

# Cybersecurity of AI: technological challenges and opportunities

Henrik Junklewitz, Ronan Hamon, Ignacio Sanchez

*European Commission Joint Research Centre (JRC)*

ENISA AI Cybersecurity Conference 2023, Brussels

# JRC - Science for policy



ANTICIPATE



INTEGRATE

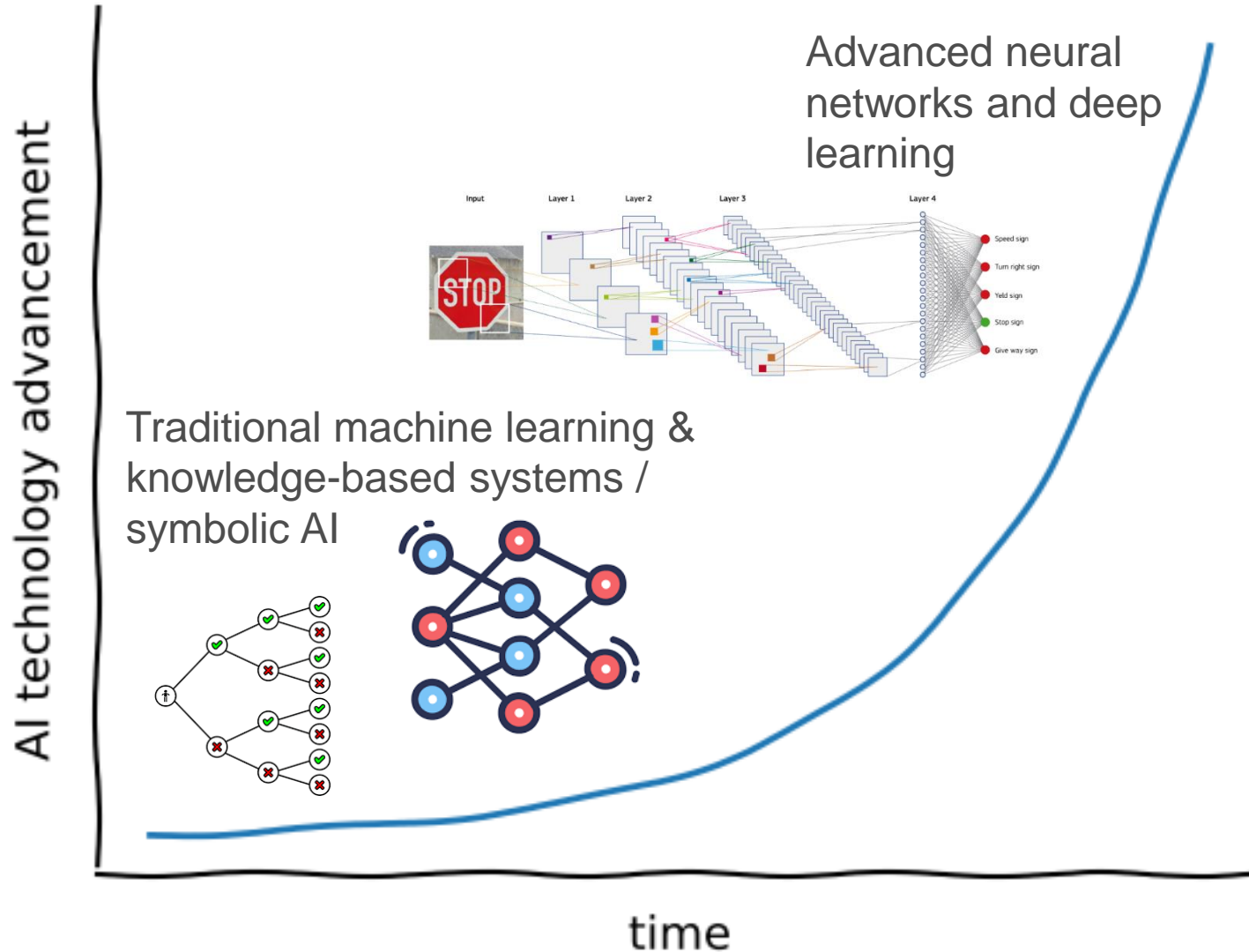


IMPACT

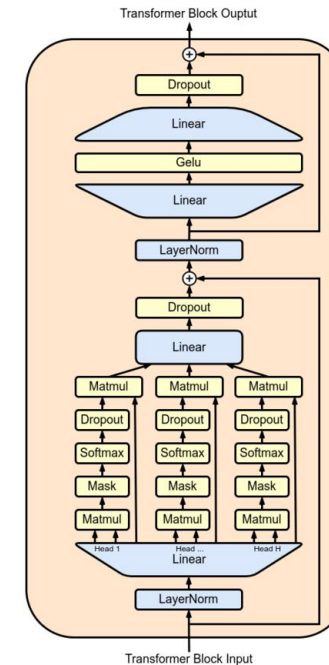
## Our purpose

The Joint Research Centre provides independent, evidence-based knowledge and science, supporting EU policies to positively impact society.

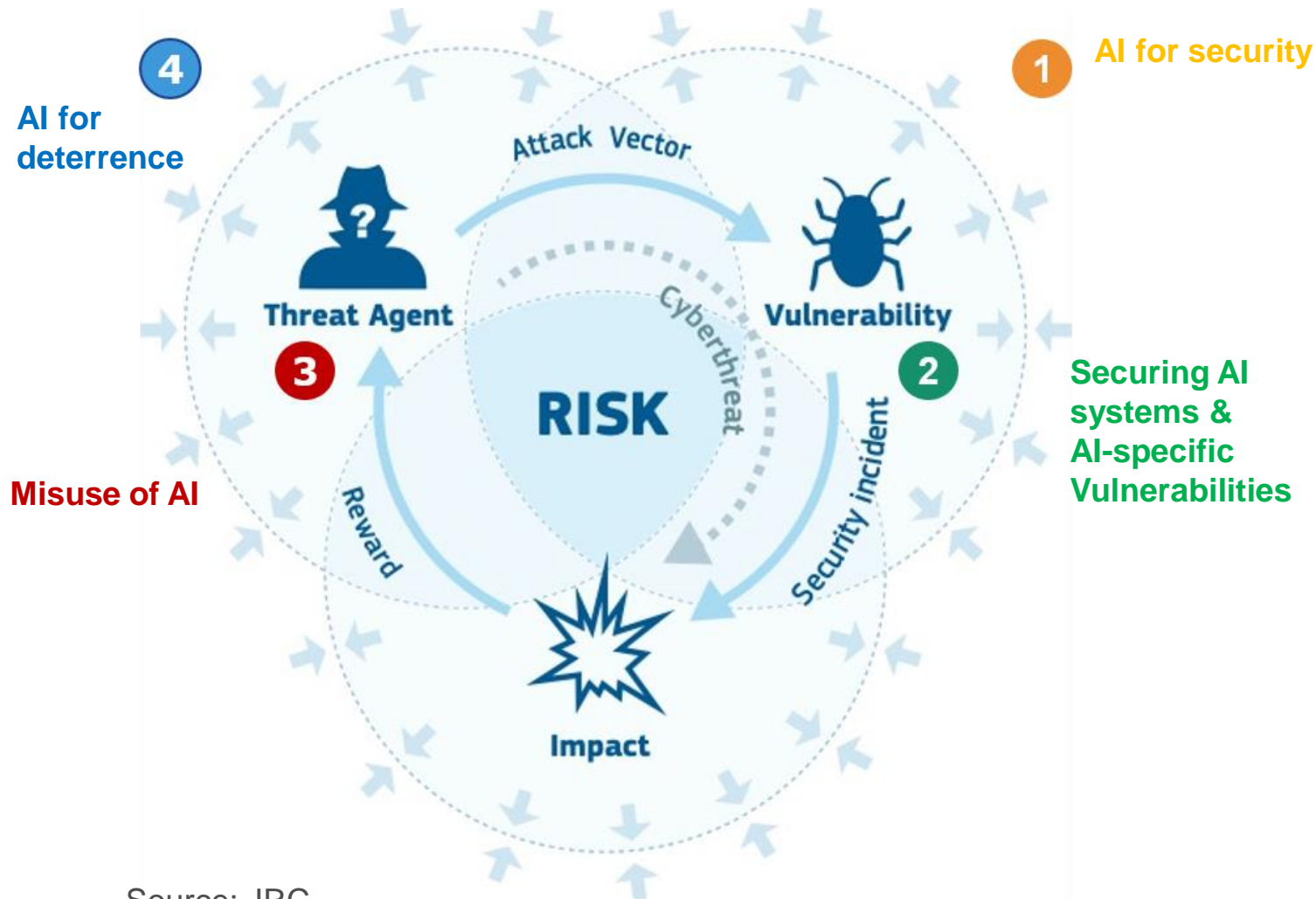
# AI is rapidly developing – many concurrent technology levels



## Large-scale models



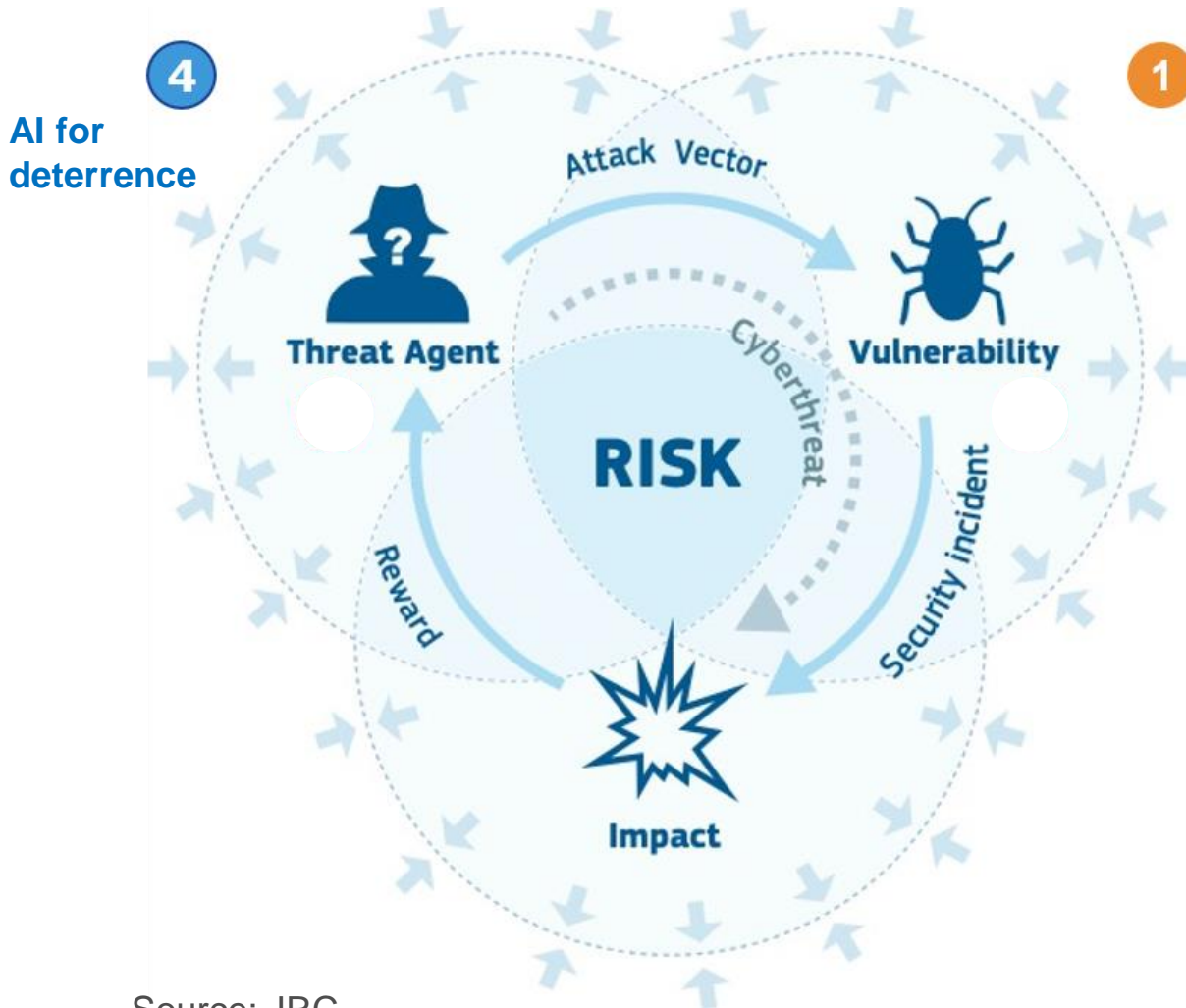
# AI Cybersecurity – Risk analysis brings challenges and opportunities



Source: JRC

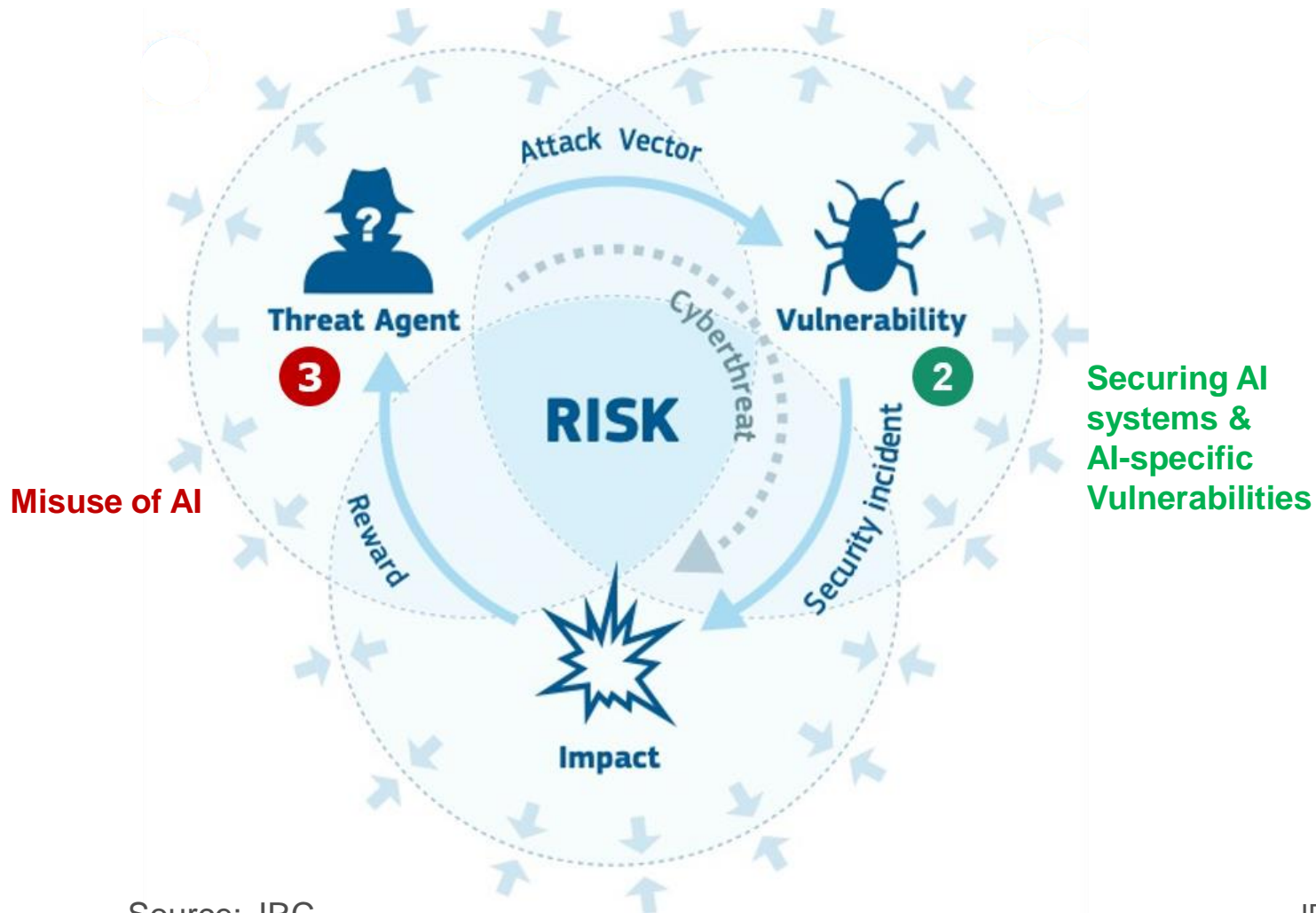
Source: [JRC Cybersecurity Flagship Report 2020](#)

# AI Cybersecurity – Technological opportunities



- **AI for cybersecurity:** AI-based security controls, threat intelligence, or defences
- **AI for deterrence:** AI tools for law enforcement, digital forensics, etc.

# AI Cybersecurity – Technological Challenges

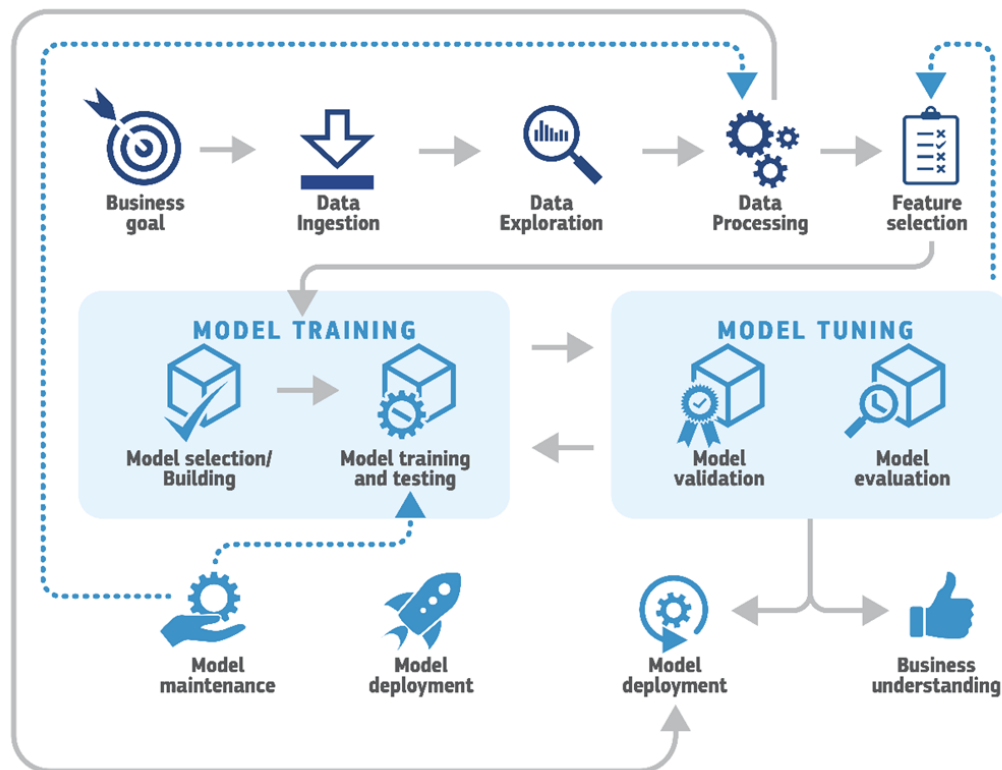


Source: JRC

- **Preventing Misuse of AI:** currently very **hard** at the technological / AI-model stage **for all technological levels of AI.**
- **Securing AI systems and AI-specific vulnerabilities:** many organisational measures possible, AI-model specific security controls and vulnerability handling **can be complex depending on the technological levels of AI.**

# Securing AI and AI-specific vulnerabilities – Organisational challenges

- **Harmonising terminologies**, taxonomies and scope of definitions across fields and standards
- **Managing AI-lifecycle security**, including **AI specific supply chain security** (large-scale data, pretrained models) and **developing system-level security controls** for AI software
- **Raising AI Cybersecurity Awareness and Competences**



Source: JRC

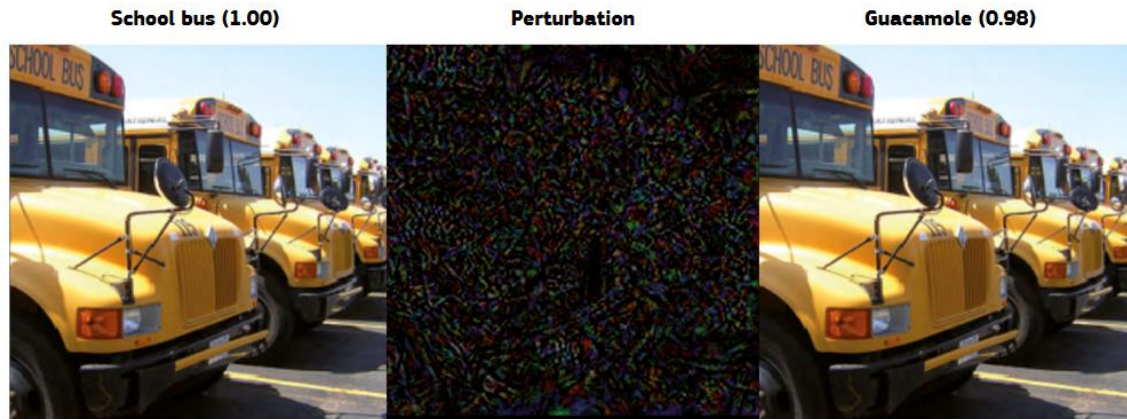
# Securing AI and AI-specific vulnerabilities – R&D challenges

- **Many active fields of research on AI-specific topics** (e.g., adversarial machine learning, prompt injection attacks) with a **host of attacks specific to ML models without proper security handling in practice** (evasion , poisoning, backdoors, model inference, ...)
- **Building AI-specific defences** and hardening against attacks
- **Defining metrics for AI cybersecurity, measuring adversarial robustness, and assessing trade-offs** with other requirements such as accuracy and transparency
- **AI threat modelling practice missing** in many cases due to the lack of experience in real-world deployment
- Technical and scientific innovation is still needed to address these challenges

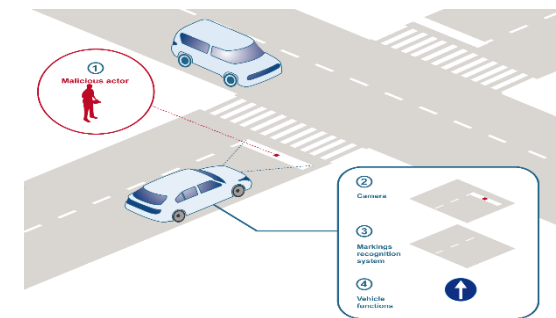
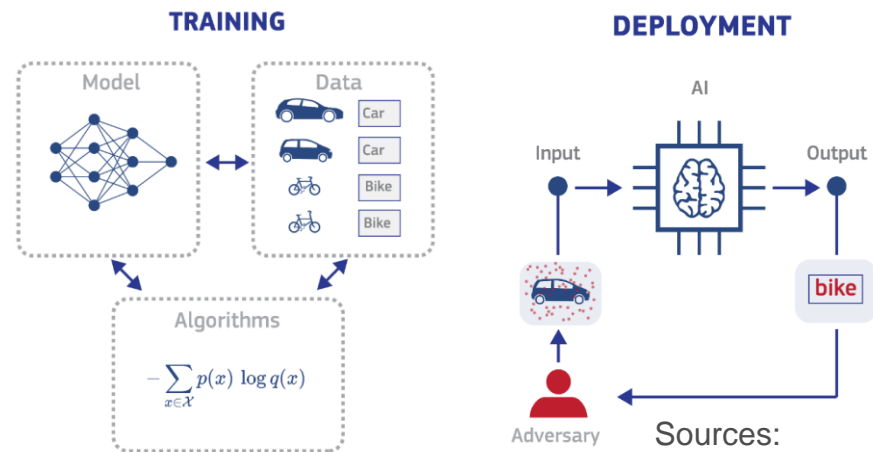


# Securing AI and AI-specific vulnerabilities – Example: Evasion Attacks

## Digital Attack

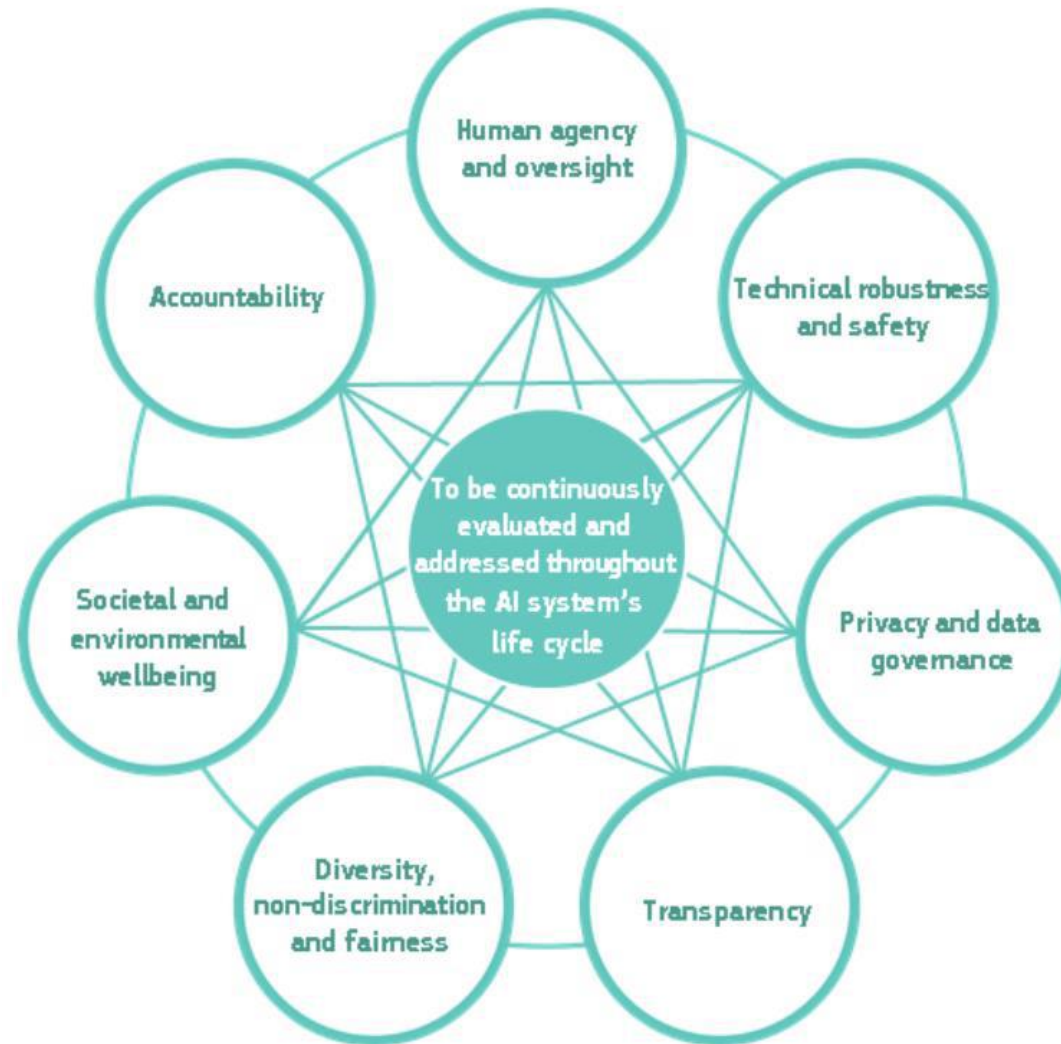


## Physical Attack



Sources:  
JRC-ENISA Report on Cybersecurity Challenges in AI and Autonomous Driving 2021  
JRC Scientific projects

# AI Regulation is coming – Trustworthy AI - challenges as an opportunity



# Thank you and keep in touch





© European Union 2023

Unless otherwise noted the reuse of this presentation is authorised under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license. For any use or reproduction of elements that are not owned by the EU, permission may need to be sought directly from the respective right holders.

## EU Science Hub

[joint-research-centre.ec.europa.eu](https://joint-research-centre.ec.europa.eu)

-  @EU\_ScienceHub
-  EU Science Hub – Joint Research Centre
-  EU Science, Research and Innovation
-  EU Science Hub
-  @eu\_science