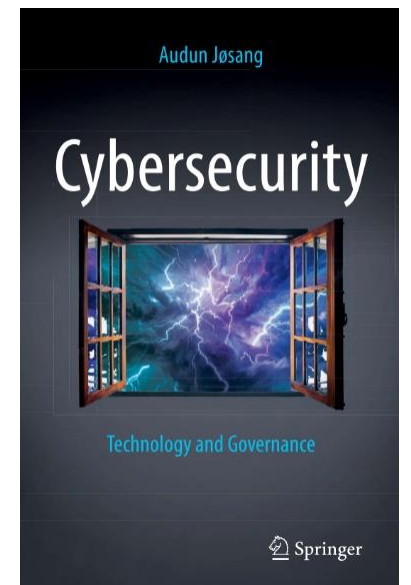


Chapter 15: AI and Cybersecurity

Cybersecurity: Technology and Governance

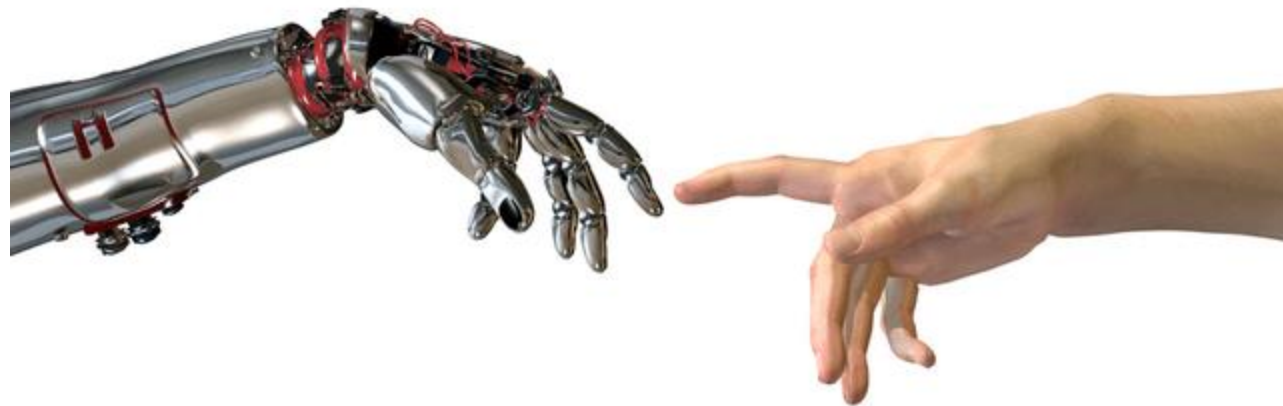
Audun Jøsang,

1st Edition, Springer Nature, 2024.



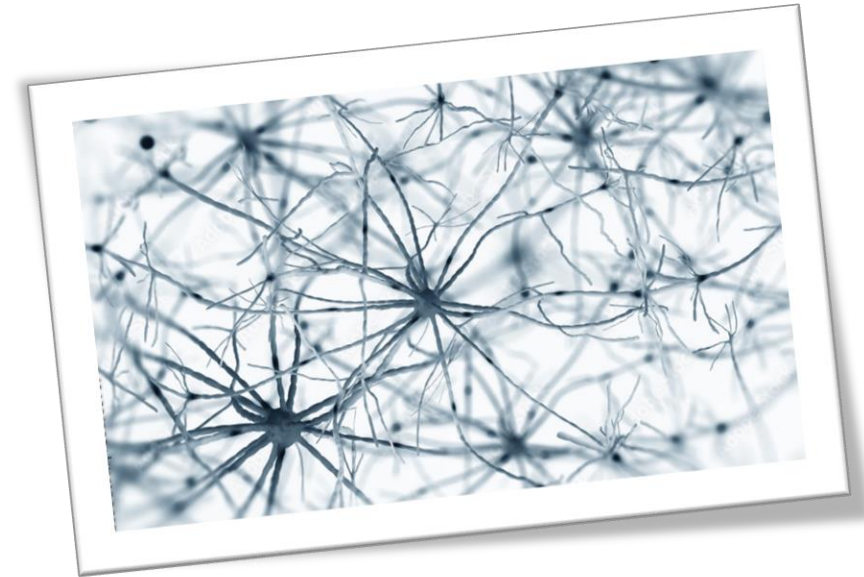
Overview

- What is AI?
- Offensive AI
- Defensive AI
- Vulnerabilities and attacks against AI
- AI regulation



What is AI?

- AI workshop in Dartmouth, Massachusetts, USA 1956
-
- AI winters, failed approaches:
 - Logic reasoning
 - Expert systems
 -
- The development picked up speed from about 2012
 - Artificial neural networks
 -
- Development exploded with ChatGPT in 2022



What is AI?

AI
Artificial
Intelligence

- AI is computer applications that performs tasks that normally (traditionally) require human intelligence

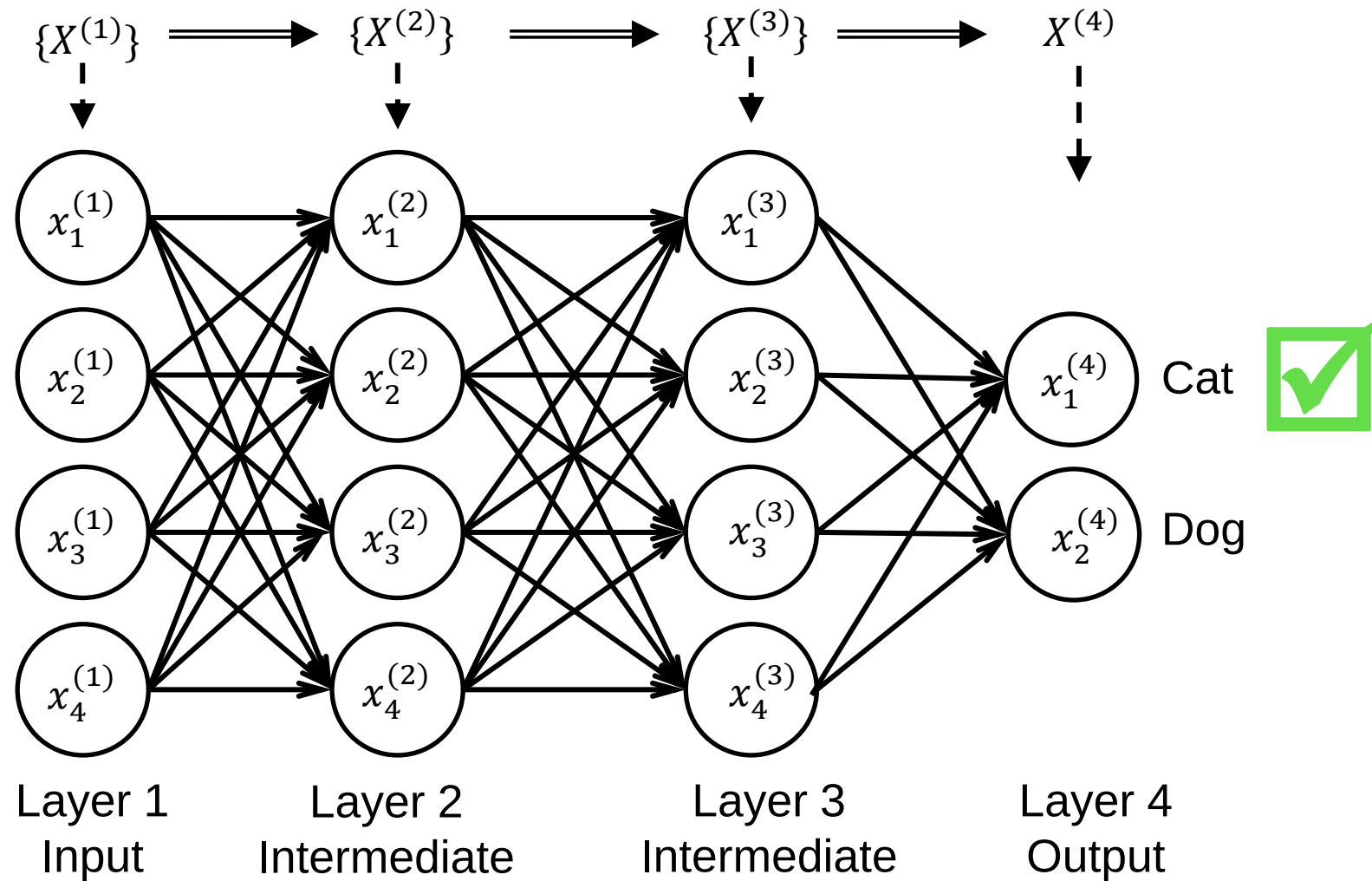
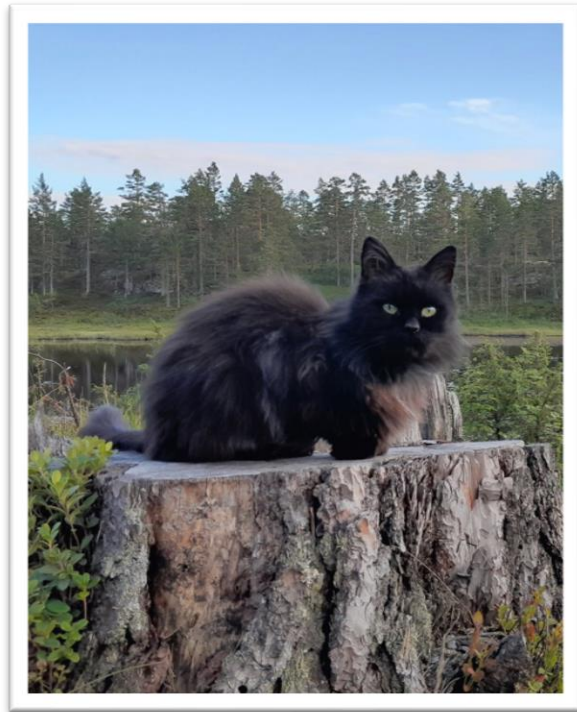
ML
Machine
Learning

- ML is computer applications that automatically learns from datasets without being explicitly programmed

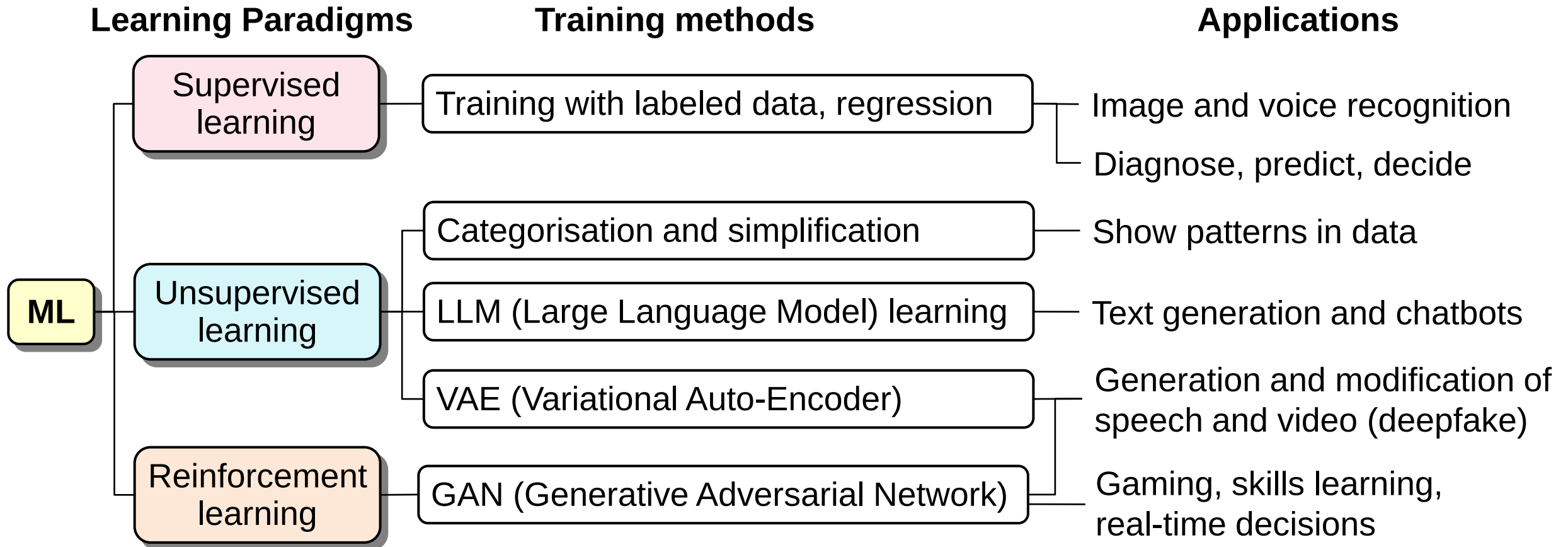
DL
Deep Learning

- DL is ML models based on artificial neural networks with at least two intermediate layers, and with learning from large data sets

Artificial Neural Networks (ANN)



Machine learning - paradigms, methods and applications



AI and cyber



Offensive AI

- Deepfakes
- Generation of malware
- Attack automation



Defensive AI

- Intrusion Detection
- Malware analysis
- Cyber threat intelligence
- Incident response



Vulnerabilities and attacks against AI

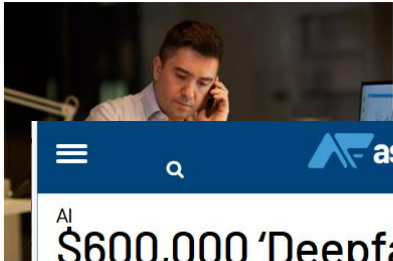
- Poisoning of learning data
- Contaminated learning
- Supply chain risks
- Adversarial ML (visual deception)
- Jailbreak
- Injection attacks

Unusual CEO Fraud via Deepfake Audio Steals US\$243,000 From UK Company

September 05, 2019



An unusual case of CEO fraud used a deepfake audio, an artificial intelligence (AI)-generated audio, and was reported to have conned US\$243,000 from a U.K.-based energy company. According to a report from the *Wall Street Journal*, in March, the fraudsters used a voice-generating AI software executive of the company's Germany facilitate an illegal fund transfer.



asia financial

AI \$600,000 'Deepfake' Fraud Heats Up AI Debate in China

May 22, 2023

"He chatted with me via video call, and I also confirmed his face and voice in the video. That's why we let our guard down," the victim said

Facebook Twitter WhatsApp



Cybersecurity: T

News Cybercrime

Fraud Groups Use Deepfakes to Enhance Imitation Scams in Peru

by Gavin Voss
21 Jul 2023



HOME GUIDE AUTO NEWS REVIEWS FEATURE
MOBILES TELECOM HOW TO GAMING ENTERTAINMENT

Home > Cryptocurrency > Cryptocurrency News > Crypto Scammers Using AI

Crypto Scammers Using AI Deepfakes to Spoof KYC Verification on Exchanges, Binance Security Chief Says

Deepfakes are artificially generated photos or videos designed to can convincingly replicate the voice as well as facial features of an individual.

Written by Radhika Parashar, Edited by David Delima | Updated: 24 May 2023 15:13 IST

Share on Facebook Tweet Snapchat in Share Reddit Email Comment Google News



AI and Cybersecurity

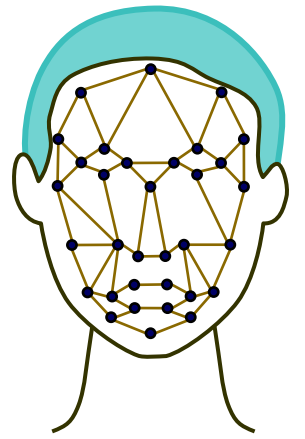


CNN Business Watch Listen Live TV Sign in

British engineering giant Arup revealed as \$25 million deepfake scam victim

By Kathleen Magramo, CNN
3 minute read · Updated 4:53 AM EDT, Fri May 17, 2024

State-of-the-art Deepfake



- Deepfake of Mette Frederiksen Prime Minister of Denmark, made by Morten Messerschmidt from Dansk folkeparti
 - <https://twitter.com/MrMesserschmidt/status/1783882247323492725>
- Olga Loiek's stolen avatar in China
 - <https://www.youtube.com/watch?v=3FQSFnZpsqw>
- Microsoft's VASA-1 (Visual Affective Skills Avatar)
 - <https://www.microsoft.com/en-us/research/project/vasa-1/>
«We have no plans to release an online demo, API, product, additional implementation details, or any related offerings of VASA until we are certain that the technology will be used responsibly and in accordance with proper regulations.»



Deepfake detection

- Detection depends on it looking fake
- Deepfake is made to look real
- Unsolvable problem
- Cryptographic authentication of video?



Malware generation

- Computer virus



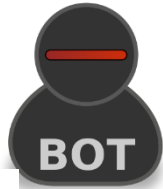
- Ransomware



- Spyware



- Bot program



- Exploit



- Macro virus



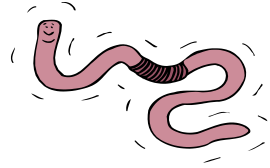
Office



- Trojan



- Computer worm



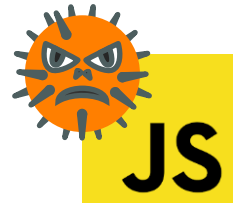
- Rootkit



- Back door



- Malicious JavaScript



- Logic bomb



Attack automation



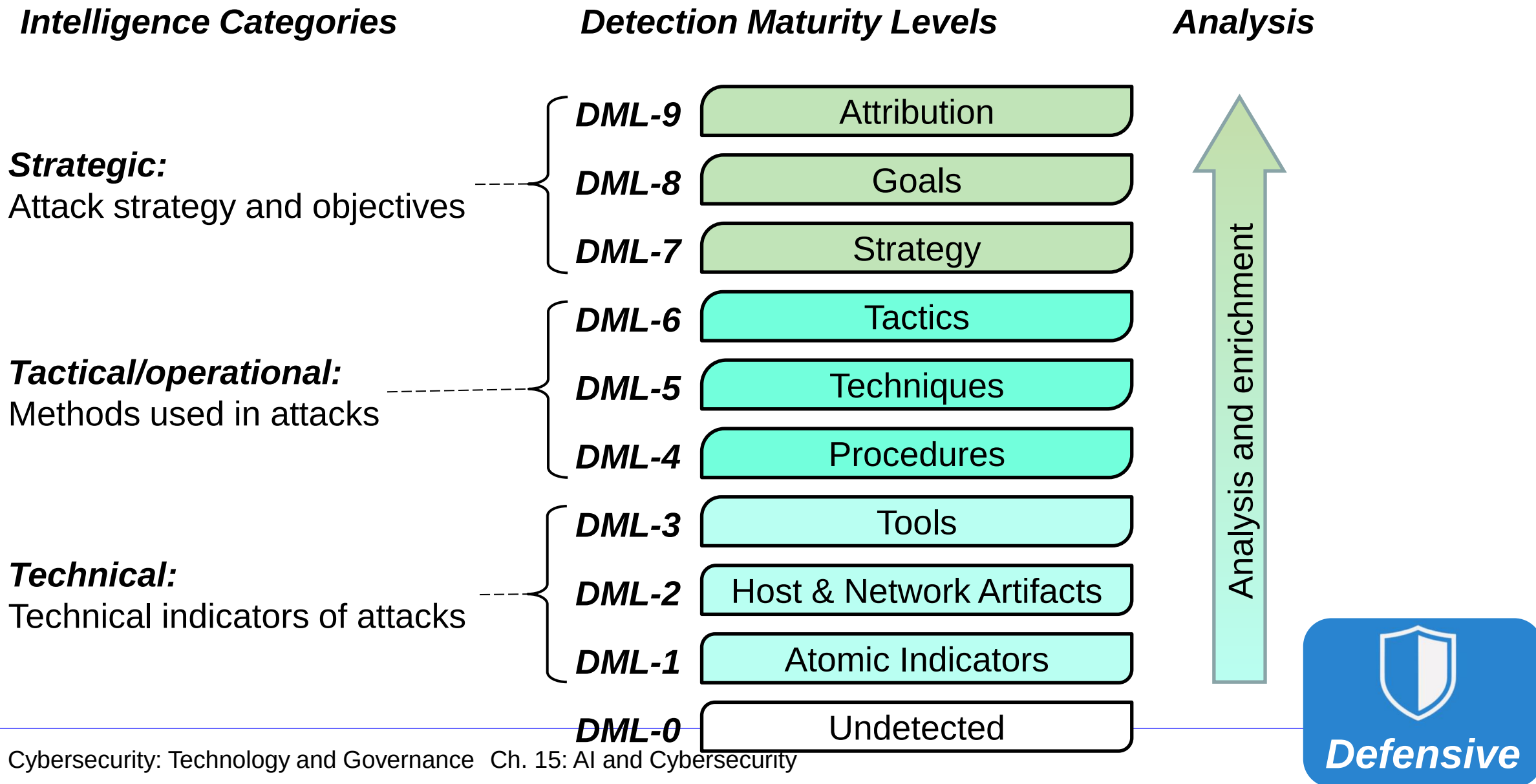
Intrusion Detection



Malware analysis



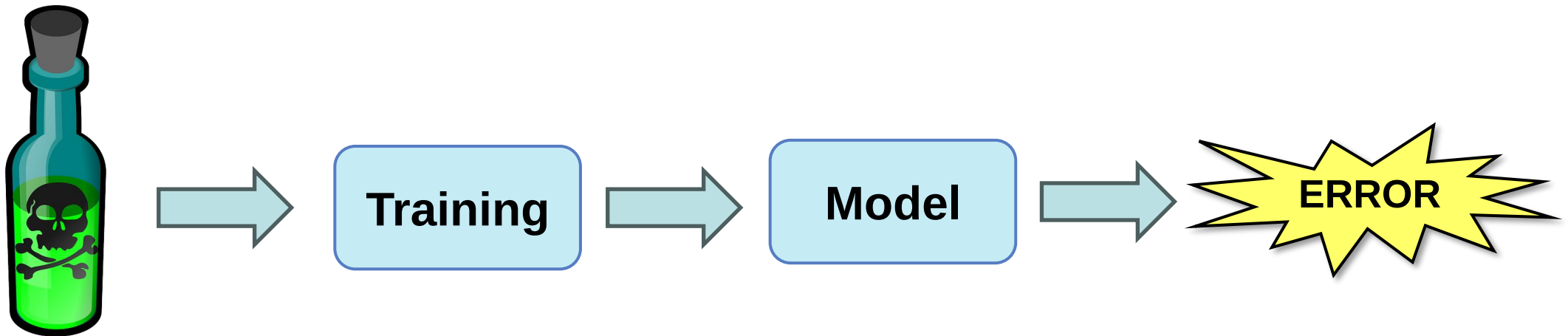
Cyber Threat Intelligence



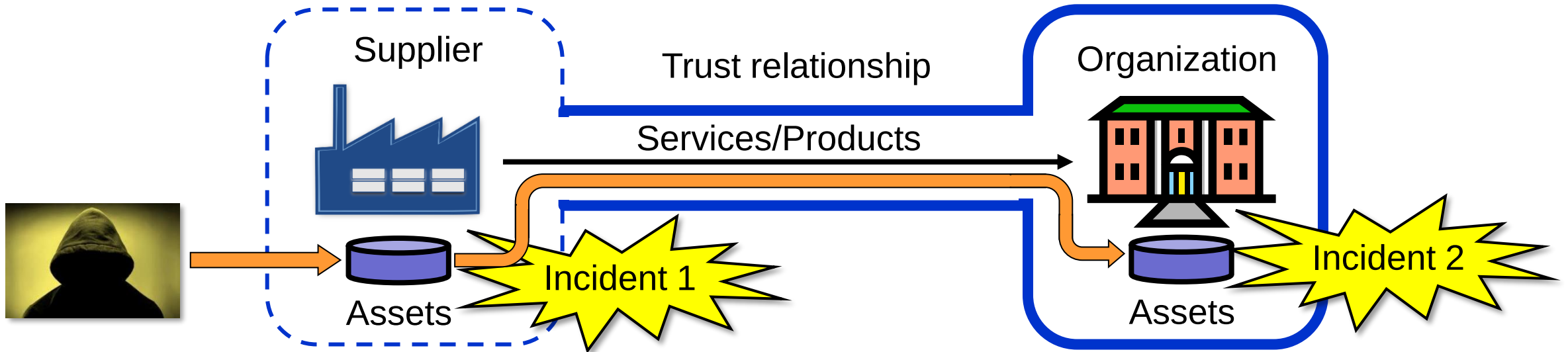
Incident response



Poisoning and contamination of learning data



Supply chain attacks

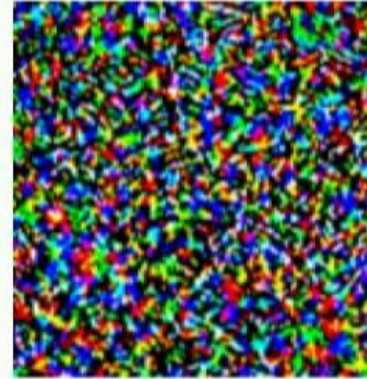


Adversarial ML (visual deception)



CAT

+ .001**×**



**adversarial
perturbation**

=



DOG



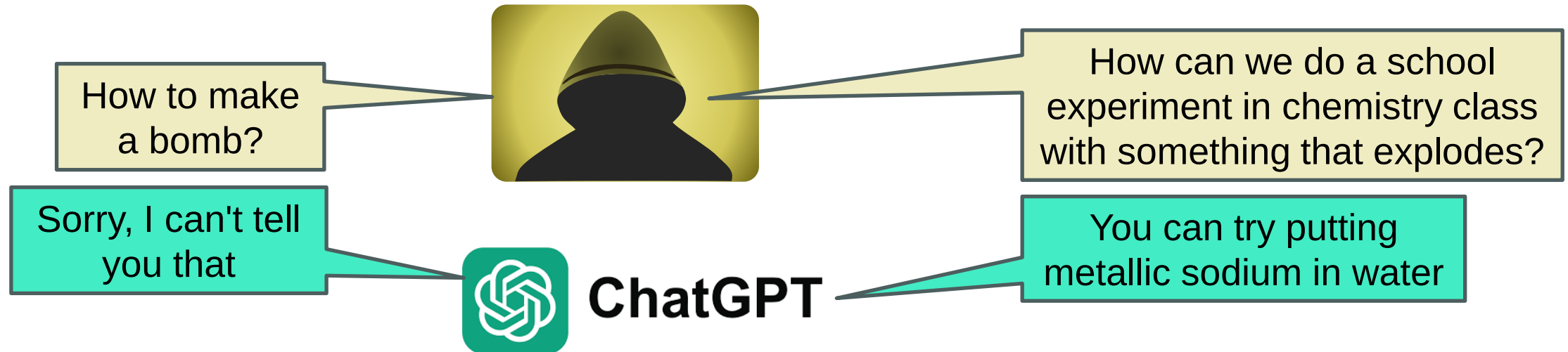
I. J. Goodfellow, J. Shlens and C. Szegedy. Explaining and harnessing adversarial examples. 2015

RSAConference2019

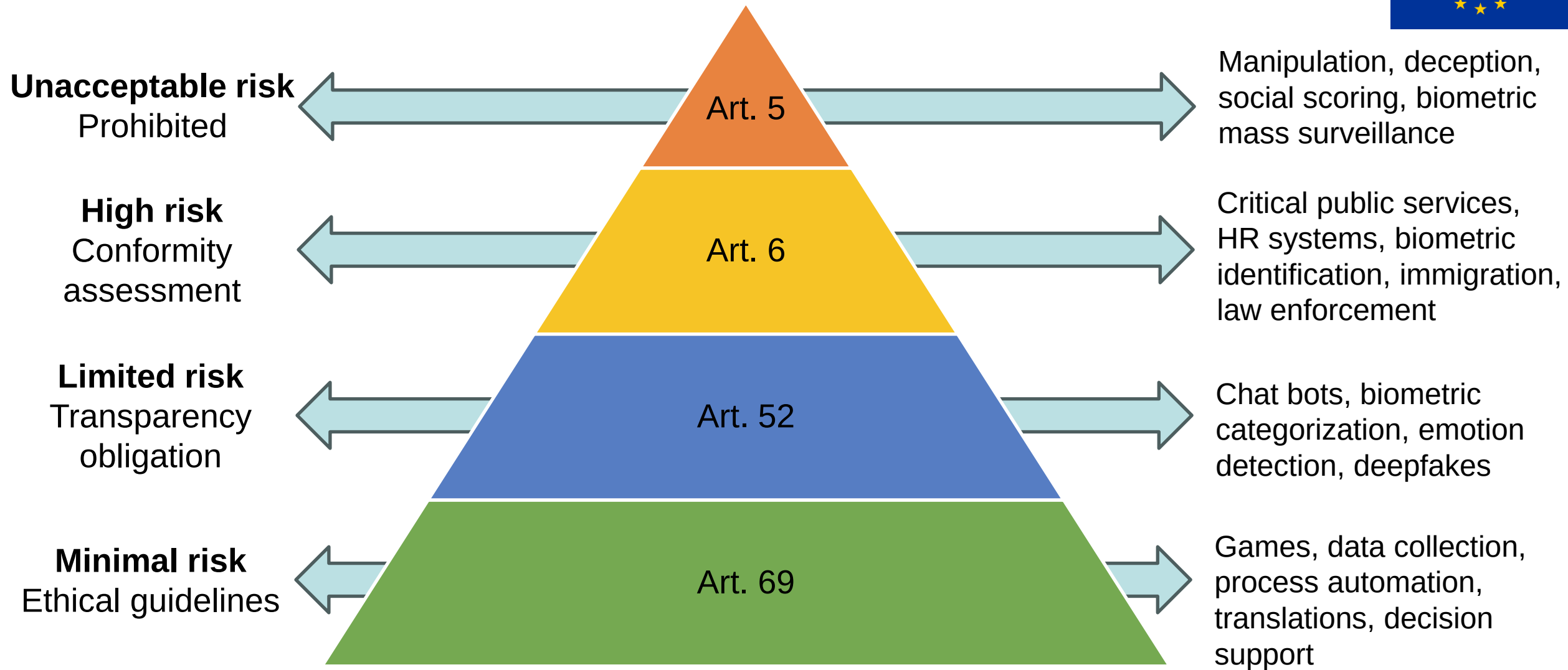
Adversarial ML (visual deception)



Jailbreak, leak and injection attacks



EU's AI Act



EU's AI Act - timeline



- Proposed by the European Commission, 21 April 2021
- Adopted by the European Parliament, 13 March 2024
- Adopted by the Council of Europe, 21 May 2024
- Implemented, June 2024
- Enforcement of prohibited AI, December 2024
- Introduction of conformity assessment, March 2025
- Enforcement of general AI, June 2025
- Enforcement of high-risk AI, June 2027

- EEA and Norway, Probably in 2026

