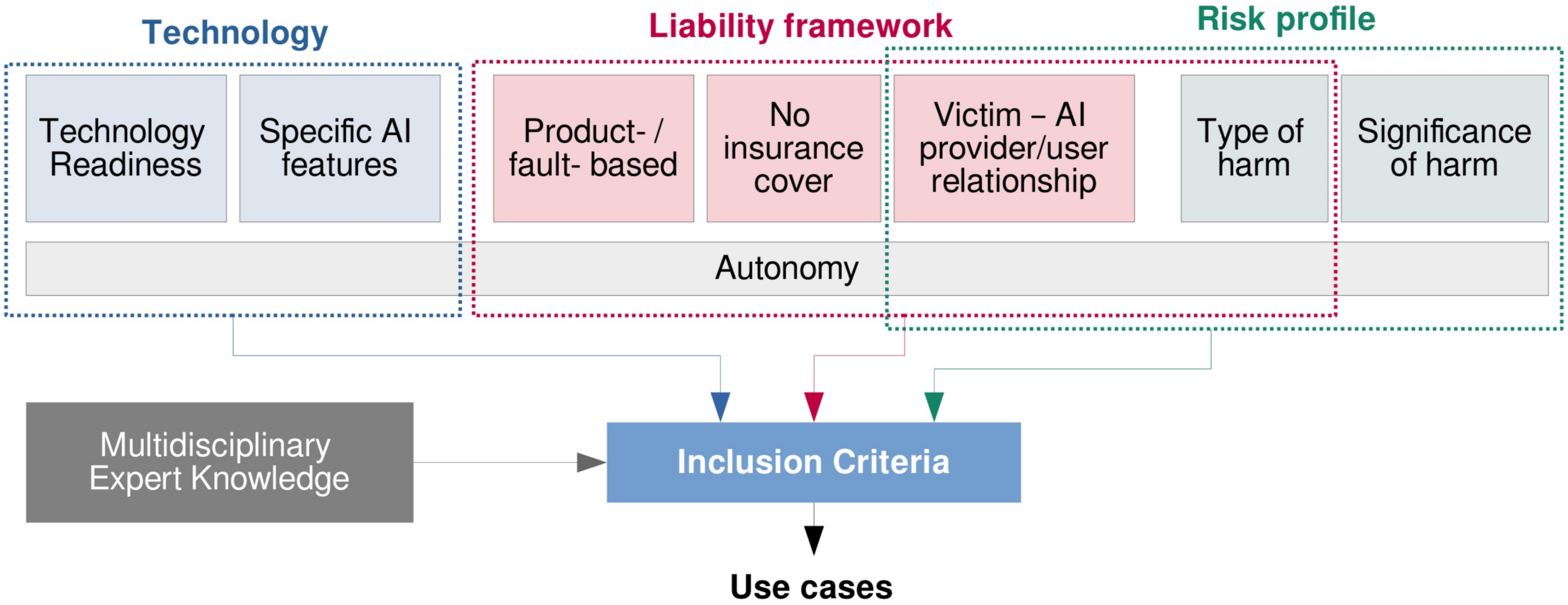
European Commission

4. Self and continuous learning

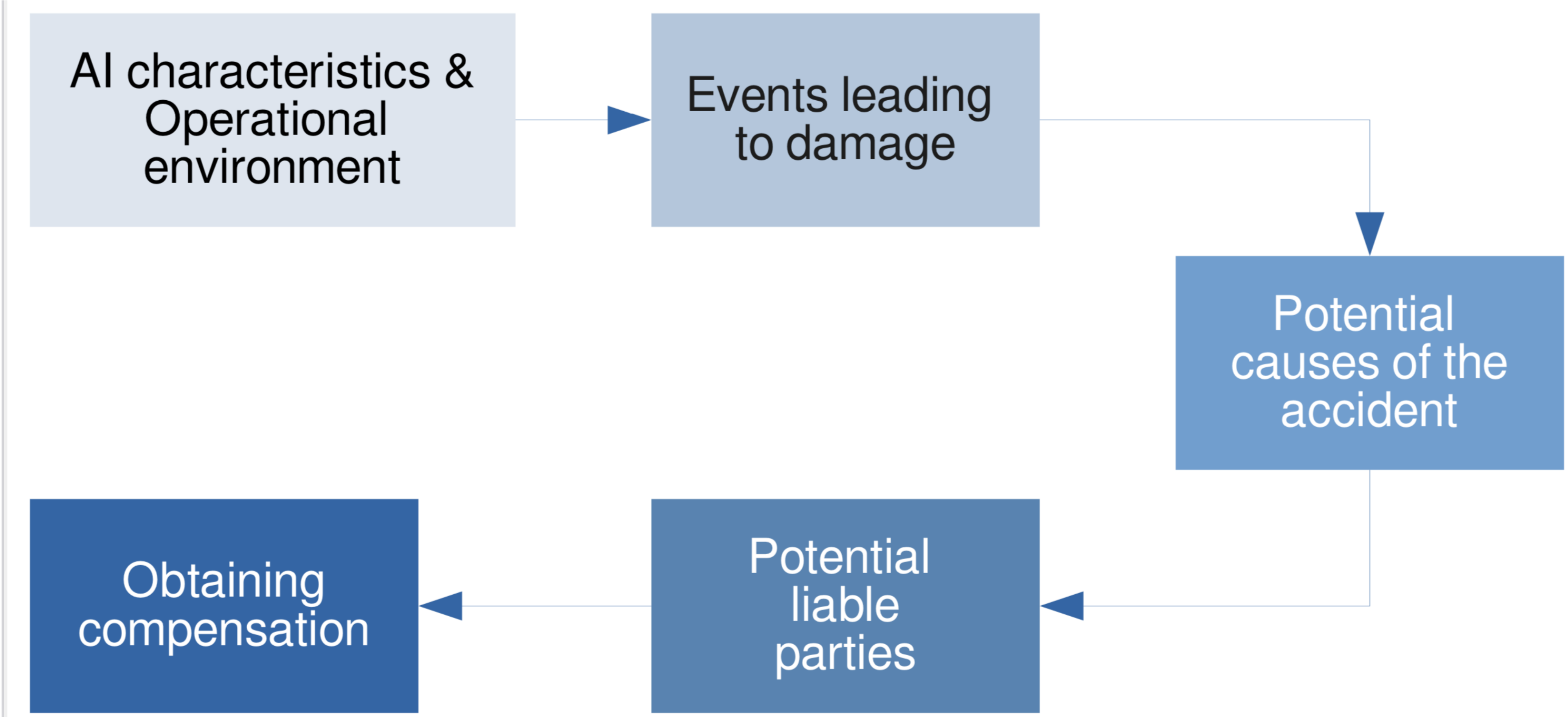


- Incremental training of the AI system during the operation phase (*online learning*).
- Catastrophic forgetting: learning new patterns can interfere model´s knowledge (French, 19909).
- Related to the question of foreseeability.
- Substantial modifications: new conformity assessment (AI Act).

Inclusion criteria for case studies



Methodology

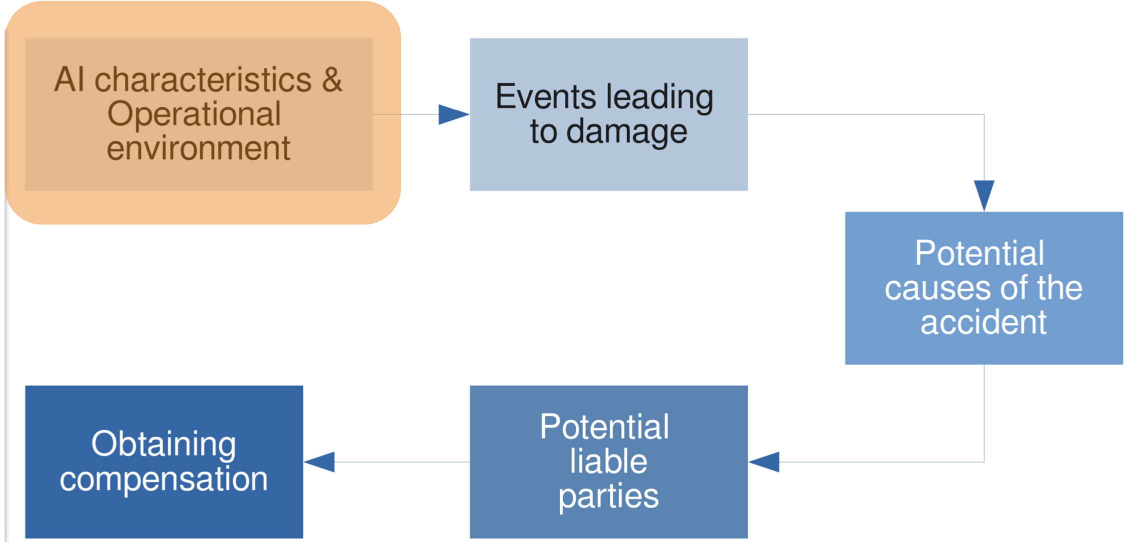


Autonomous Urban Cleaning robot

Figure 5: From left to right, three examples of the current state of this kind of technology: the systems developed by ENWAY (ENWAY, 2021), Trombia (Trombia, 2020) and Boschung (Boschung, 2020).

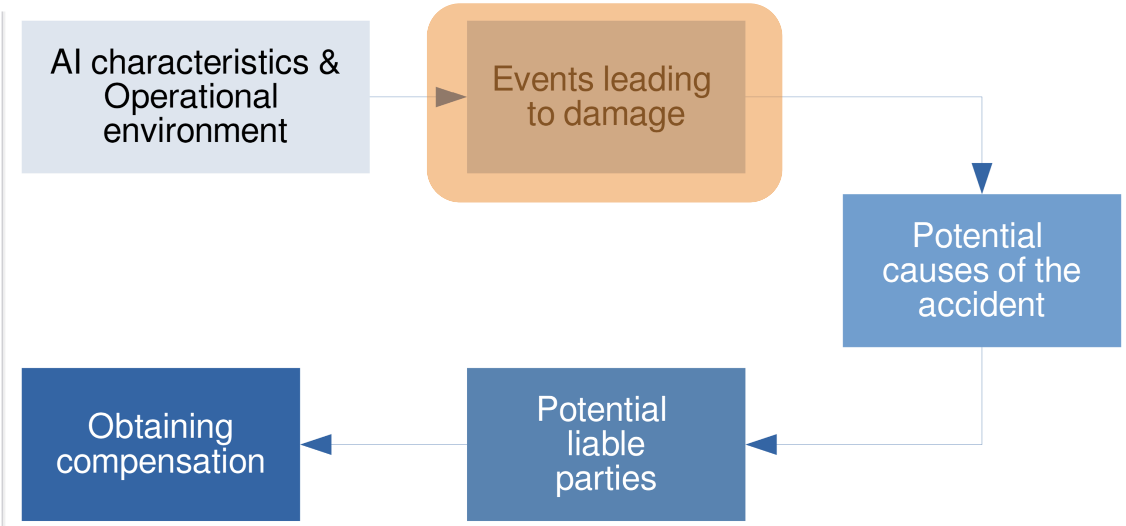
Autonomous Urban Cleaning Robot

- Product: sensors, digital information, connectivity features, communication systems, actuators.
- AI/ML systems: perception systems, robot localization and mapping, detection of obstacles, trajectory planning, lateral and longitudinal control of the platform, etc.
- Human operator: supervisory role.



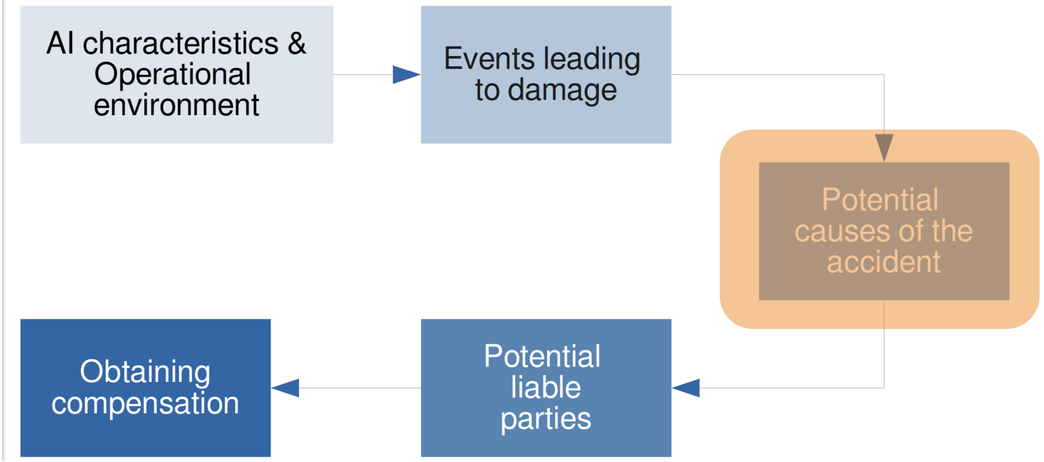
Autonomous Urban Cleaning Robot

A colourful baby stroller is parked in front of an advertising banner with similar colour patterns while the baby´s guardian looks at a nearby shop window. One of the cleaning robots .. collides with it. The stroller is damaged and the baby slightly injured.



Autonomous Urban Cleaning Robot

- Failure of a component: perception module (wrong image segmentation), decision making and control (wrong reaction time), sensors,..
- Potential causes:
  - Mislabelling in training data, inadequate lighting, unfavourable weather conditions,….
  - Deliberate attack potentially exploiting vulnerabilities.

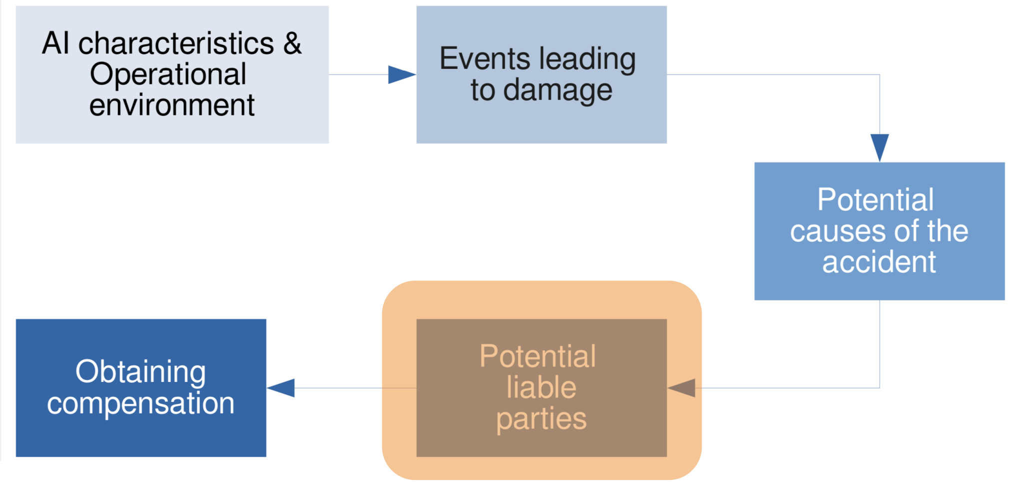


Autonomous Urban Cleaning Robot

Due to a single component, several components or faulty integration.

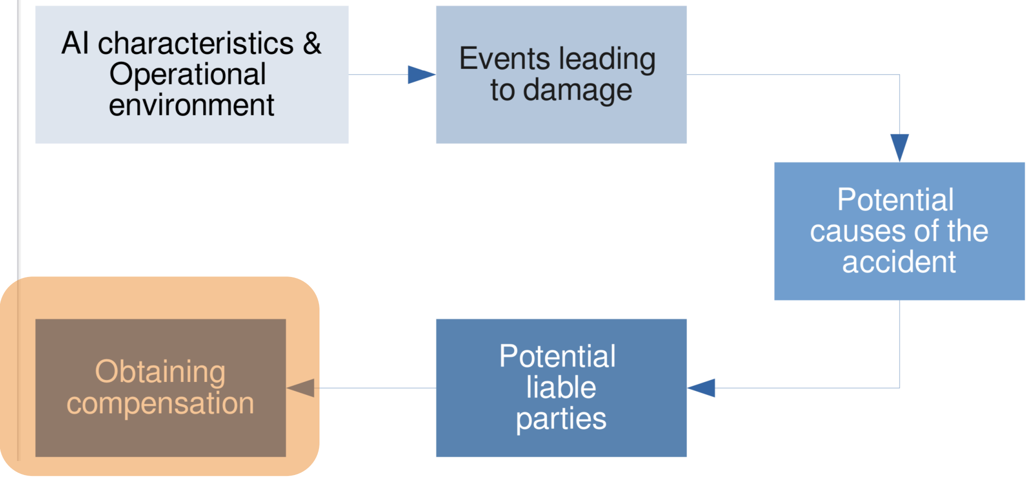
Parties:

1. Robot manufacturer.
2. Provider of individual AI components.
3. Professional user or operator (e.g. municipality).
4. Adversaries attacking the system.



Autonomous Urban Cleaning Robot

- Experts should assess all potential causes to establish *prima facie* evidence.
- Correlation proved: we cannot discard alternative sources of the damage.
- Impossibility to infer a clear causal link input – harmful output.
- Expert information needs: logs, technical documentation.



Use cases

Autonomous delivery drones: physical harm, property dammage.

Robots in education: physical harm, property dammage, phychological harm.

Figure 6: From left to right, three examples of the current state of this kind of technology: the systems developed by Wing (Wing, 2022), Amazon Prime Air (Amazon Prime Air, 2022) and Zipline (Zipline, 2022).

Figure 7: From left-to-right, top-to-bottom, five different robotic platforms in the context of education: De-Enigma (De-Enigma, 2019), Pepper (BBC News, 2021), QTrobot from LuxAI (LuxAI, 2019), Nao robot (Zhang et al., 2019) and Haru (Charisi et al., 2020).

Conclusions

- We highlighted the technical difficulties that an expert opinion would face in trying to prove defect or negligence, and the causal link to damage.
- Liability regimes should be revised to alleviate the burden of proof on victims in cases involving AI systems.

AI liability directive

- Part of a package: AI Act, revision of product safety rules.
- *Harmonisation of national liability claims based on the fault of any person with a view of compensating any type of damage.*
  - measures to ease the burden of proof:
    - Disclose of evidence (Article 3) on high-risk AI systems: technical documentation, logs.
    - Rebuttable presumption of causal link in the case of fault (Article 4)
  - a review mechanism to re-assess the need for harmonising strict liability for AI use cases with a particular risk profile (possibly coupled with a mandatory insurance).

https://commission.europa.eu/business-economy-euro/doing-business-eu/contract-rules/digital-contracts/liability-rules-artificial-intelligence_en

Liability regimes in the age of AI: a use-case driven analysis of the burden of proof

David Fernández-Llorca, Vicky Charisi, Ronan Hamon,

Ignacio Sánchez, Emilia Gómez

Joint Research Centre, European Commission

collaboration with DG JUST

European Commission