

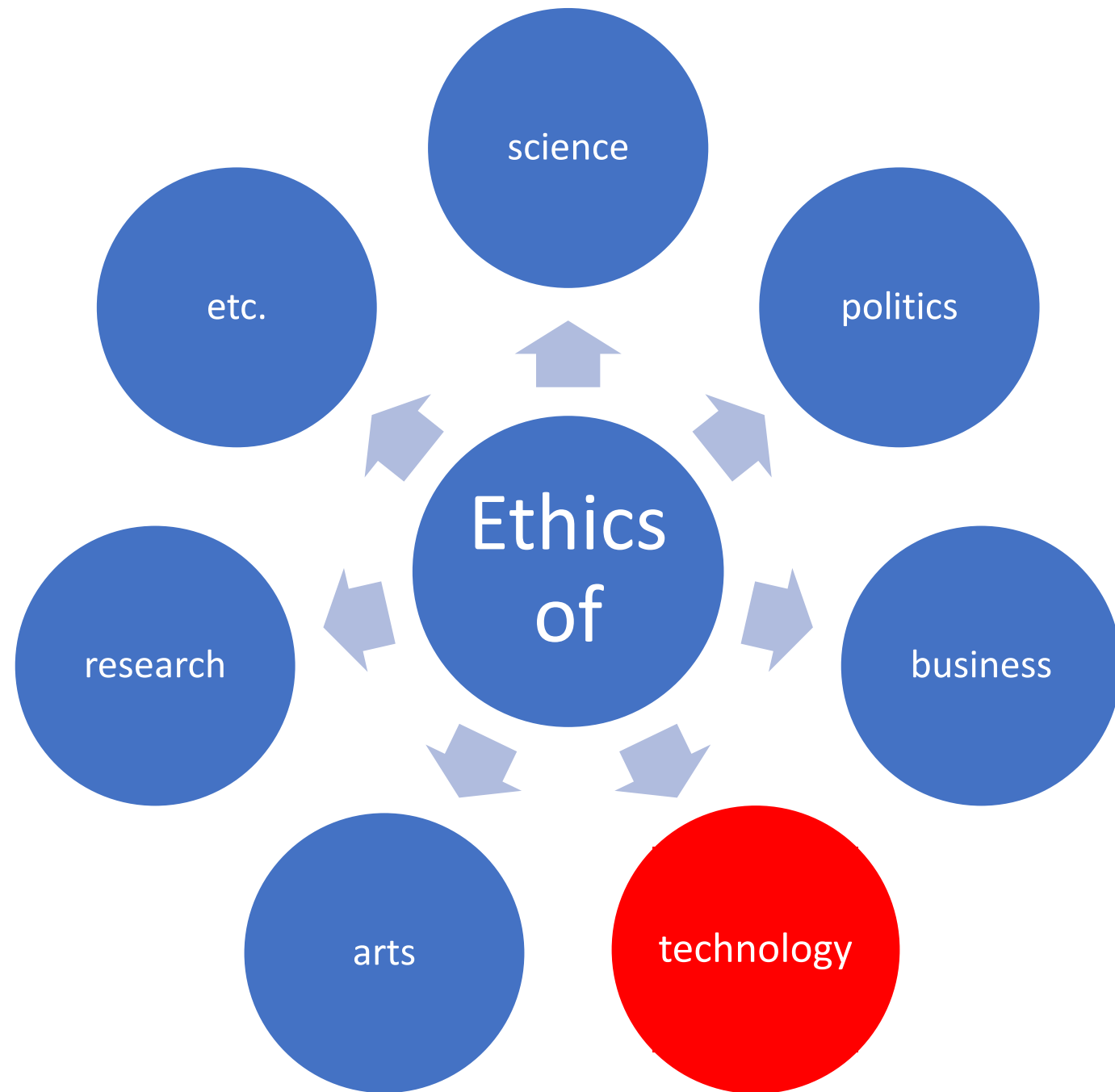
Main Ethical Problems of AI

Migle Laukyte

Tenure Track Professor of Cyberlaw and Cyber Rights,
Law Department, UPF

What ethics is
all about?

About what is **right**
and what is **wrong**



Ethics of technology

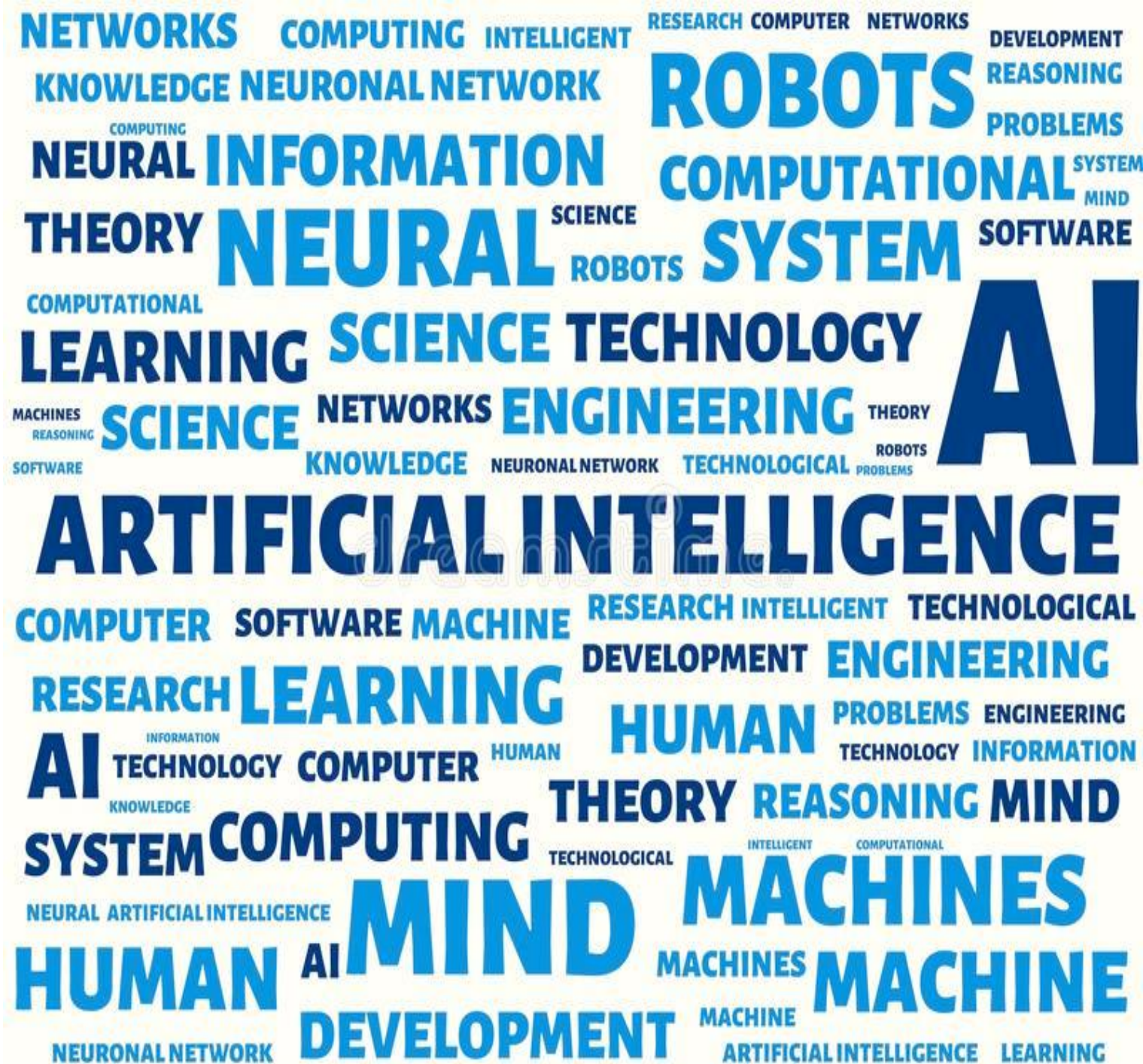
- Machine ethics
- Ethics of nanotechnology
- Data ethics
- Information ethics
- Ethics of biotechnology
- Engineering ethics
- Internet ethics
- **Ethics of Artificial Intelligence**
- ...

Ethics of AI: set of principles on what is right and what is wrong **in or with (in relation to)** AI (its development and use)



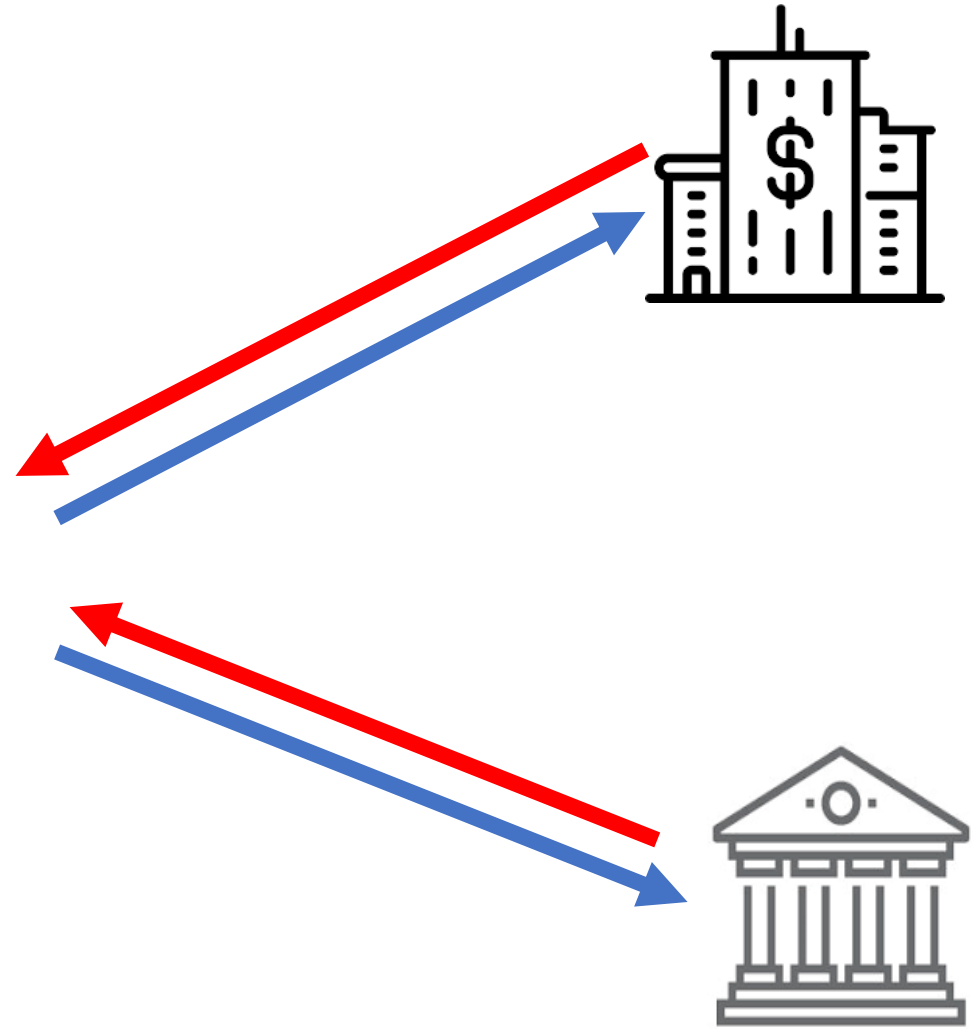
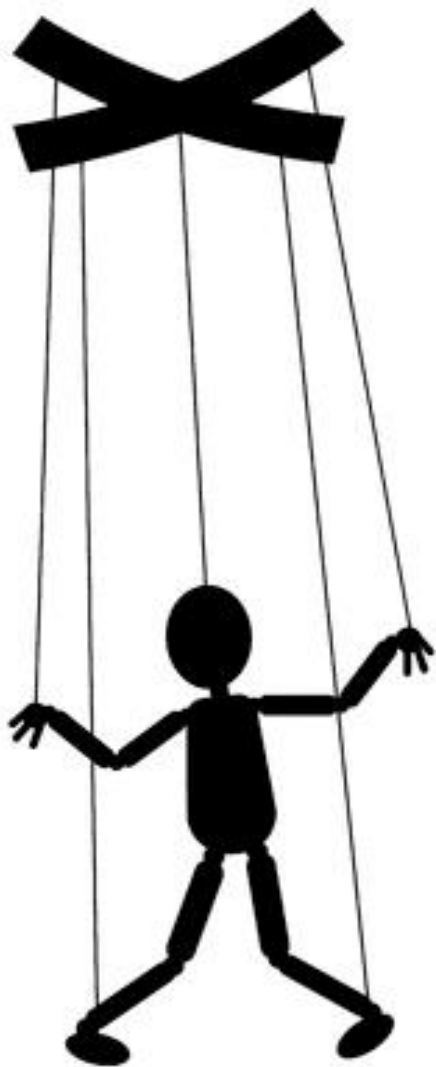
What we talk about when we talk about AI?

- Data and Datasets
- Machine learning
- Deep learning
- Neural networks
- Algorithms
- Pattern recognition
- NLP
- Autonomous machines
- Smart technologies
- ...



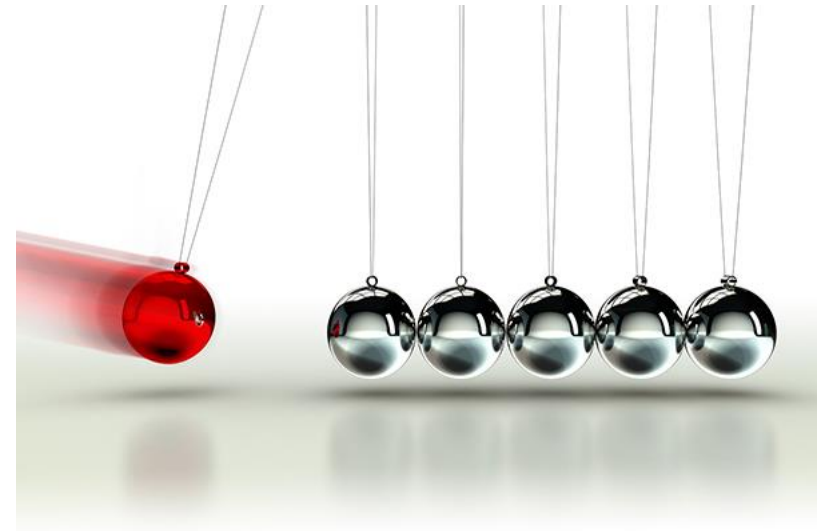
Main problem with data-driven technologies: **human no longer matter, we are used as means to an end, but not as an end in itself**

- Human beings as **data-generating machines**
- Machines to **continuously generate data** which are **used by businesses and governments/regimes** to control and manipulate
 - For the sake of **economic benefit (for a few)** or
 - as a **fuel to keep governments/regimes in place and justify violations** and limitations of fundamental rights and entitlements



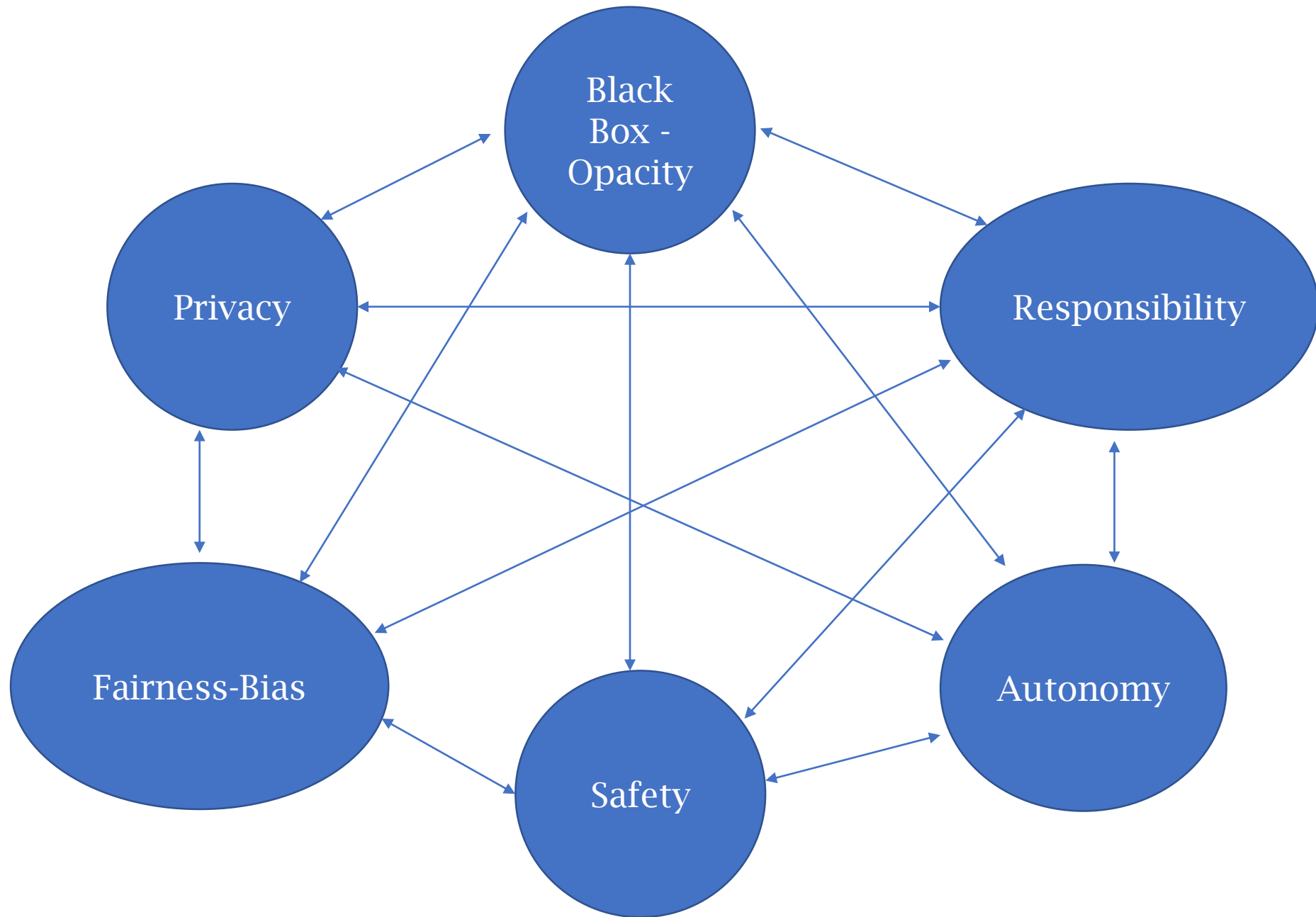
Where does AI come into play?

- all the domains of our lives
- **faster processing** of data, **development** of new services, apps, ...
- which leads to **content generation**, **recommendations**, **suggestions**, **decision-making**, ...
 - **From** tool that **support** humans **towards** tools that **substitute** human
- Impact on human lives and wellbeing of the rest of the planet.



A Few Ethical Problems

1. Privacy and personal data protection - manipulation - surveillance
2. “black boxes” - transparency/opacity - explainability
3. Accountability - responsibility - liability
4. Fairness - bias - no discrimination
5. Safety
6. Autonomy of AI versus autonomy of humans
7. ...



Ethics as a lense to make sure that **AI is working for humans** and **not the other way round**



1. Privacy and personal data protection – manipulation - surveillance

- **Data:** personal, sensitive (biometric)
- Privacy rules are *conditio sine qua non* for any AI system
- **Rules** of data minimization, data quality, privacy by design and by default, purpose limitation, consent ...
- **Manipulation** by businesses:
 - Tell me what you want/like/look at/buy, I tell you what you need:
 - Perhaps something you do not even know that you need?
 - You are what you buy/look at/like/... : **algorithmic prison**
- **Surveillance** by governments
 - to guarantee safety and security of citizens but have to do that without violating human rights and becoming Big Brother



BEHAVIOR ANALYSIS

Situational and contextual behavior analysis understands actions and responses to appropriately alert operators during instances of necessary heightened awareness. Specific types of behavior can be analyzed and understood in real-time.

2. “Black boxes” – transparency/opacity – explainability

- **Black box algorithms** are ML algorithms that may produce results that cannot be explained:
 - Reasoning chain from data to conclusion is **opaque** and code is **unaccessible**
- **Lack of transparency:** how can we trust algorithms which we do not understand and
 - If we do not understand, how can we make sure that the decisions they take are based on **correct data**?
- Transparency is not enough: to see the code does not mean to **understand it**
- Human oversight is **worthless**
- **Explainable AI**

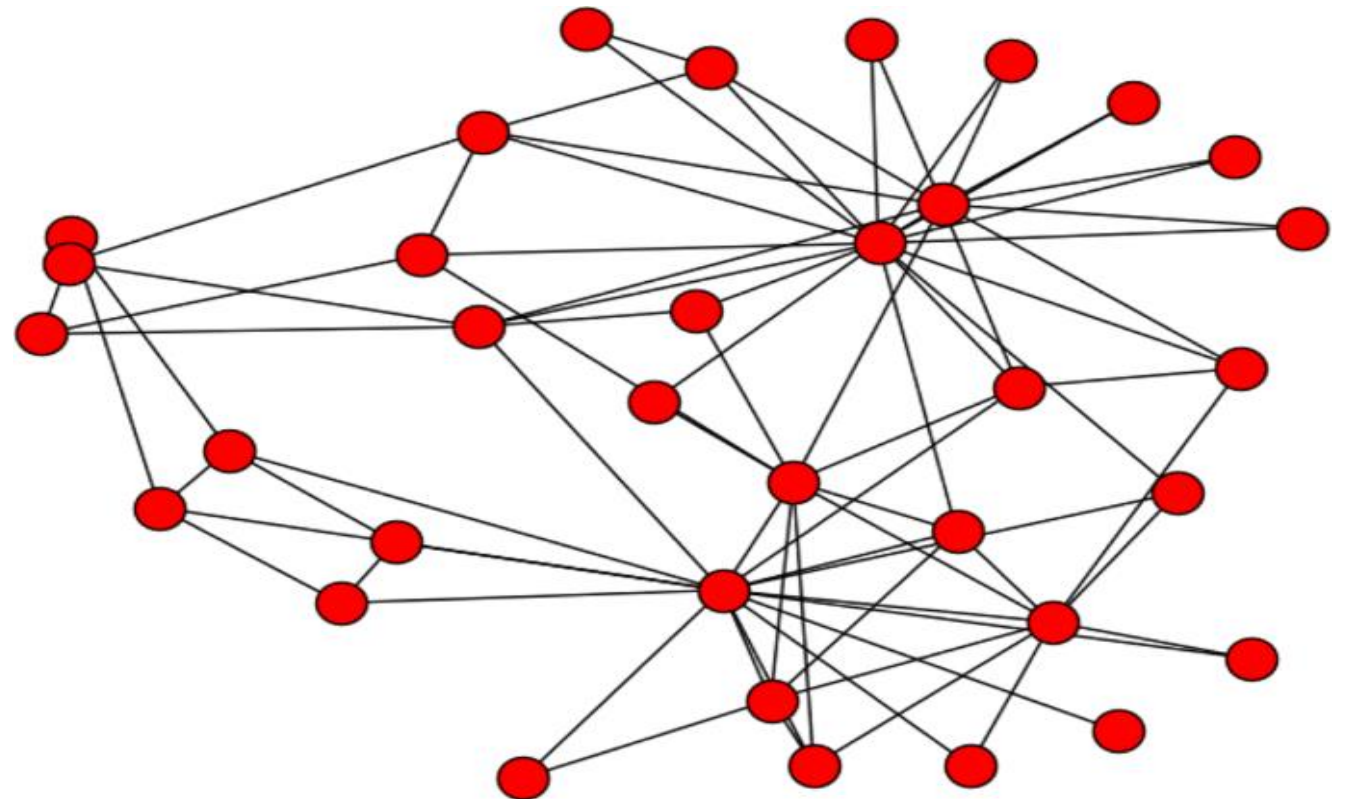
Increase **AI efficiency** and user trust

Boost your decision-making process thanks to transparent AI

 Start implementing Explainable AI for free

3. Accountability - responsibility - liability

- **Who is responsible if something goes wrong?**
- Responsibility attribution --> responsibility distribution among
 - **Users:**
 - And owners
 - **Businesses:**
 - Developers
 - Third parties (outsourcing)
 - Manufacturers
 - Maintenance/updates
 - **Government:**
 - Public administrations



4. Fairness – bias – no discrimination

- We should be all treated equally and not discriminated on the basis of **race, sex, gender, age, disability, ...**
- In any domain whatsoever, ... i.e. **employment:**
 - Reference of successful candidate in many companies is still **white and male:**
 - Which was reflected by a AI-based hiring system
 - **Facial recognition** does not take into account people with speech impediments or other difficulties
 - Measuring employee **performance by keystrokes:**
 - If you have arthritis or other motoric difficulties, your score of keystrokes could be low



amazon

Recruitment 2021

Technical Support

Work From Home

5. Safety

- Some domains where AI could bring most benefits are also the most vulnerable domains: i.e. **healthcare**
- AI has to be safe to use, technologically robust and resilient to cyber attacks, accidental misuses, or any other use that could negatively affect the person:
 - **Prediction** of disease evolution
 - **Prescription** of drugs and treatment
 - Patient record **management**
 - **NLP** to keep the record of physician-patient interaction
 - **Computer vision** and image interpretation
- **Human oversight**

Artificial Intelligence–Based Breast Cancer Nodal Metastasis Detection

Insights Into the Black Box for Pathologists

Yun Liu, PhD; Timo Kohlberger, PhD; Mohammad Norouzi, PhD; George E. Dahl, PhD; Jenny L. Smith, DO; Arash Mohtashamian, MD; Niels Olson, MD; Lily H. Peng, MD, PhD; Jason D. Hipp, MD, PhD; Martin C. Stumpe, PhD

• **Context.**—Nodal metastasis of a primary tumor influences therapy decisions for a variety of cancers. Histologic identification of tumor cells in lymph nodes can be laborious and error-prone, especially for small tumor foci.

Objective.—To evaluate the application and clinical implementation of a state-of-the-art deep learning–based artificial intelligence algorithm (LYmph Node Assistant or LYNA) for detection of metastatic breast cancer in sentinel lymph node biopsies.

Design.—Whole slide images were obtained from hematoxylin-eosin–stained lymph nodes from 399 patients (publicly available Camelyon16 challenge dataset). LYNA was developed by using 270 slides and evaluated on the remaining 129 slides. We compared the findings to those obtained from an independent laboratory (108 slides from 20 patients/86 blocks) using a different scanner to measure reproducibility.

Results.—LYNA achieved a slide-level area under the receiver operating characteristic (AUC) of 99% and a

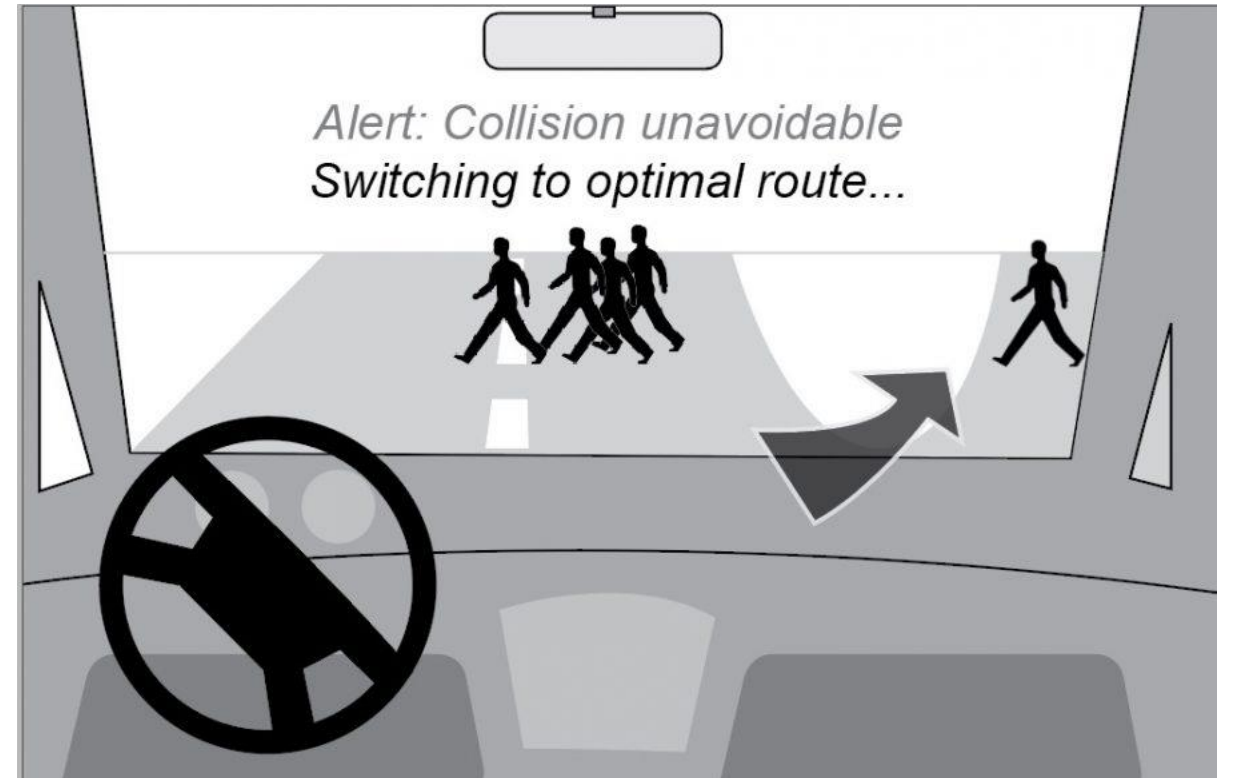
tumor-level sensitivity of 91% at 1 false positive per patient on the Camelyon16 evaluation dataset. We also identified 2 “normal” slides that contained micrometastases. When applied to our second dataset, LYNA achieved an AUC of 99.6%. LYNA was not affected by common histology artifacts such as overfixation, poor staining, and air bubbles.

Conclusions.—Artificial intelligence algorithms can exhaustively evaluate every tissue patch on a slide, achieving higher tumor-level sensitivity than, and comparable slide-level performance to, pathologists. These techniques may improve the pathologist’s productivity and reduce the number of false negatives associated with morphologic detection of tumor cells. We provide a framework to aid practicing pathologists in assessing such algorithms for adoption into their workflow (akin to how a pathologist assesses immunohistochemistry results).

(*Arch Pathol Lab Med.* 2019;143:859–868; doi: 10.5858/arpa.2018-0147-OA)

6. Autonomy

1. What **levels of autonomy** to which AI-based tools and systems?
2. Where does the **human autonomy** fit in?
3. Who decides in cases of autonomy of
 - vehicles
 - robots
 - weapons





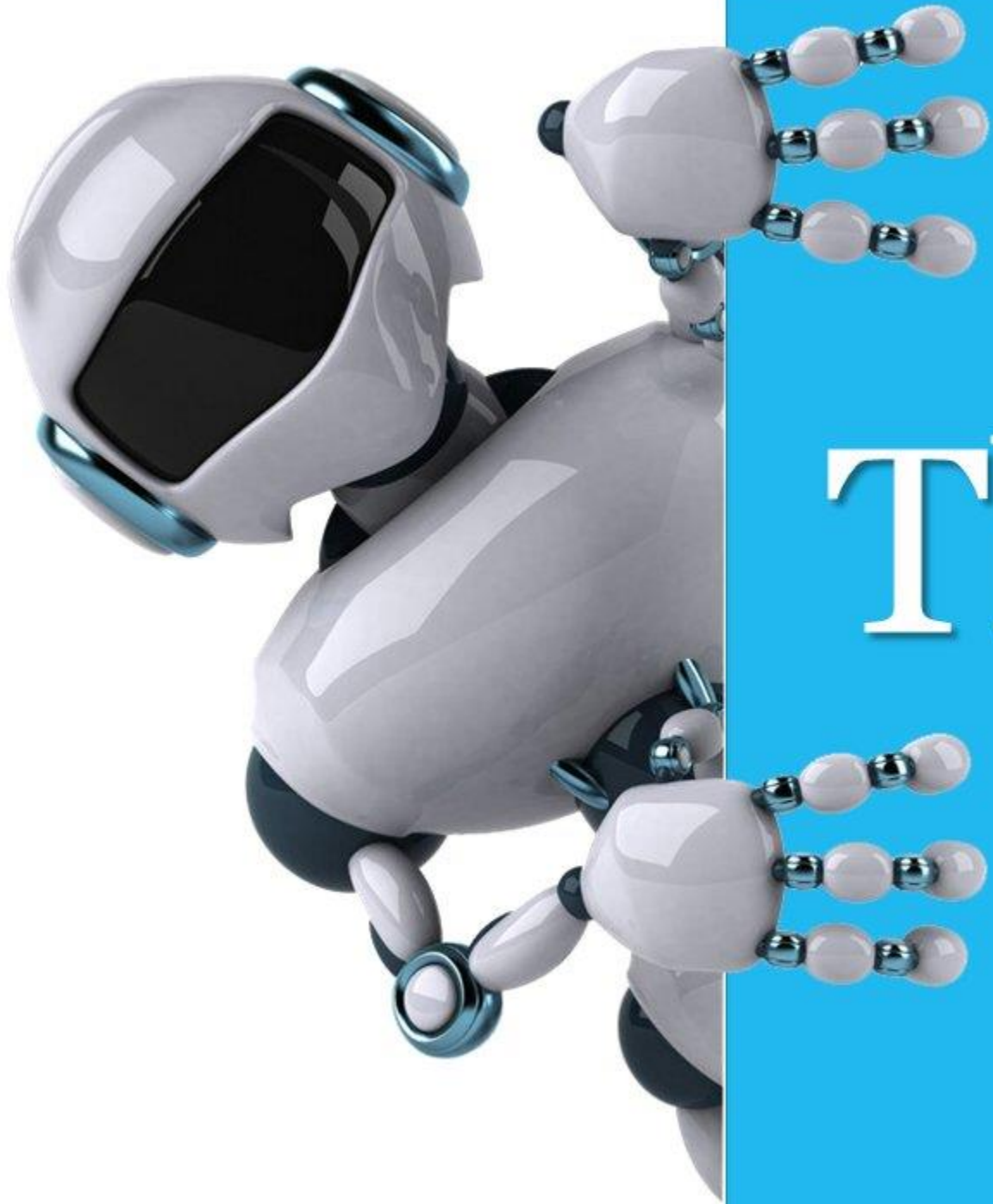
CAMPAIGN TO STOP
KILLER ROBOTS

STOP KILLER ROBOTS.ORG

CAMPAIGN
KILLER

The world is complex and each of us is a world on its own: ethics is about **letting us and not AI** decide who we are and who we want to be





Thank you