Data, algorithms and AI in healthcare and medicine: Reflection on cybersecurity and cyber resilience

7th eHEALTH SECURITY CONFERENCE

Copenhagen, Denmark

10th October 2022

Vittorio Reina
vittorio.reina@ec.europa.eu

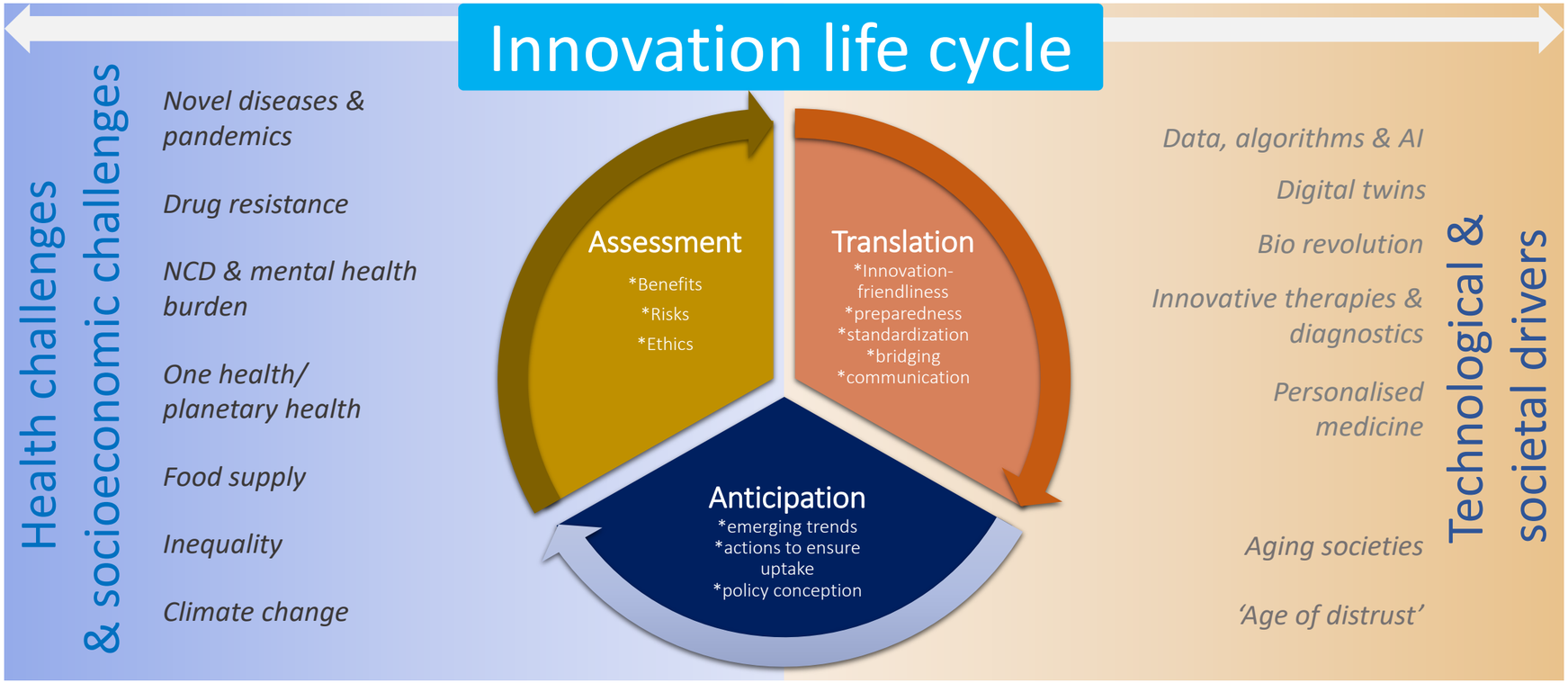
Claudius Griesinger
claudius.griesinger@ec.europa.eu

EU Commission – Joint Research Centre

AI, cyber resilience and health at the JRC

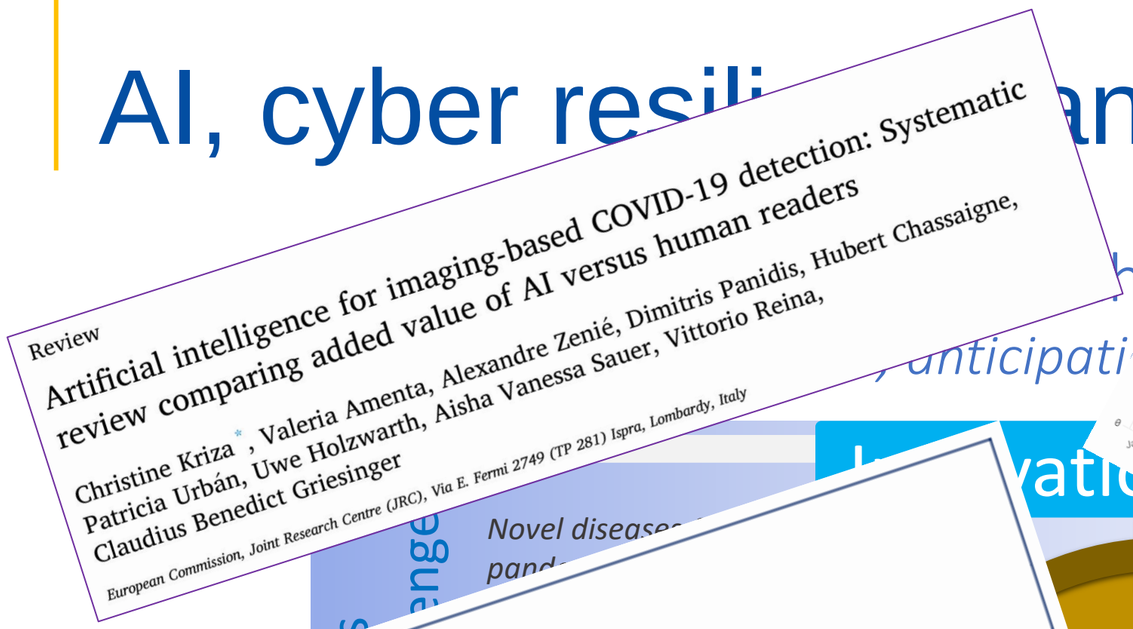
Innovation in life and health sciences:

assessment, translation, anticipation



European Commission

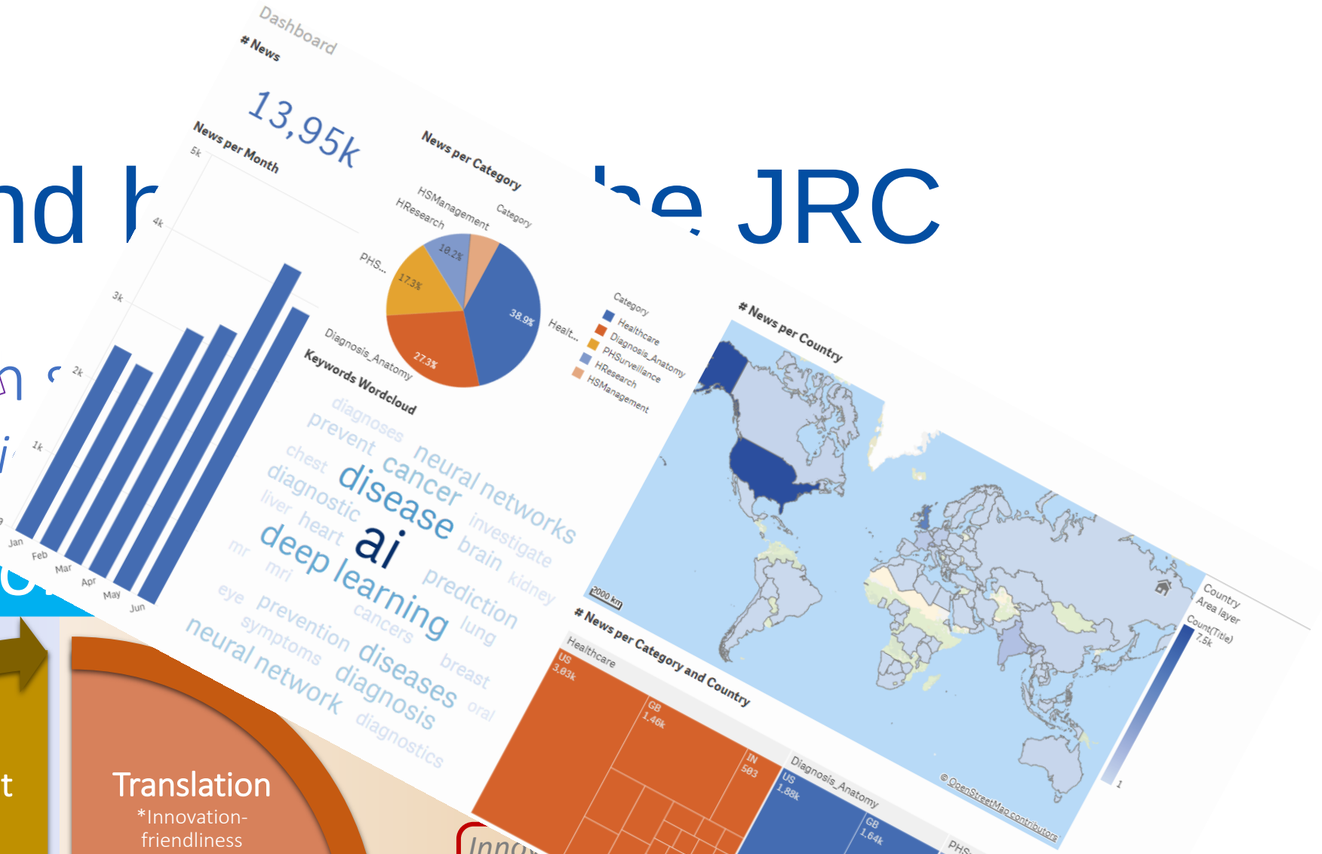
AI, cyber resili… …and h… …he JRC


Review

Artificial intelligence for imaging-based COVID-19 detection: Systematic review comparing added value of AI versus human readers

Christine Kriza *, Valeria Amenta, Alexandre Zenié, Dimitris Panidis, Hubert Chassaigne, Patricia Urbán, Uwe Holzwarth, Aisha Vanessa Sauer, Vittorio Reina, Claudius Benedict Griesinger

European Commission, Joint Research Centre (JRC), Via E. Fermi 2749 (TP 281) Ispra, Lombardy, Italy



MDCG 2019-16
Guidance on Cybersecurity for medical devices

December 2019

European Commission

IT and data create vulnerabilities

Highest likelihood risks ***of the next ten years*** *are extreme weather, climate action failure and human-led environmental damage; as well as digital power concentration, digital inequality and*
cybersecurity failure*.*

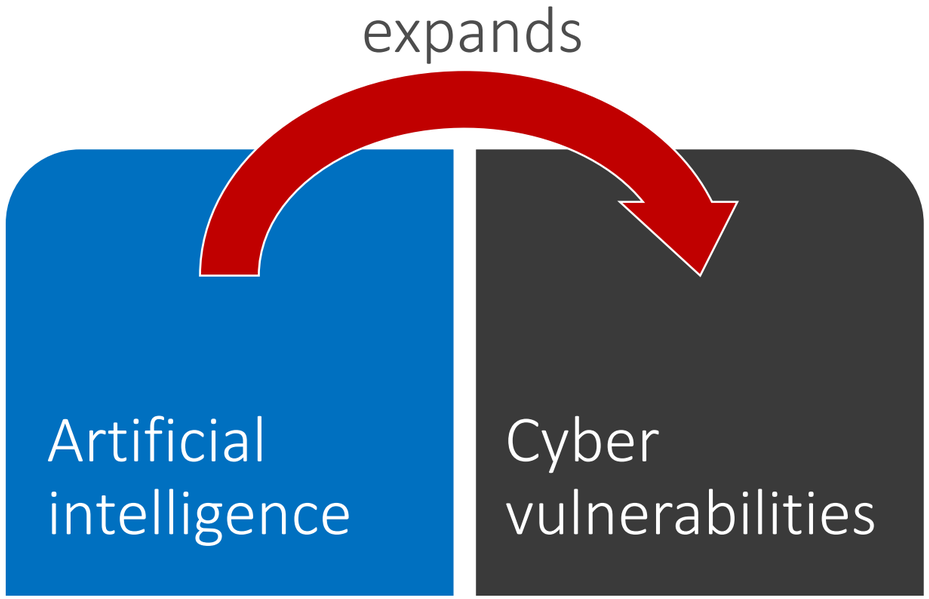
Highest impact risks ***of the next decade****, infectious diseases are in the top spot, followed by climate action failure and other environmental risks; as well as weapons of mass destruction, livelihood crises, debt crises and*
IT infrastructure breakdown*.*

(World Economic Forum's Global Risks Report 2021)

Global Risks Perception Survey (GRPS) 2021
https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2021.pdf

European Commission

Artificial Intelligence's Janus face



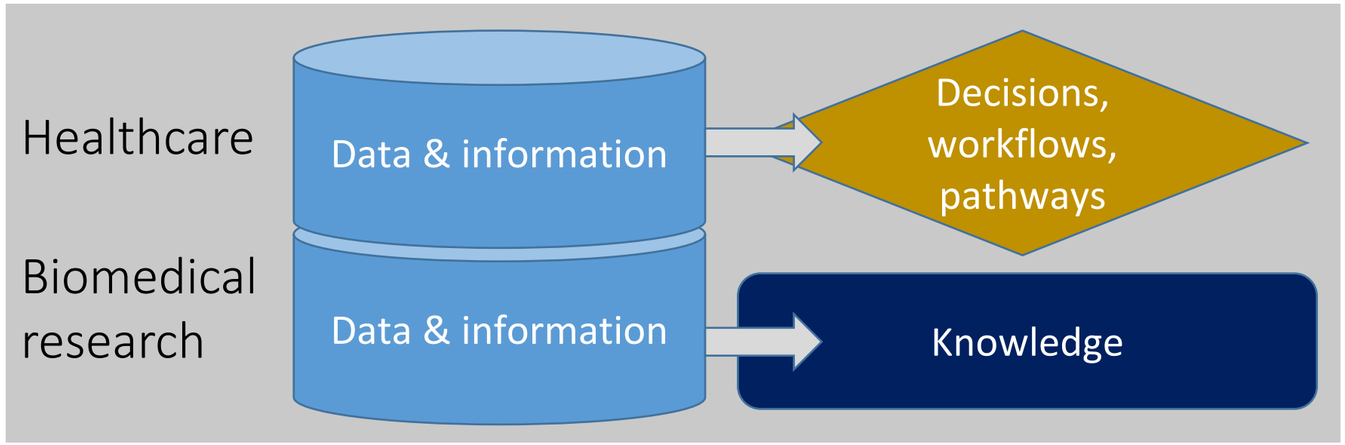
Healthcare & medicine

AI for critical functions and services

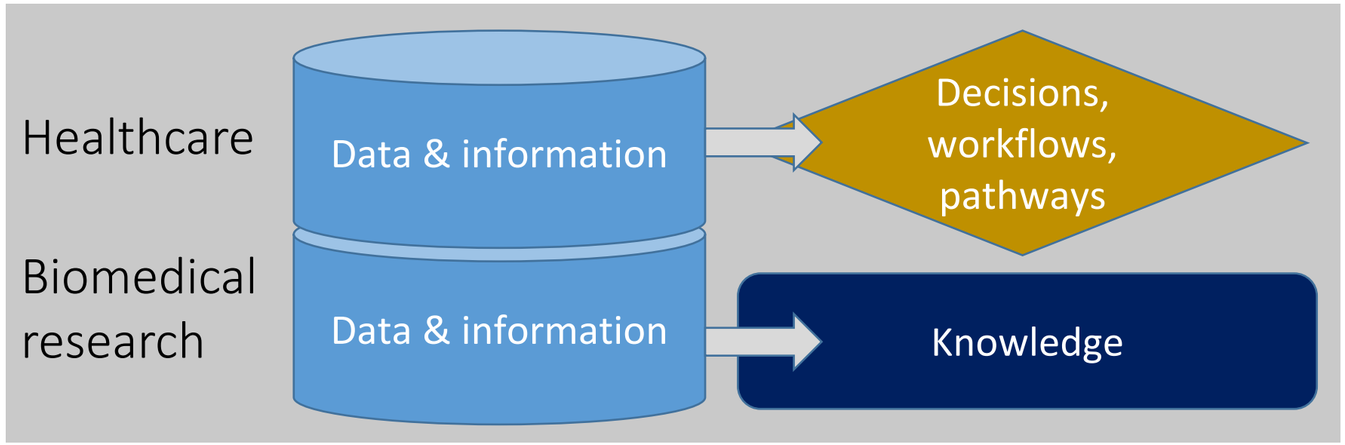
“Increasing dependence on AI for critical functions and services will not only create greater incentives for attackers to target those algorithms, but also the potential for each successful attack to have more severe consequences.”

https://www.brookings.edu/research/how-to-improve-cybersecurity-for-artificial-intelligence/

AI in medicine and healthcare: many diverse applications



AI in medicine and healthcare: many diverse applications



1) Healthcare

- Diagnosis & prediction-based diagnosis
- Clinical care & disease management pathways
- Active implantable devices, wearables etc.
- Robotic surgery

2) Health systems management

- Administrative workflow
- Logistics, procurement
- Chatbots & virtual nursing assistants
- emedicine: care at home

Data

3) Public health & surveillan

- Disease outbreaks monitoring
- Pandemic preparedness
- Health promotion & disease prevention

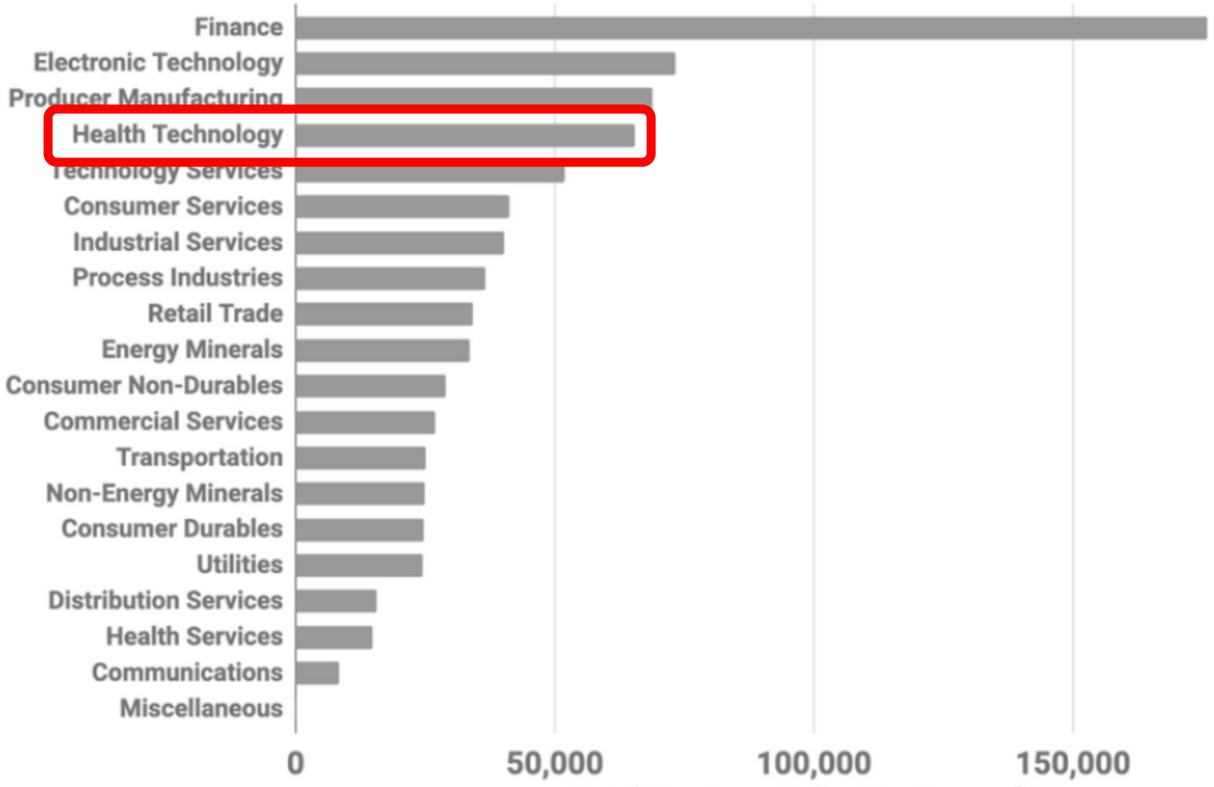
lth research

- Health data for research & development (including AI)
- Electronic health records: optimisation of clinical care
- Drug / Vaccine development & repurposing
- Genomic medicine & personalised medicine

Ethics and governance of artificial intelligence for health: WHO guidance (2021)

AI in medicine and healthcare: many diverse applications

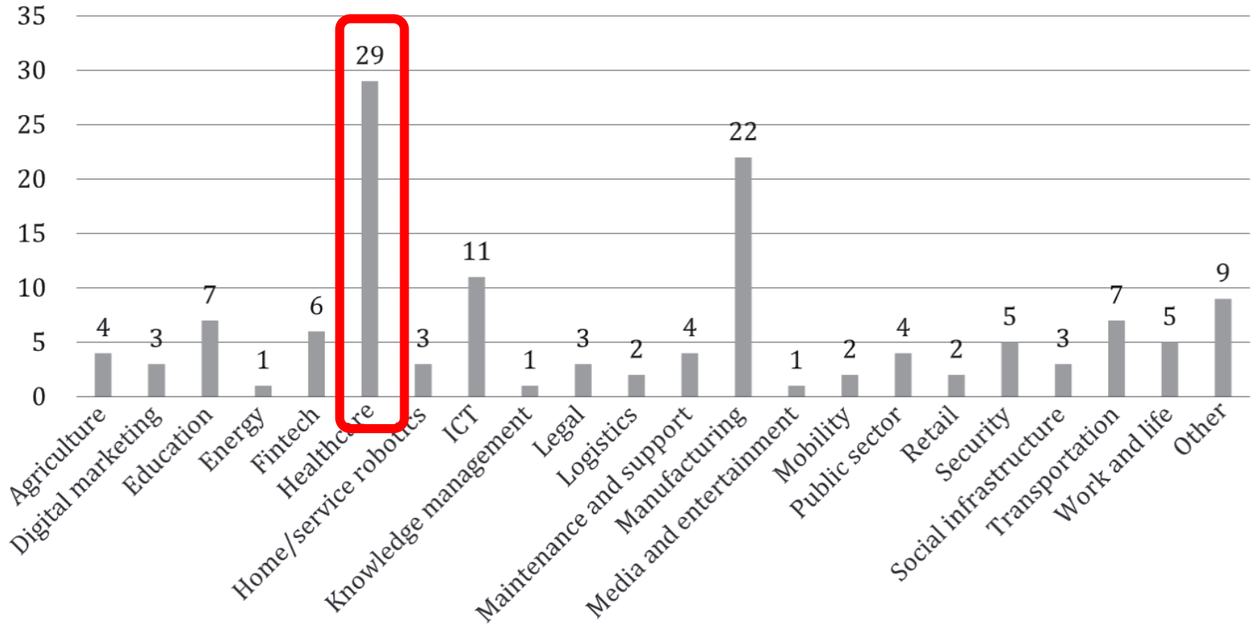
OECD Framework for the classification of AI systems OECD (2022)



Mentions of AI in earning calls by sector 208-2019

Stanford AI Index 2021, https://aiindex.stanford.edu/report/.

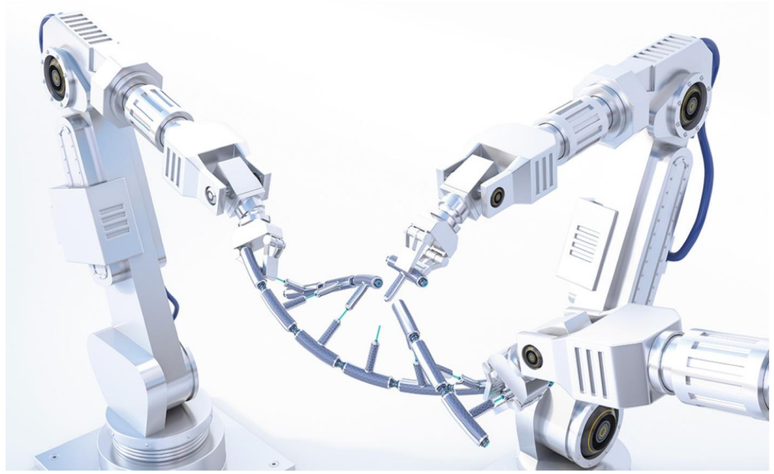
Information technology - Artificial intelligence (AI) - Use cases ISO/IEC TR24030 (2021)



Distribution of use cases by application domain

European Commission

AI in medicine and healthcare: many diverse applications (1/2)



Real examples of AI tasks and methods in the health sector:

- **COVID-19 diagnosis**
- Chatbots
- Identifying risk factors in health
- Heart disease diagnosis
- Breast cancer management
- Cervical cancer diagnostics
- Human fall detection

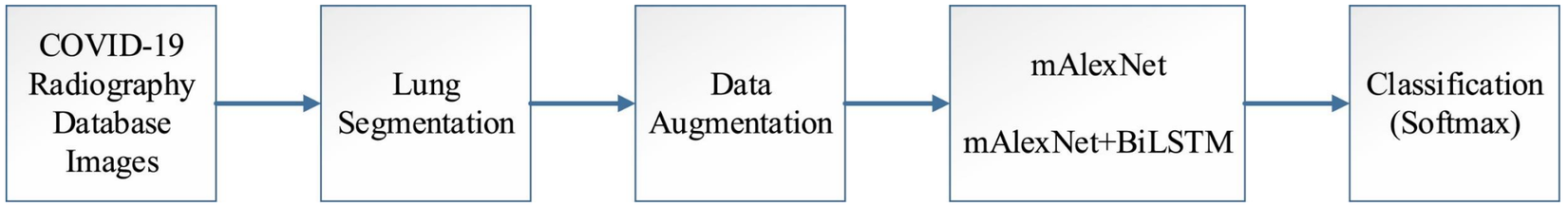
AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems - Iqbal H. Sarker (2022)
https://doi.org/10.1007/s42979-022-01043-x

European Commission

AI in medicine and healthcare: many diverse applications (2/2)

Convolutional Neural Networks (CNN) based approach for COVID-19 infection detection

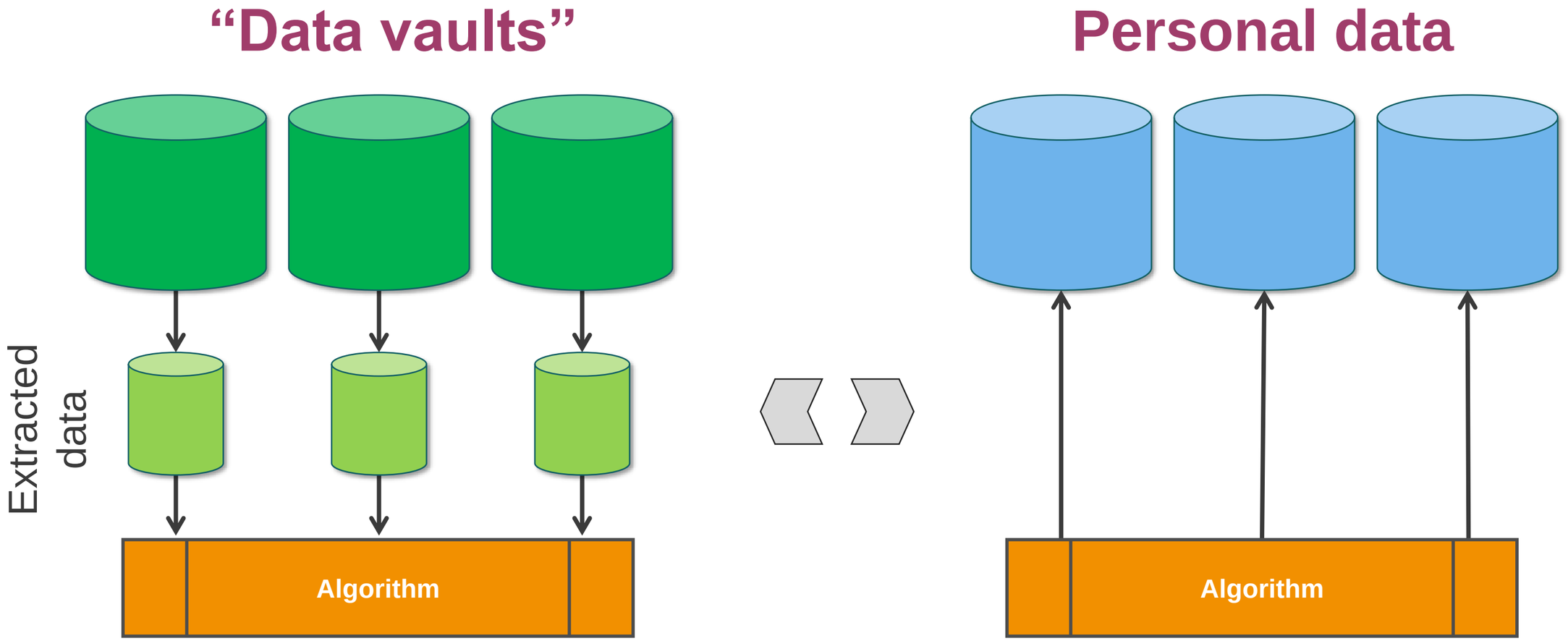
1. Open-access database covering the posterior-to-anterior chest X-ray images
2. Noises or irrelevant patterns are auomaticaly removed from raw X-ray images
3. Data augmented in computer environment to increase the classification success
4. Chest X-ray images are classified using a transfer learning-based modified architecture (mAlexNet)
5. Classification is completed using Softmax



CNN-based transfer learning–BiLSTM network: A novel approach for COVID-19 infection detection - Muhammet Fatih Aslan, Muhammed Fahri Unlersen, Kadir Sabanci and Akif Durduc (2021)
DOI: 10.1016/j.asoc.2020.106912

European Commission

Data or algorithm flow for AI systems in healthcare & medicine

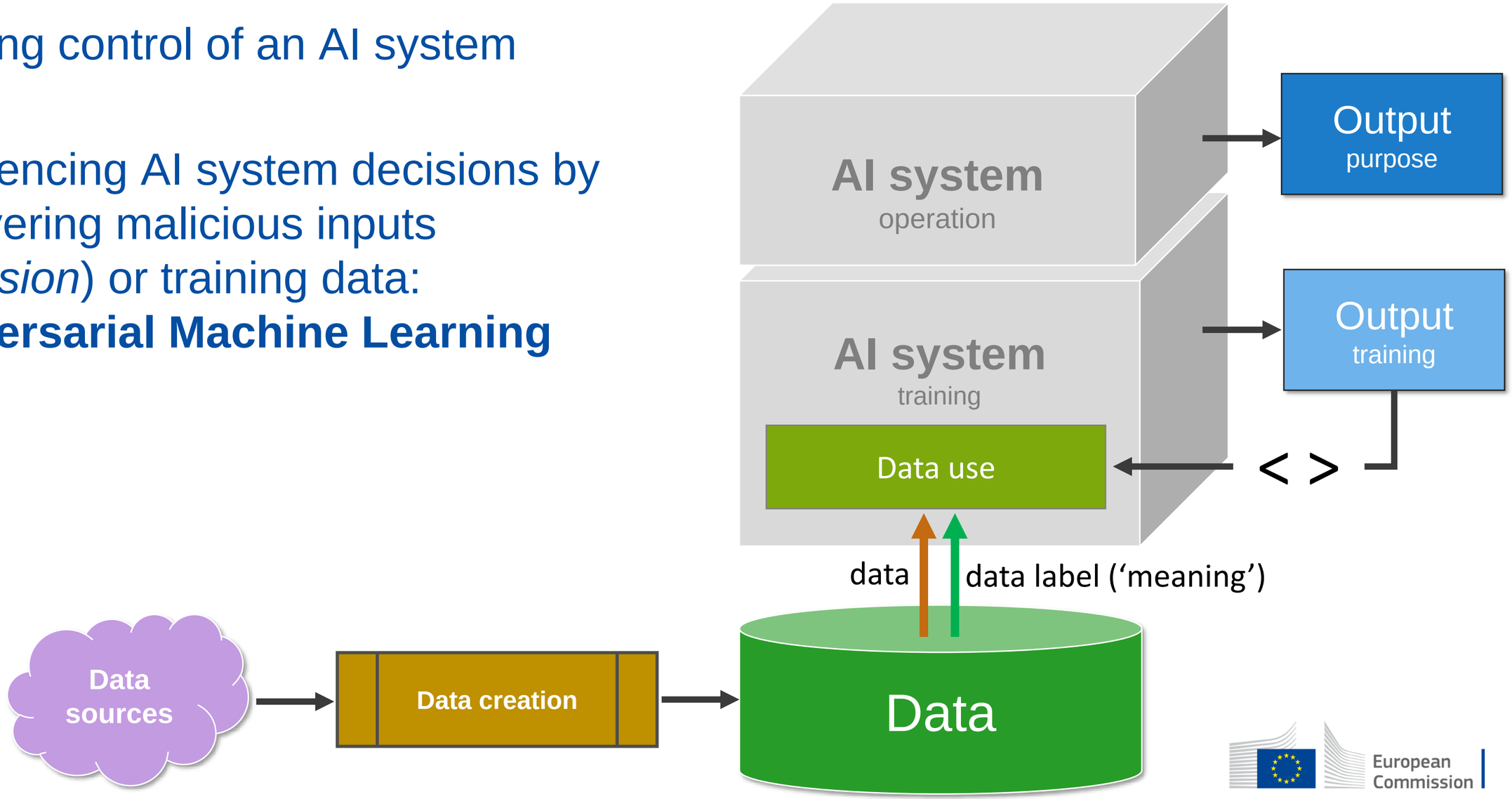


European Commission

Compromising model integrity

1. Taking control of an AI system

2. Influencing AI system decisions by delivering malicious inputs (*evasion*) or training data: **Adversarial Machine Learning**

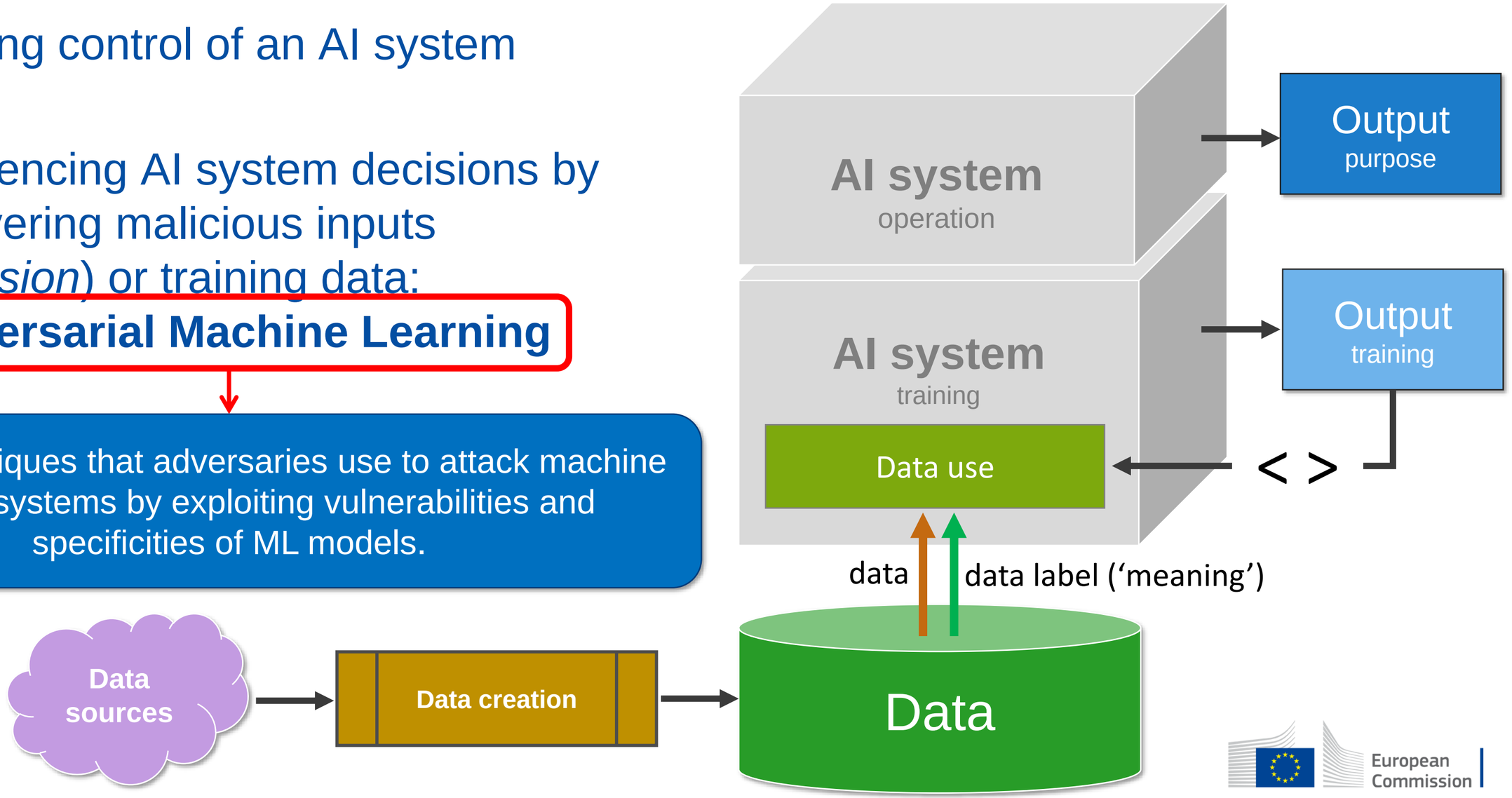


Compromising model integrity

1. Taking control of an AI system

2. Influencing AI system decisions by delivering malicious inputs (*evasion*) or training data:
   Adversarial Machine Learning

A set of techniques that adversaries use to attack machine learning systems by exploiting vulnerabilities and specificities of ML models.



European Commission

Adversarial Machine Learning: types of attacks

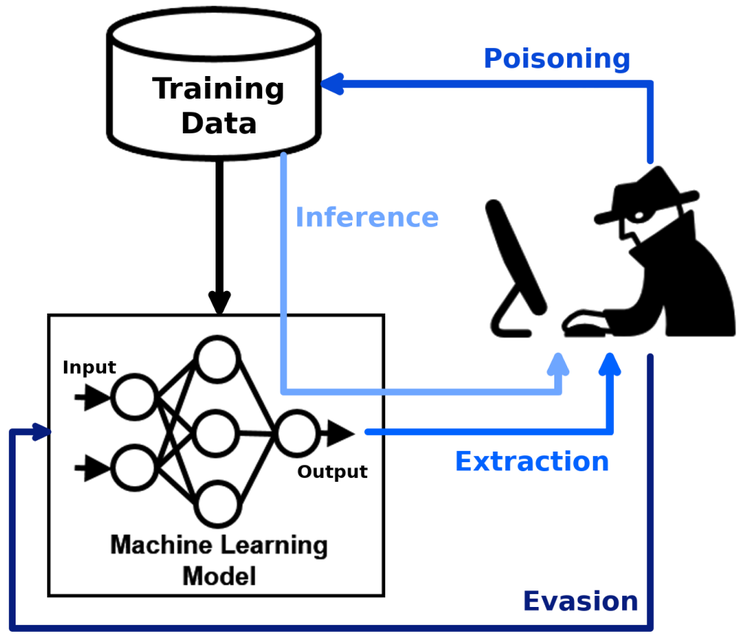
Poisoning attacks

Evasion attacks

Adversarial machine learning

Extraction attacks

Inference attacks



https://adversarial-robustness-toolbox.readthedocs.io/en/latest/

https://www.csoonline.com/article/3664748/adversarial-machine-learning-explained-how-attackers-disrupt-ai-and-ml-systems.html

European Commission

Adversarial Machine Learning: types of attacks

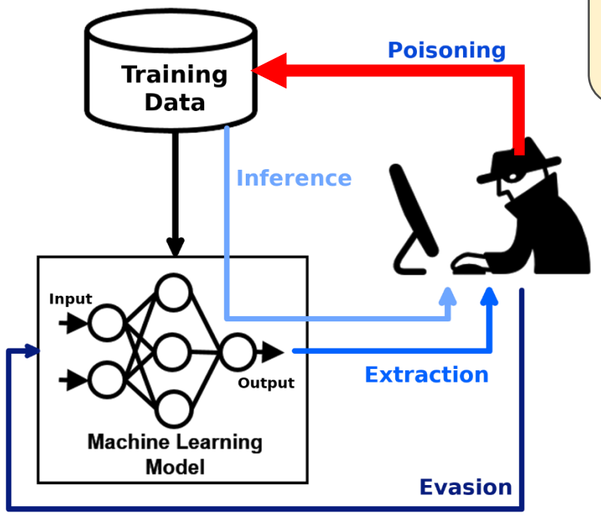
Poisoning attacks

Extraction attacks

Evasion attacks

ference attacks

Contaminating the training dataset inserting corrupt data to compromise a target machine learning model during training.



European Commission

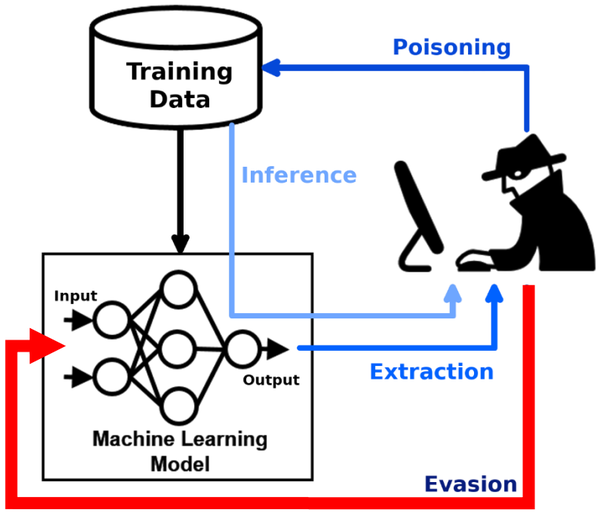
Adversarial Machine Learning: types of attacks

Poisoning attacks

Extraction attacks

Evasion attacks

Inference attacks



Adversaries insert a small perturbation (in the form of noise) into the input of a machine learning model to make it classify incorrectly.

European Commission

Adversarial Machine Learning: types of attacks

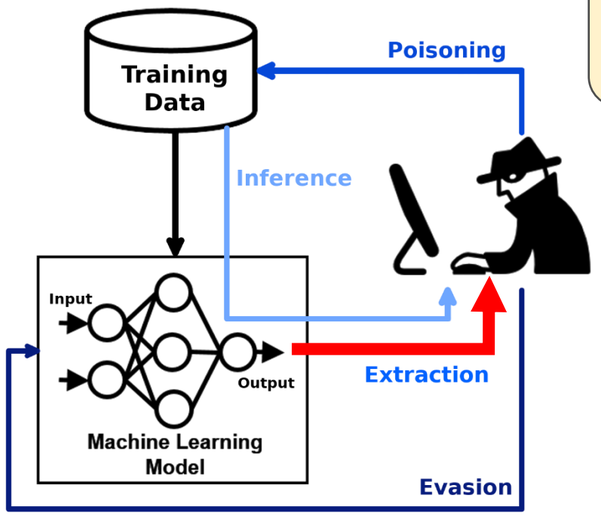
Poisoning attacks

Extraction attacks

Probing a blacl-box machine learning system in order to either reconstruct the model or extract the data it was trained on (e.g. query a model in a mathematically guided fashion)

Evasion attacks

ference attacks



European Commission

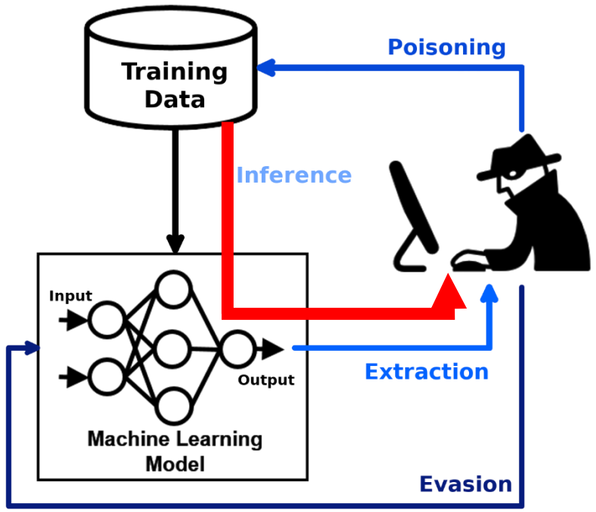
Adversarial Machine Learning: types of attacks

Poisoning attacks

Extraction attacks

Evasion attacks

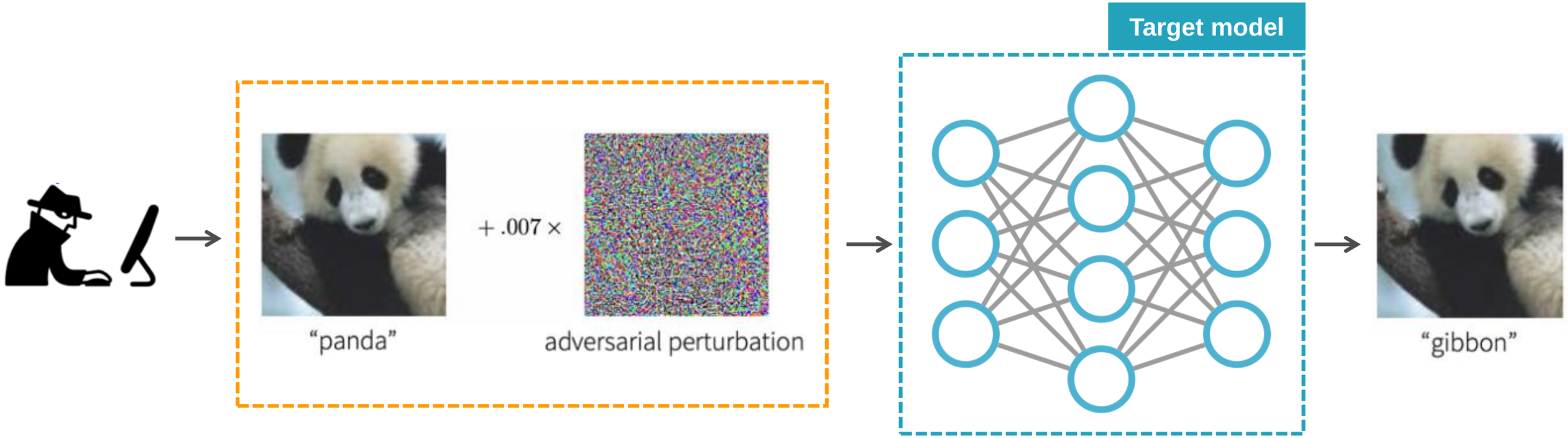
Inference attacks



Inferring whether a data record was used to train a target model by probing a machine learning model with different input data and weighing the output.

Adversarial Machine Learning: types of attacks

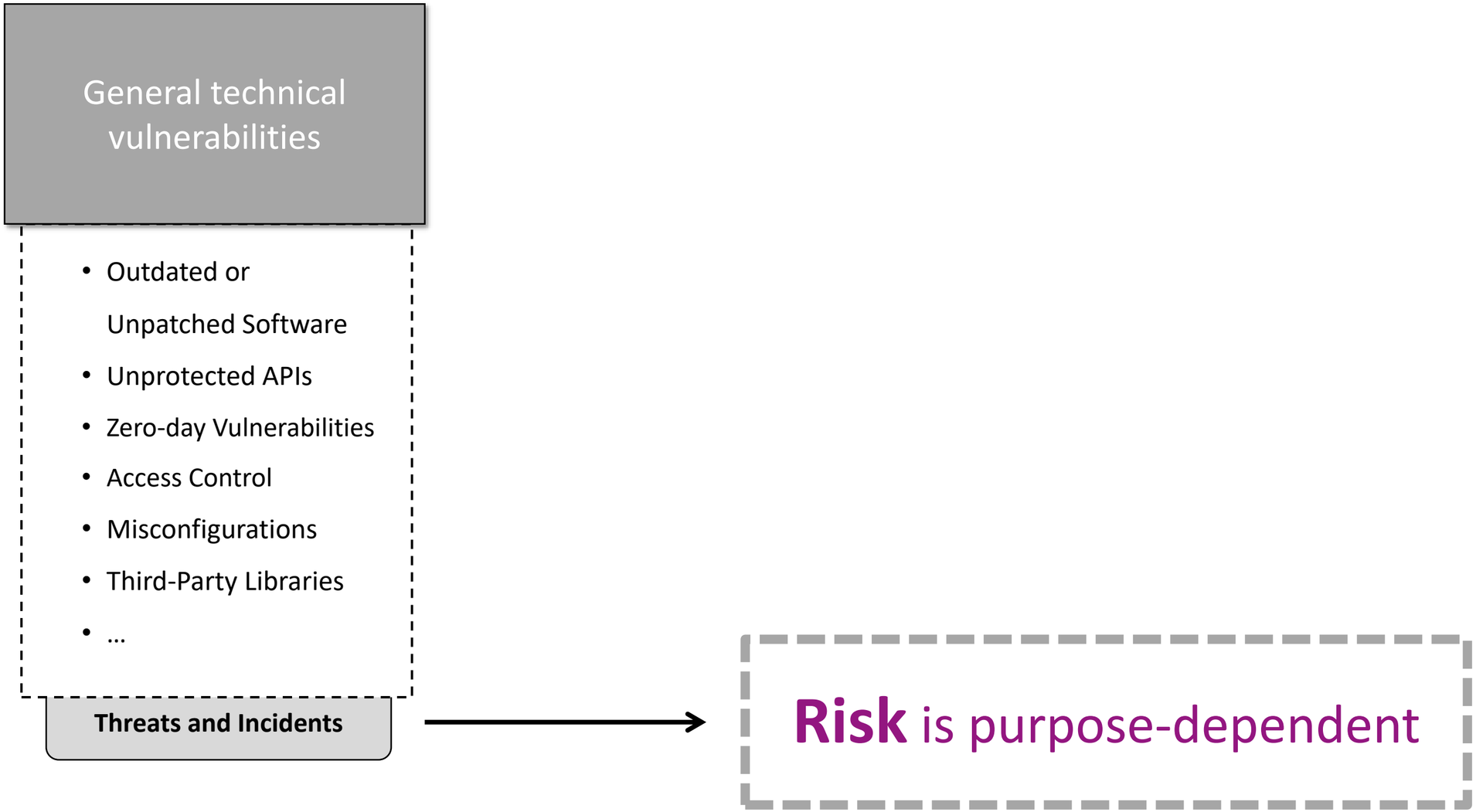
Example of an evasion attack: an image is manipulated to fool a neural network and lead to unexpected erroneous behavior on seemingly benign inputs

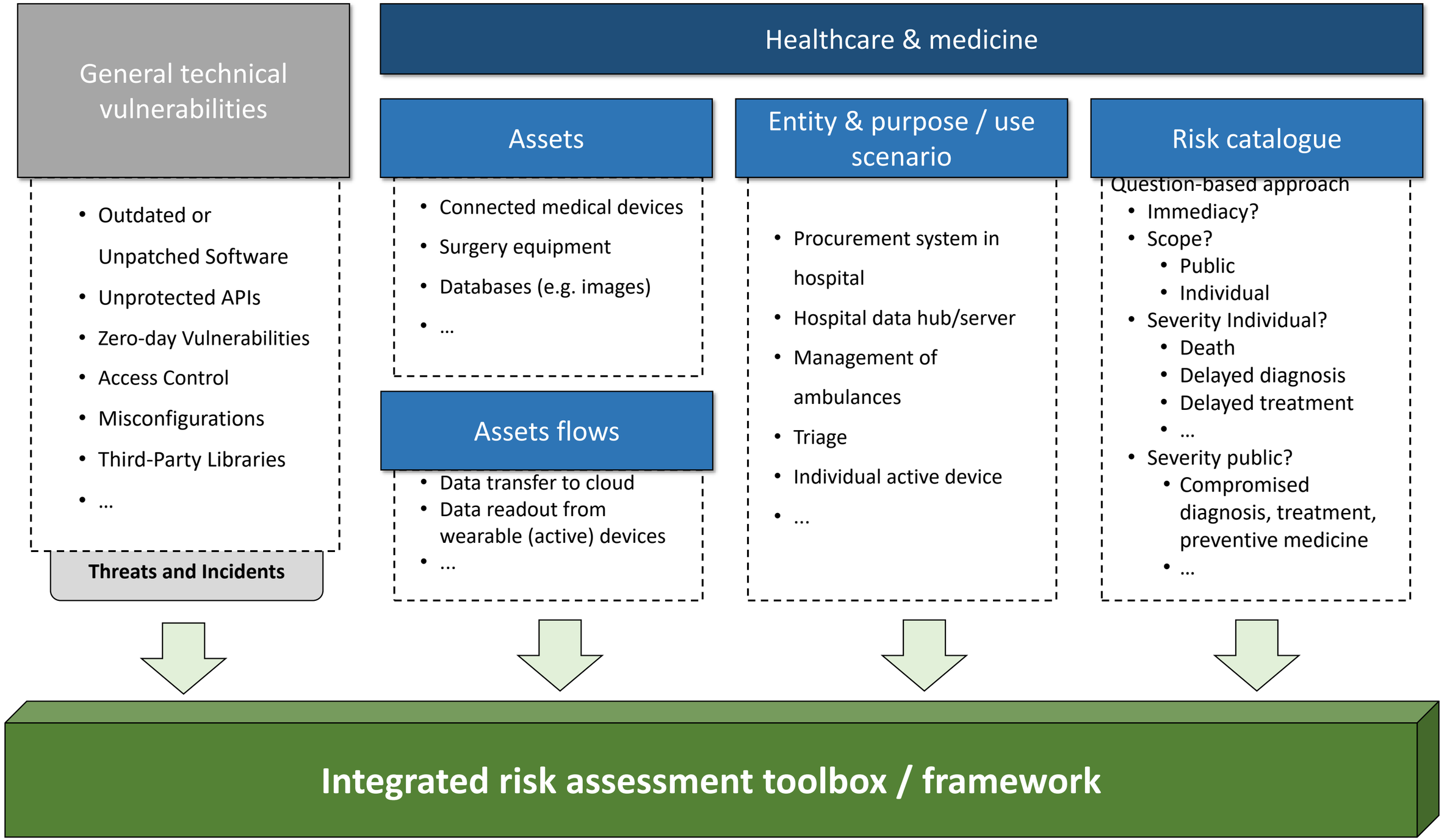


Adapted from “Explaining and harnessing adversarial examples” - Goodfellow et al (2015)

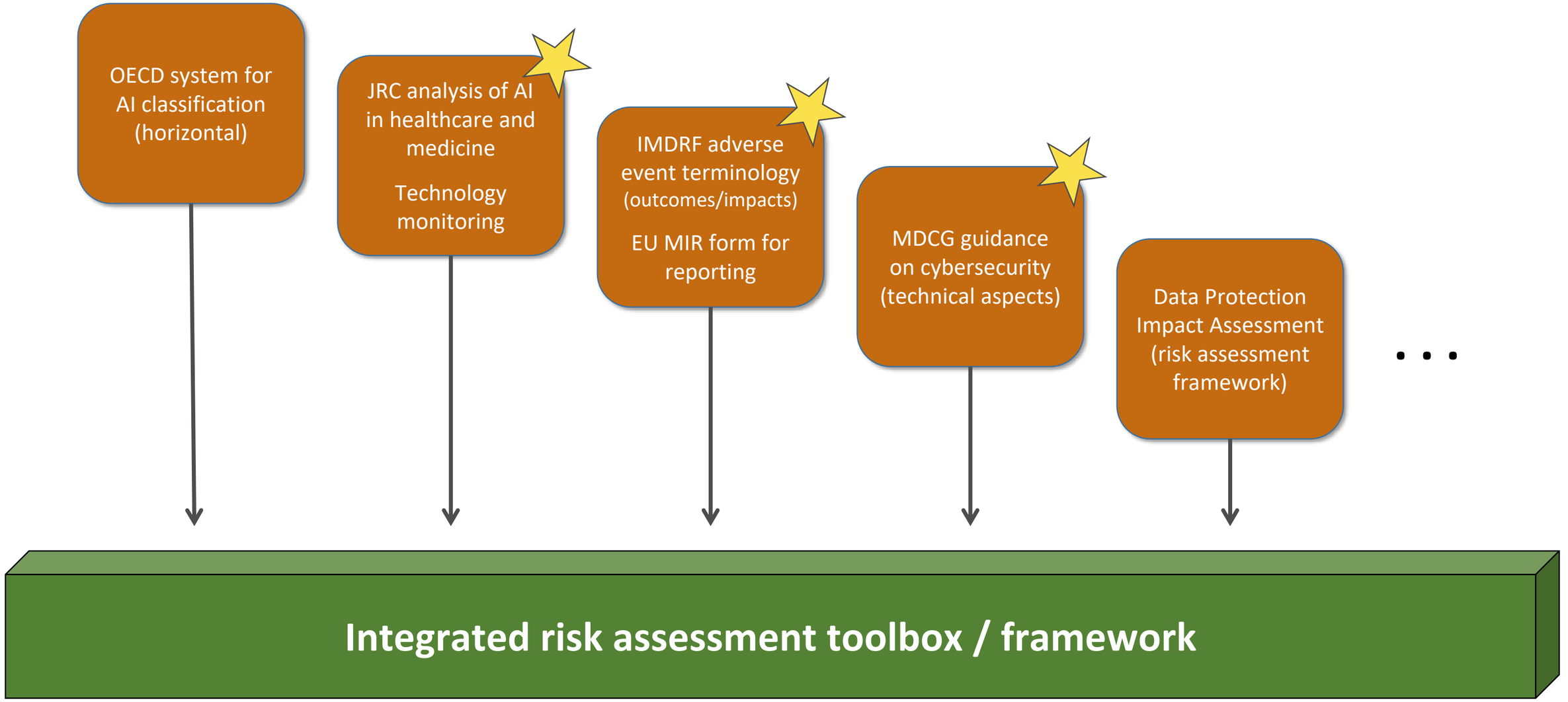
Dissecting the cloudy sky of cybersecurity and health

European Commission


General technical vulnerabilities
• Outdated or Unpatched Software
• Unprotected APIs
• Zero-day Vulnerabilities
• Access Control
• Misconfigurations
• Third-Party Libraries
• …
Threats and Incidents
Risk is purpose-dependent


General technical vulnerabilities
Outdated or Unpatched Software
Unprotected APIs
Zero-day Vulnerabilities
Access Control
Misconfigurations
Third-Party Libraries
…
Threats and Incidents
Healthcare & medicine
Assets
Connected medical devices
Surgery equipment
Databases (e.g. images)
…
Assets flows
Data transfer to cloud
Data readout from wearable (active) devices
…
Entity & purpose / use scenario
Procurement system in hospital
Hospital data hub/server
Management of ambulances
Triage
Individual active device
…
Risk catalogue
Question-based approach
Immediacy?
Scope?
Public
Individual
Severity Individual?
Death
Delayed diagnosis
Delayed treatment
…
Severity public?
Compromised diagnosis, treatment, preventive medicine
…
Integrated risk assessment toolbox / framework

Elements that may help establishing the toolbox



Toolbox may help dealing with…

…prevention

- Minimize the access bad actors have to training data within confidential computing
- Control over the training datasets that are used to build AI models
- Perturbation based defense mechanisms: input perturbation-base and output perturbation-based approaches
- Dynamic testing
- Red teaming
- Security Development Lifecycle

Toolbox may help dealing with…

…detection

- Automated defenses (AI)
- Dynamic analysis
- Human defenders, human threat hunters.
- Security teams to stay alert for suspicious activity or unanticipated machine learning behaviors which can help identify attacks like these
- Control over the data inputs

European Commission

Thank you

© European Union 2020
Unless otherwise noted the reuse of this presentation is authorised under the CC 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.
Photos, source: Freepik, Unsplash, Derek Leung – Icons, source: Flaticon