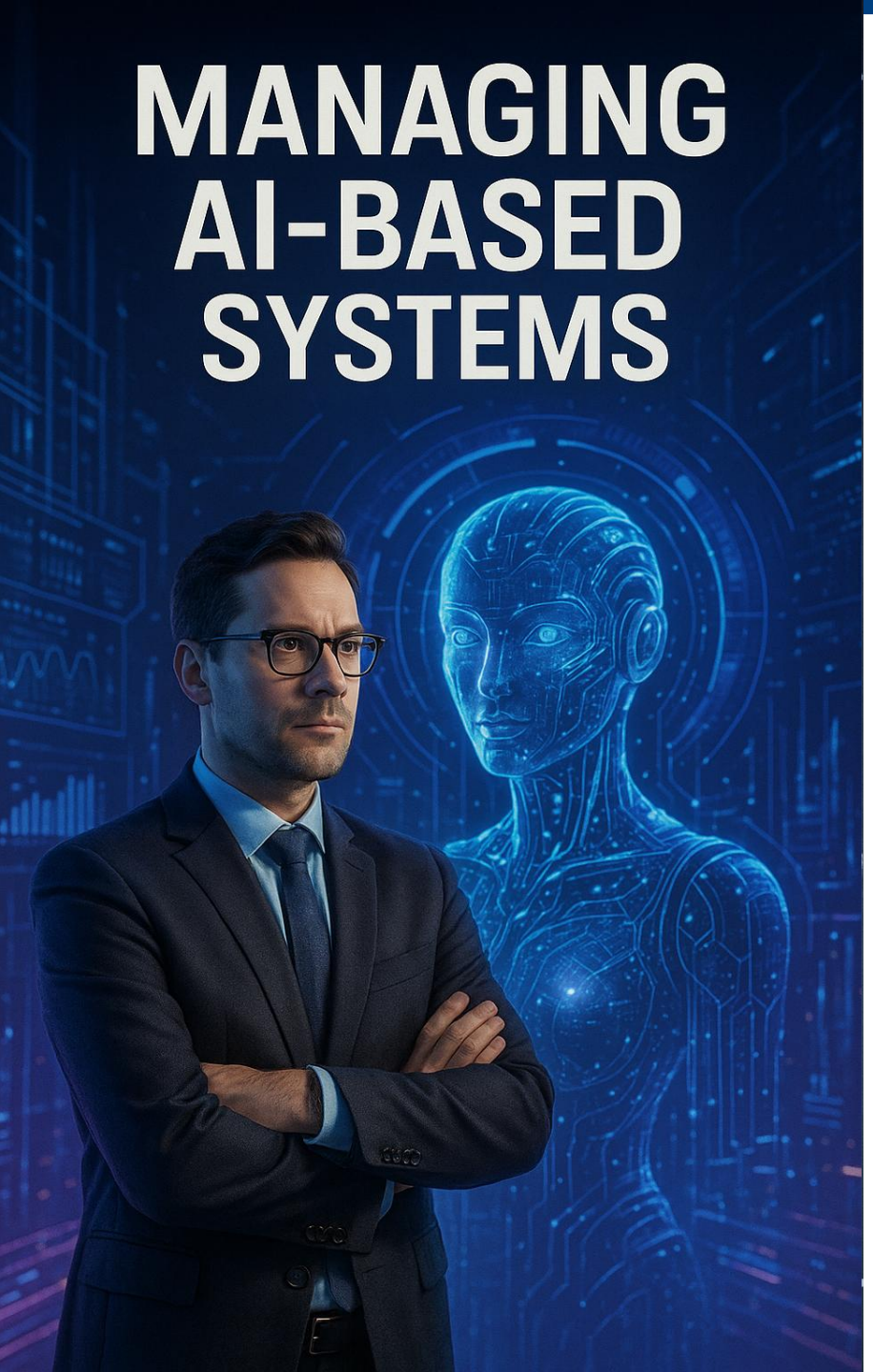



# Session 13: Ethical, legal and social implications of AI

## Managing AI-based Systems

**Prof. Dr. Nils Urbach**

Frankfurt University of Applied Sciences,
Research Lab for Digital Innovation & Transformation

FIM Forschungsinstitut für Informationsmanagement

Fraunhofer-Institut für Angewandte Informationstechnik FIT,
Institutsteil Wirtschaftsinformatik

www.ditlab.org
www.fim-rc.de
www.wirtschaftsinformatik.fraunhofer.de

# Creative Commons Copyright

# Course navigator

2 3 4 Basic understanding of AI and its business potential

## AI Ideation

5 Application potentials of AI technologies

6 Identification, design and evaluation of AI use cases

## AI Strategizing

7 Evaluation an organization's AI readiness

8 Management and governance of AI

## AI Design & Development

10 Architectures of AI applications

11 Data Management and Model Transparency

12 Design of human-AI interactions

## AI Operations at Scale

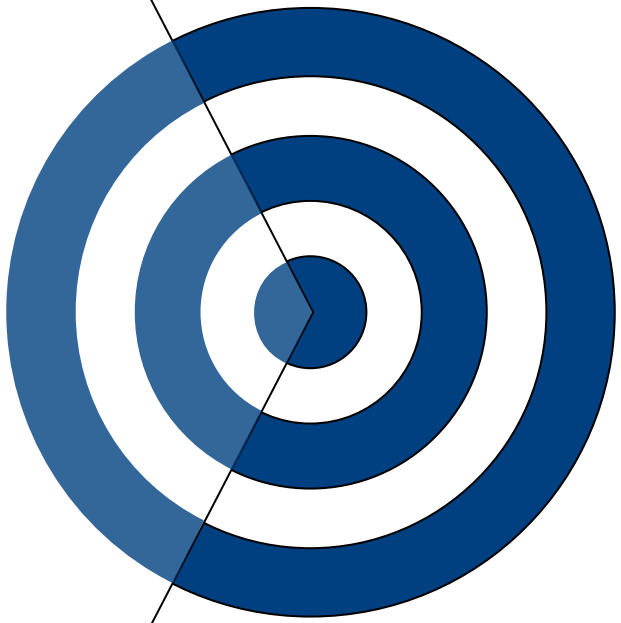
12 Monitoring and KPI-based control

13 **Ethical, legal and social implications of AI**

Implementation

# Objectives of today's lecture

1. Experience ethical decision dilemmas
2. Understand the importance of moral and trustworthy AI
3. Get to know about the legal implications (also for your own use with GenAI)



# Some AI use cases cause ethical dilemmas


The development of autonomous cars has risen ethical questions, like in case of car accident, where the car cannot prevent to hit somebody, should it rather hit a kid or an old man.

**What do you think should the car rather hit?**

Go to
www.menti.com
Enter the code
16 00 75 6

| Decision dimensions | | |
|---|---|---|
| Psychological (social) | Ethical | Legal |

How should such decision situations be implemented in an artificially intelligent application?

# Ethical dilemma - Age as the differentiator

Rather hit the…

…young kid or…

…the old man

Vote now

# Ethical dilemma - Gender as the differentiator

Rather hit the…

…young man or…

…the young woman

Vote now

# Ethical dilemma - Species as the differentiator

Rather hit the…

…dog or…

…the old man

Vote now

# Agenda

01 | Ethics of AI

02 | Moral agency

03 | Trustworthy AI

04 | Ethic standards

05 | Legal implications

# Agenda

01 | Ethics of AI

02 | Moral agency

03 | Trustworthy AI

04 | Ethic standards

05 | Legal implications

# What is Ethics and why is it important for AI?

**Definition of Ethics:** A system of principles/ rules/ guidelines that assists to determine what is good or bad and deals with moral obligations of entities (e.g., humans, intelligent robots, etc.)

**Ethics of AI:** Subgroup of ethics of advanced technology with a focus on intelligent agents and can be divided into roboethics and machine ethics

- Transparency
  - Ethical Standards
- Data Security and Privacy
  - Automation and Job Replacement
- Autonomy, Intentionality and Responsibility
  - Human Rights Laws
- Human Bias
  - Accessibility
- Accountability
  - Democracy and Civil Rights

**Ethics and AI is important to combine in order to ensure that there is no harm for humans in the Human-AI interaction (e.g. through deep fakes, fake news, unfairness to minorities ,…)**

*(1) Siau & Wang, 2020*

# What is Ethics and why is it important for AI?

**Definition of Ethics:** a system of principles/ rules/ guidelines that assists to determine what is good or bad and deals with moral obligations of entities (e.g., humans, intelligent robots, etc.)

**Ethics of AI:** subgroup of ethics of advanced technology with a focus on intelligent agents and can be divided into roboethics and machine ethics

**Roboethics: D**eals with human moral behavior in the design, construction, use and interaction with AI agents (human bias in data, privacy, unemployment) but also robot right if they become conscious one day

**Machine ethics: C**oncerned with moral behaviors of Artificial Moral Agents (AMAs) in order to get ethical AI (Asimov's three laws of robotics)

**Ethics and AI is important to combine in order to ensure that there is no harm for humans in the Human-AI interaction (e.g. through deep fakes, fake news, unfairness to minorities ,…)**

*(1) Siau & Wang, 2020*

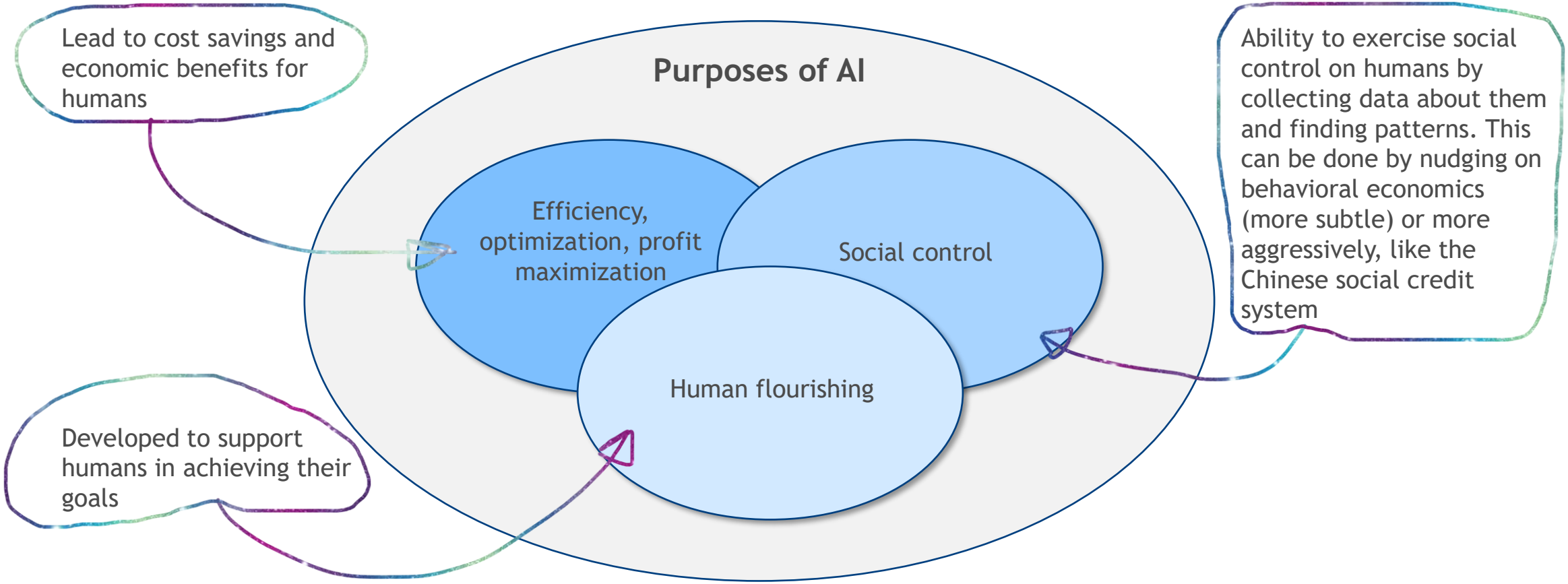
# Isaac Asimov’s three laws of robotics

**Three laws of robotics**

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm
2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law

*Asimov, 1950*

# The purposes of AI




**The three different purposes of AI can be contradictory and synergetic. But the ethically right thing can be best described with the purpose of human flourishing.**

*Stahl & Stahl, 2021*

# Agenda

01 | Ethics of AI

02 | Moral agency

03 | Trustworthy AI

04 | Ethic standards

05 | Legal implications

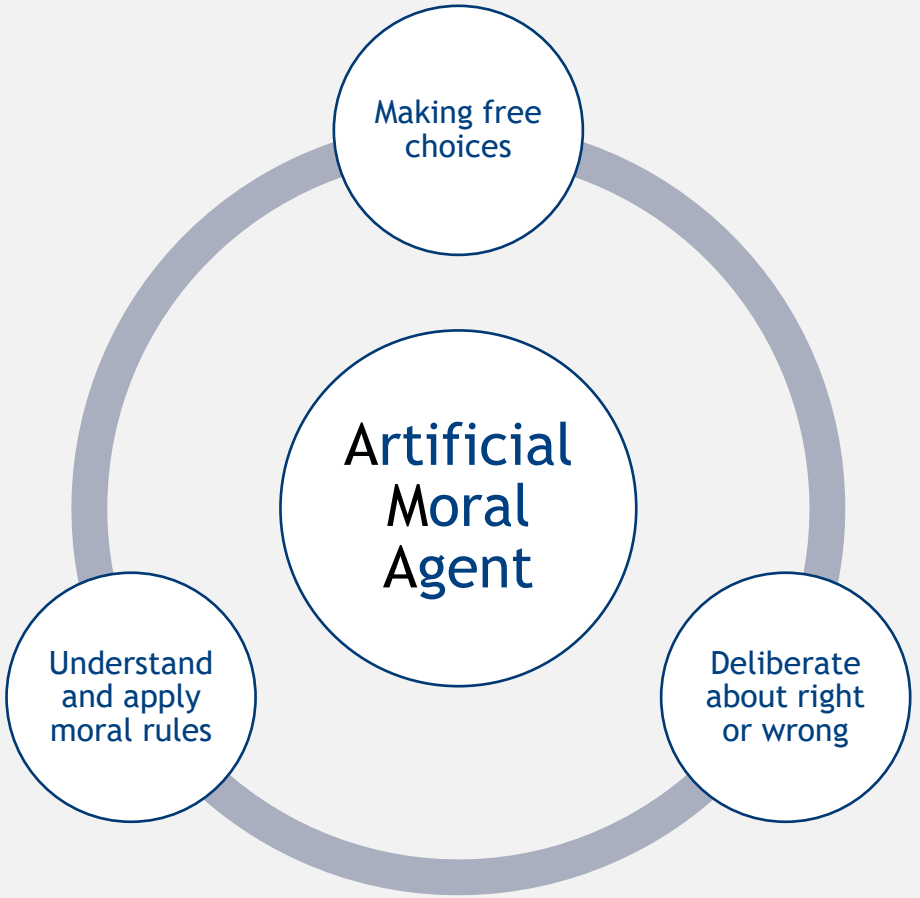
# What is moral agency and what is needed to achieve AMA?

**A**rtifical intelligence should not diminish our agency but we can use it as a tool to enhance it.

**M**oral agency states the idea of being accountable for one's behavior, which is an important consideration for Human-AI interaction.

**A**gents that act morally are governed by moral standards, which come with moral obligations.

**Necessary characteristics for moral agency**




*(1) Himma, 2009, (2) Boddington, 2020*

# What speaks for and what against building AMAs?

| Pros | Cons |
| --- | --- |
| 1. Preventing harm | 1. Seems to be impossible to build them |
| 2. Complexity | 2. We want robots/ AIs to remain our “slaves” |
| 3. Public trust | 3. No universal agreement in ethics |
| 4. Preventing immoral use | 4. Safe machines are enough |
| 5. Machines are better than us | 5. Existential concerns |
| 6. Understanding morality better | 6. Morality forbids it |
| | 7. Moral deskilling |
| | 8. Domain-specific concerns |
| | 9. Responsibility concerns |

*Formosa & Ryan, 2021*


Can artificial moral agents improve trustworthiness?

# Agenda

01 | Ethics of AI

02 | Moral agency

03 | Trustworthy AI

04 | Ethic standards

05 | Legal implications

# When is an AI system trustworthy?

**Trustworthiness:** "if the humans and machines involved can rely on each other, self-organize to take advantage of each others' strengths and mitigate their weaknesses, and can be held accountable for their actions (Ramchurn et al., 2021, P.3)

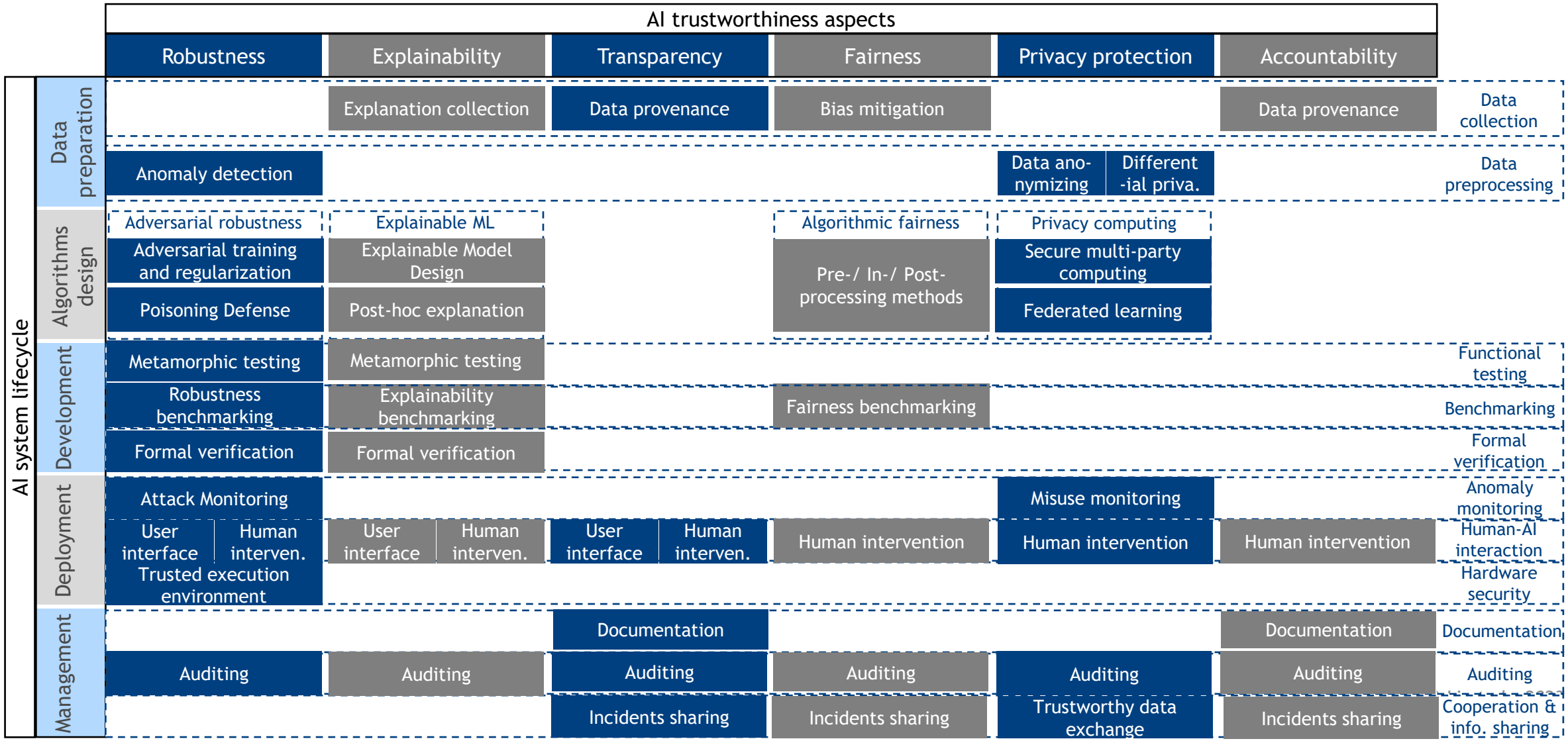
**Trustworthy Artificial Intelligence**

Trustworthy AI should be robust and comply with legal and ethical requirements.

| Respect for Human Control | Prevention of Harm | Fairness | Explainability |
|---|---|---|---|
| ▪ Human Agency and Oversight | ▪ Technical Robustness & Safety<br>▪ Privacy & Data Governance<br>▪ Societal Well-Being<br>▪ Accountability | ▪ Diversity, Non-Discrimination & Fairness | ▪ Transparency |

*(1) Ramchurn et al., 2021, (2) Kaur et al., 2020, (3) Kumar et al., 2020*

# Approaches to trustworthiness across the AI lifecycle



# Approaches to trustworthiness across the AI lifecycle

Trustworthy AI: From Principles to Practices

BO LI, JD Technology, China and Tsinghua University, China
PENG QI, Amazon AWS AI Labs, USA
BO LIU, Walmart Inc., USA
SHUAI DI, JD Technology, China
JINGEN LIU, JD Technology, USA
JIQUAN PEI, JD Technology, China
JINFENG YI, Frontis.AI, China
BOWEN ZHOU, Tsinghua University, China and Frontis.AI, China

177

The rapid development of Artificial Intelligence (AI) technology has enabled the deployment of various systems based on it. However, many current AI systems are found vulnerable to imperceptible attacks, biased against underrepresented groups, lacking in user privacy protection. These shortcomings degrade user experience and erode people's trust in all AI systems. In this review, we provide AI practitioners with a comprehensive guide for building trustworthy AI systems. We first introduce the theoretical framework of important aspects of AI trustworthiness, including robustness, generalization, explainability, transparency, reproducibility, fairness, privacy preservation, and accountability. To unify currently available but fragmented approaches toward trustworthy AI, we organize them in a systematic approach that considers the entire lifecycle of AI systems, ranging from data acquisition to model development, to system development and deployment, finally to continuous monitoring and governance. In this framework, we offer concrete action items for practitioners and societal stakeholders (e.g., researchers, engineers, and regulators) to improve AI trustworthiness. Finally, we identify key opportunities and challenges for the future development of trustworthy AI systems, where we identify the need for a paradigm shift toward comprehensively trustworthy AI systems.

CCS Concepts: • **Computing methodologies** → **Artificial intelligence**; **Machine learning**; • **General and reference** → **Surveys and overviews**;

Additional Key Words and Phrases: Trustworthy AI, robustness, generalization, explainability, transparency, reproducibility, fairness, privacy protection, accountability

**ACM Reference format:**
Bo Li, Peng Qi, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi, and Bowen Zhou. 2023. Trustworthy AI: From Principles to Practices. *ACM Comput. Surv.* 55, 9, Article 177 (January 2023), 46 pages.
https://doi.org/10.1145/3555803

Authors' addresses: B. Li, JD Technology, Beijing, China and Tsinghua University, Beijing, China; email: prclibo@gmail.com; P. Qi, Amazon AWS AI Labs, Seatle, WA, USA; email: qipeng.thu@gmail.com; B. Liu, Walmart Inc., Mountain View, CA, USA; email: kfliubo@gmail.com; S. Di, JD Technology, Beijing, China; email: dishuai@jd.com; J. Liu, JD Technology, Mountain View, CA, USA; email: jingenliu@gmail.com; J. Pei, JD Technology, Beijing, China; email: peijiquan@jd.com; J. Yi, Frontis.AI, Beijing, China; email: jinfengyi.ustc@gmail.com; B. Zhou (corresponding author), Tsinghua University, Beijing, China and Frontis.AI, Beijing, China; email: zhoubowen@tsinghua.edu.cn.

0360-0300/2023/01-ART177 $15.00
https://doi.org/10.1145/3555803



For more information regarding the different factors, we can suggest following reading, Chapter 3

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., ... & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, *55*(9), 1-46. https://dl.acm.org/doi/pdf/10.1145/3555803

# Agenda

01 | Ethics of AI

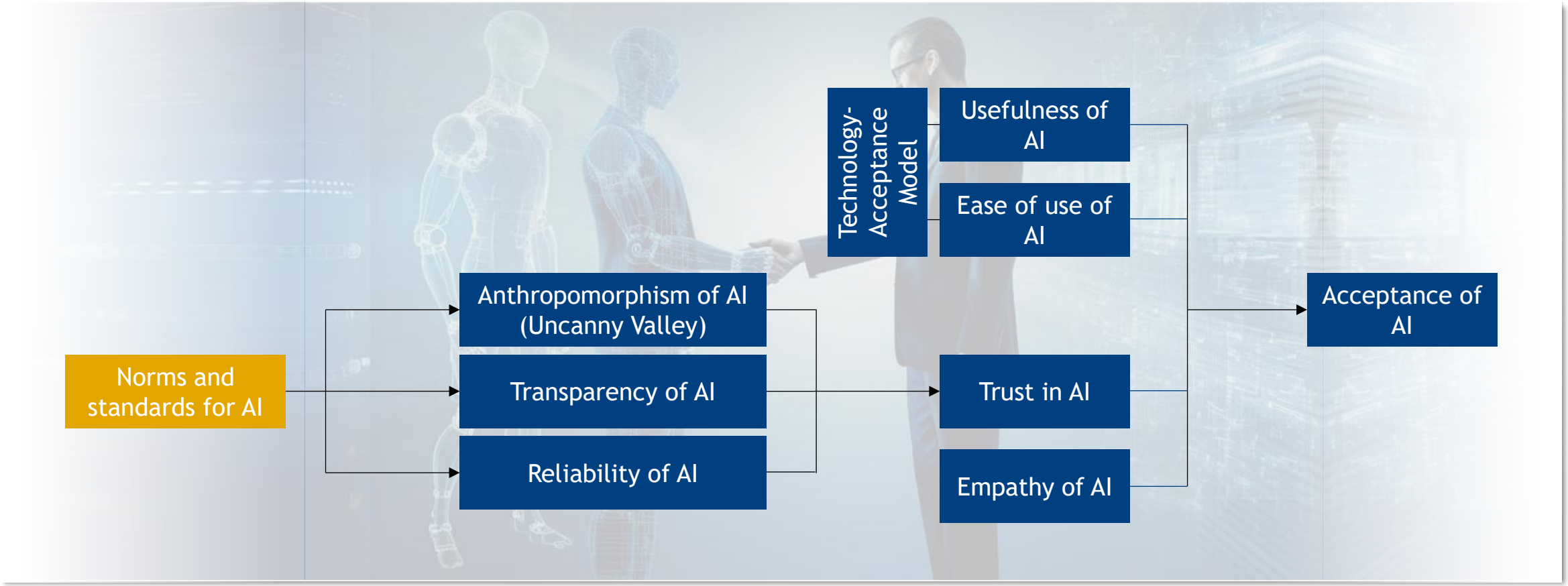
02 | Moral agency

03 | Trustworthy AI

04 | Ethic standards

05 | Legal implications

# Do you remember norms and standards as a base for trust?




**Norms and standards do not only exist from a technical standpoint but also from an ethical one**

*Scheuer (2020), Zhang et al. (2010), Siau & Wang (2018), Donner (2021)*

# Ethical standards for AI are a hot topic for Germany and EU


Die Bundesregierung

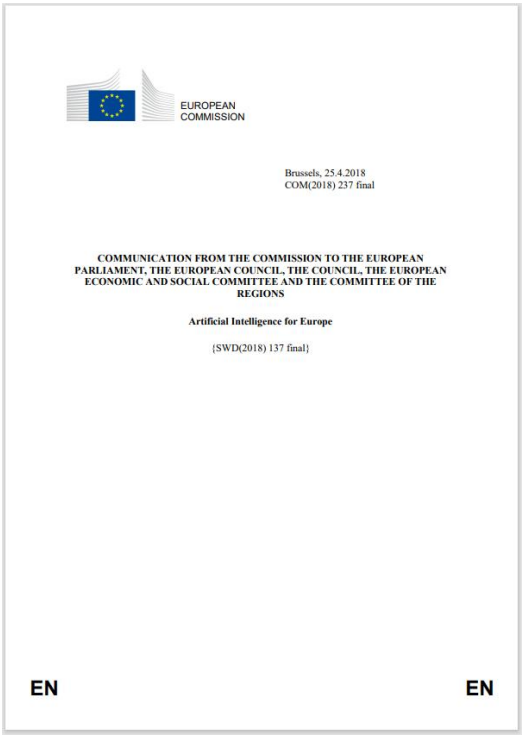
KI Nationale Strategie für Künstliche Intelligenz
AI Made in Germany

Strategie Künstliche Intelligenz der Bundesregierung

Stand: November 2018

www.ki-strategie-deutschland.de

Ein wichtiges Augenmerk wird die Bundesregierung auf Ethikaspekte in der Normung und Standardisierung für KI in autonomen Maschinen und Fahrzeugen setzen (*„ethics by design“*). Dies werden wir auch bei der Prüfung des Aufbaus von Strukturen und Prozessen zur Auditierung der Entwicklung und Anwendung von KI berücksichtigen (siehe auch Handlungsfeld 3.5).


EUROPEAN COMMISSION

Brussels, 25.4.2018
COM(2018) 237 final

**COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE EUROPEAN COUNCIL, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS**

**Artificial Intelligence for Europe**

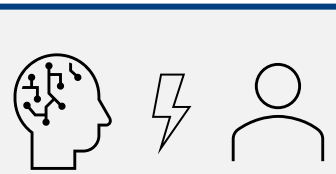
{SWD(2018) 137 final}

EN EN

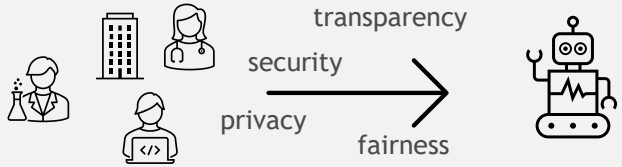
**Launching a European initiative on AI**

The European Council of October 2017 stated that the EU needs a sense of urgency to address emerging trends such as AI "while at the same time ensuring a high level of data protection, digital rights and ethical standards" and invited "the Commission to put forward a **European approach to artificial intelligence**".[7] The European Parliament made wide-

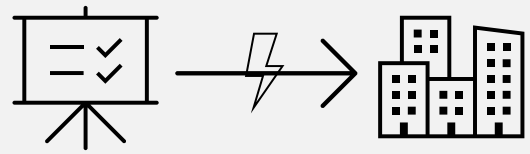
# Value-based engineering (VBE) with the IEEE 7000 standard



- Fear of strong AI
- Ethical dilemmas of existing systems




- Scientists and institutions demand value principles to guide (privacy, security, transparency, fairness, …)



- BUT: tough parts of ethics (translating theories into practices) is not solved by the means of value principles

**Value-based engineering**

- ➢ Provides processes to facilitate value reflection for system development teams and managers
- ➢ Anticipates the intricate value framework of upcoming systems and contemplates them to serve the stakeholders' interests
- ➢ Identified values are prioritized and translated into system requirements by using “Ethical Value Requirements”

**VBE ensures that value harms and benefits are addressed systematically by basing innovator’s systems on a positive value mission**

*(1) Spiekermann, 2021, (2) Spiekermann & Winkler, 2022*

# Agenda

01 | Ethics of AI

02 | Moral agency

03 | Trustworthy AI

04 | Ethic standards

**05 | Legal implications**

# 3 characteristics of AI that raise legal issues

**Autonomy**

Acting without human intervention

*Can an AI be held guilty?*

**Learning**

Improving through data and experience

*Who is liable in connected AI systems?*

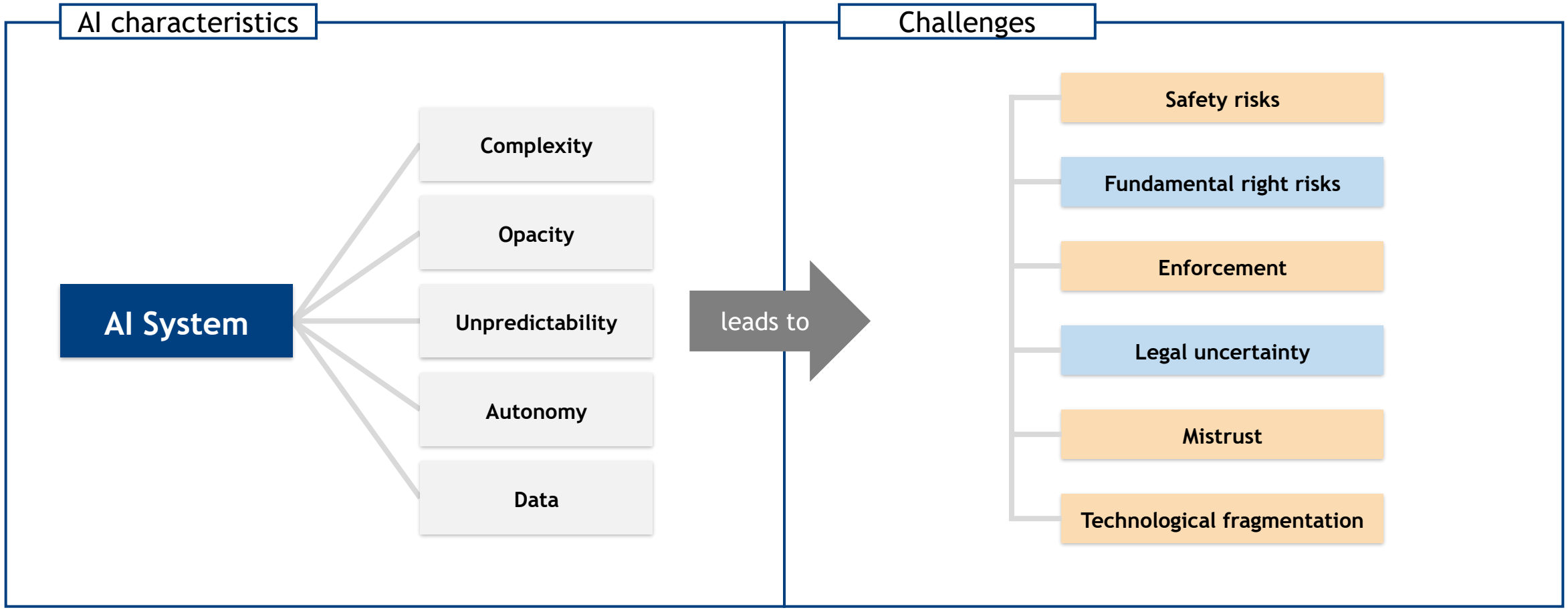
**Inscrutability**

Being unintelligible to multiple audiences

*Can decisions made by an AI be explained and understood at all?*

*Berente et al., 2021*

# Why do we regulate AI use cases?

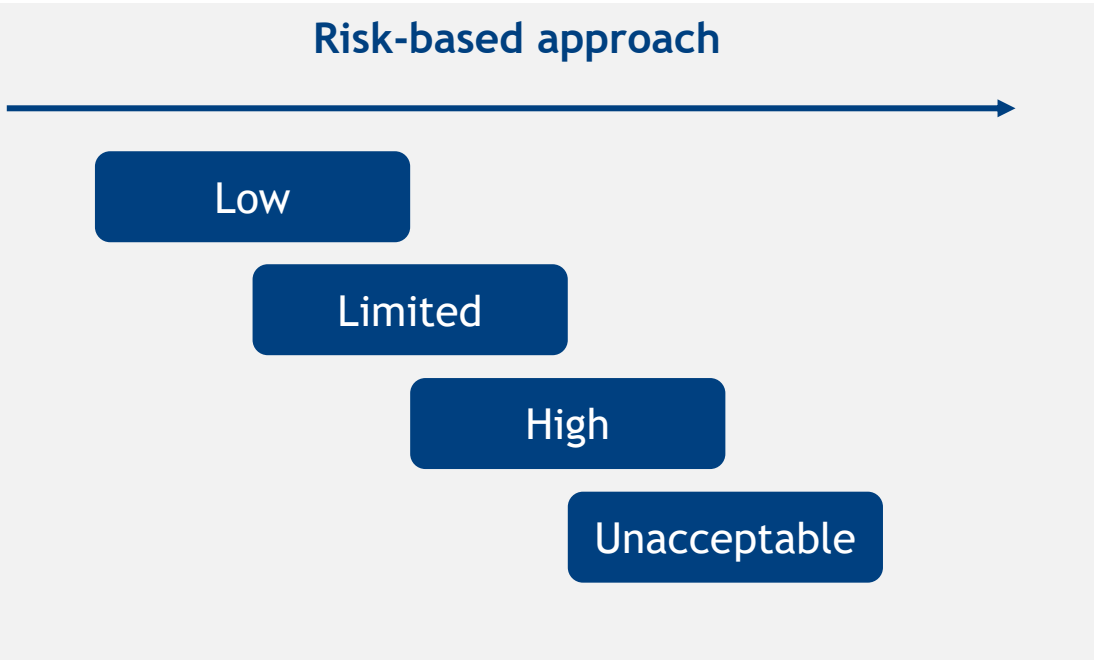



*Sioli, L. (2021)*

# The European Commission's "AI Act" as a legal framework to address AI risks

## What is the AI Act?

- Draft regulation establishing harmonized rules for the placing on the market, putting into service and use of artificial intelligence systems
- Included: High compliance requirements and transparency requirements for certain AI applications
- Addressed to providers and users of AI systems (relates to development and commissioning)

# The European Commission's AI Act as a legal framework to address AI risks

## Risk-based regulation of AI use-cases

| Low | Limited | High | Unacceptable |
|---|---|---|---|
| • AI-enabled video games<br>• Spam filters | • E.g., chatbots<br>• Image generators<br>• Most other generative AI applications | • Individual profiling<br>• Automated processing of personal data<br>• Operation of critical infrastructure | • Manipulation/deception with significant harm<br>• Exploiting vulnerabilities<br>• Social scoring<br>• Biometric categorization |
| Unregulated | Transparency obligations: inform end-user about interaction with AI | Risk management, governance and documentation obligations | Ban |

EU AI act pushes most obligations and responsibilities on **developers**, while **deployers** face only little regulations and **end-users** are almost completely exempt from any duties.

*EU (2024)*

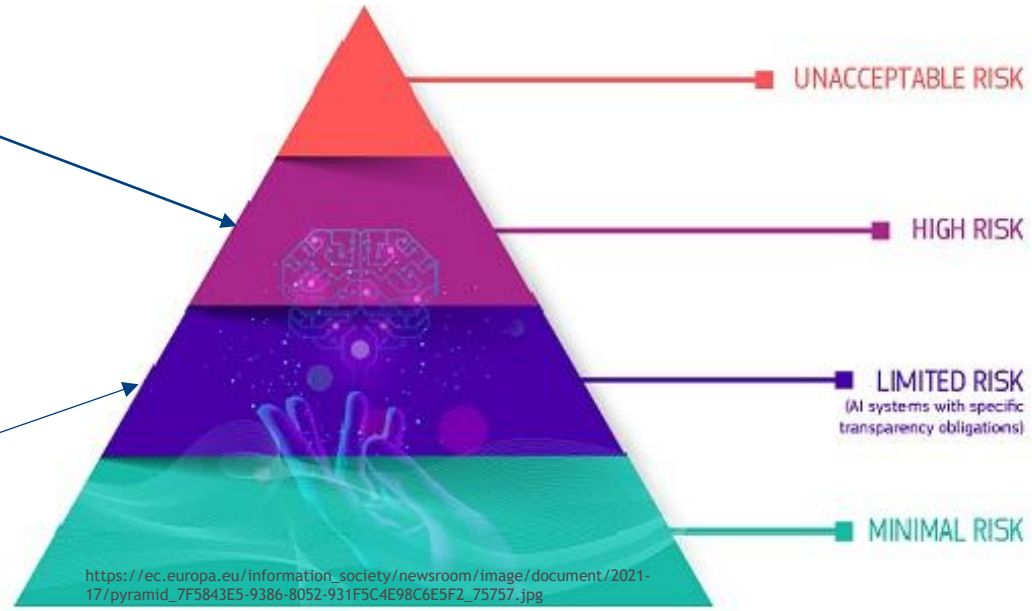
# Extract from the EU AI Act and its transparency requirements

## Art. 13 AI Act (High Risk AI Systems)

- "shall be designed and developed in such a way to ensure that their **operation is *sufficiently* transparent to enable deployers to interpret the system's output and use it appropriately**"
- "shall be accompanied by **instructions for use** […] that include concise, complete, correct and clear information that is relevant, accessible and **comprehensible to users**", including i.e. **characteristics, capabilities, and limitations of performance of the AI system**

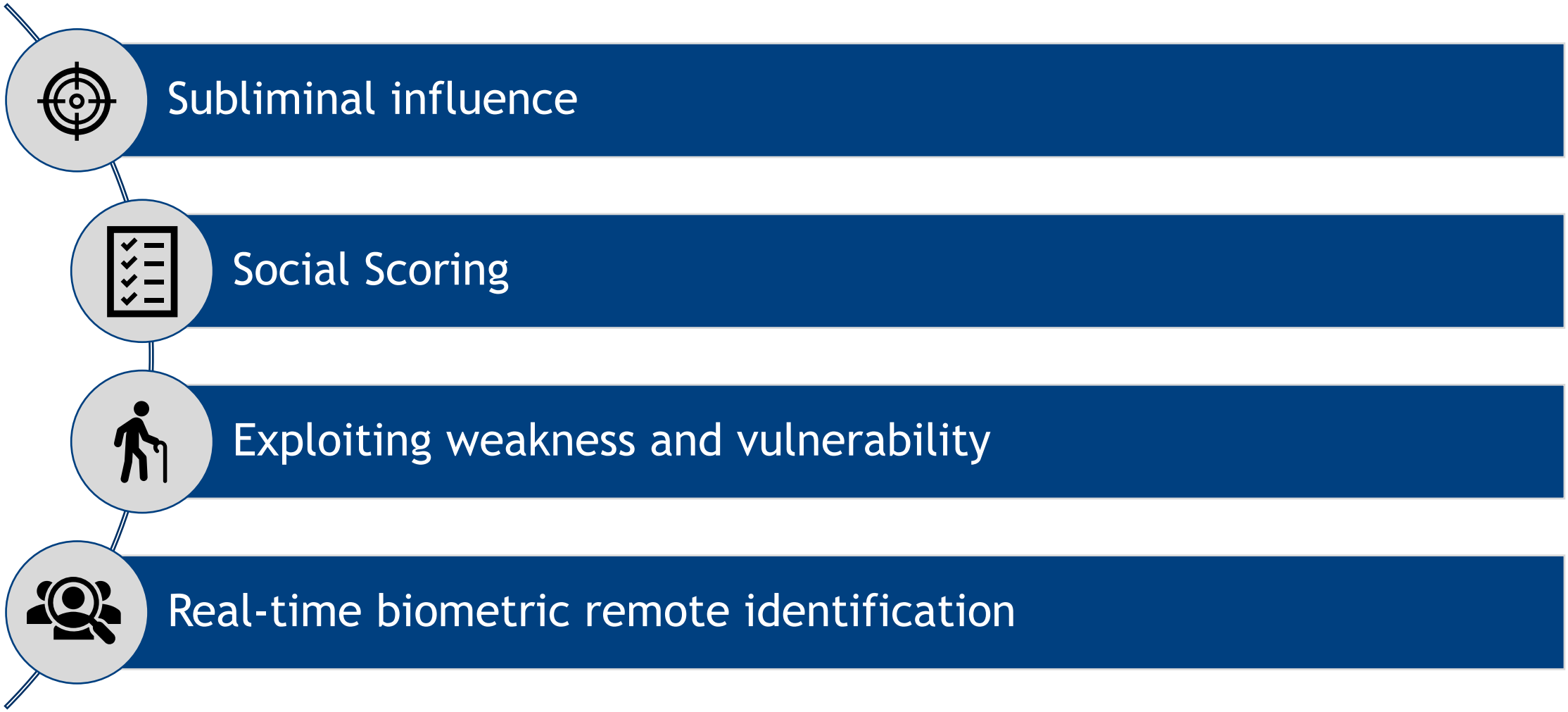
## Art. 52 AI Act (Limited Risk AI Systems)

- "shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that **natural persons are informed that they are interacting with an AI system**, unless this is *obvious from the circumstances and the context of use*"

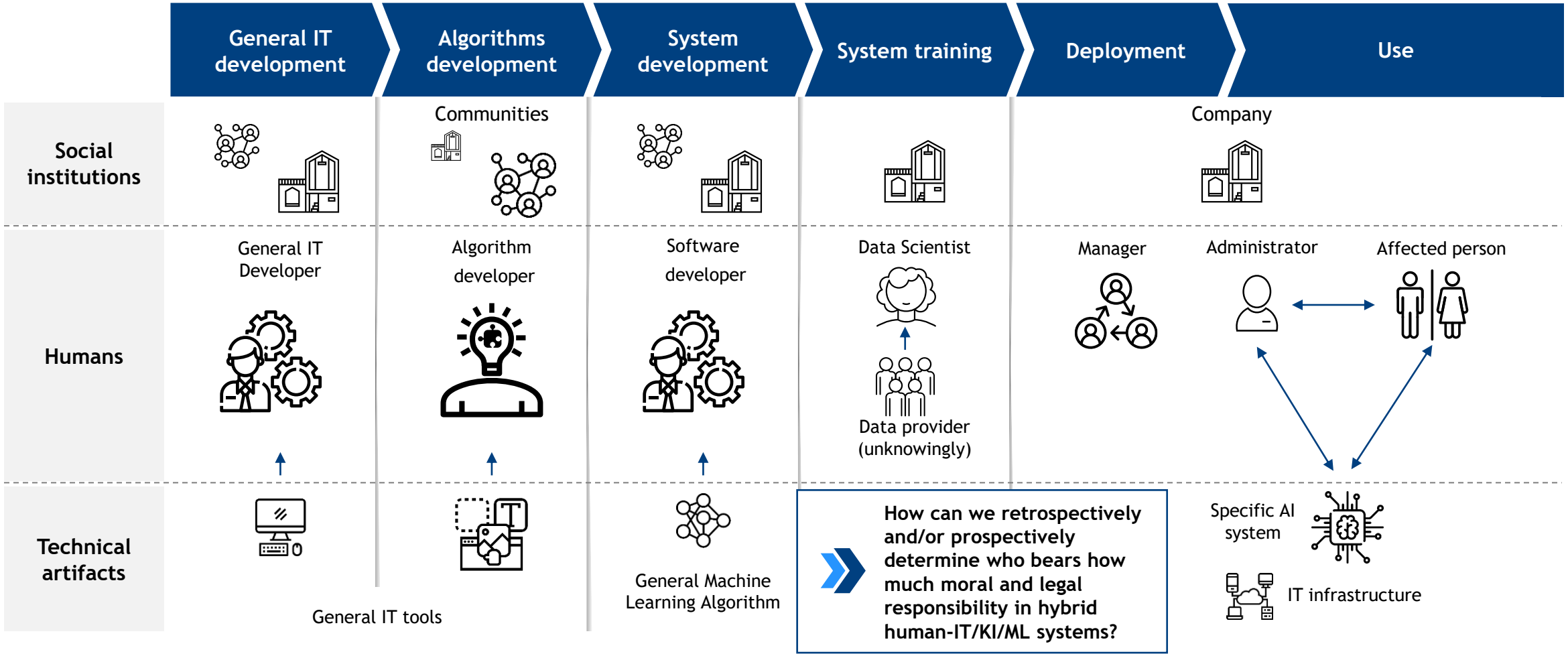



*EU (2024)*

# In addition to compliance requirements and transparency regulations, the AI Act also prohibits AI practices



# A stylized perspective on the system/network that leads to immoral behavior/outcomes




*(1) Flaticon, (2) Fritz et al. (2021), (3) Schnaak & Gimpel, 2023*

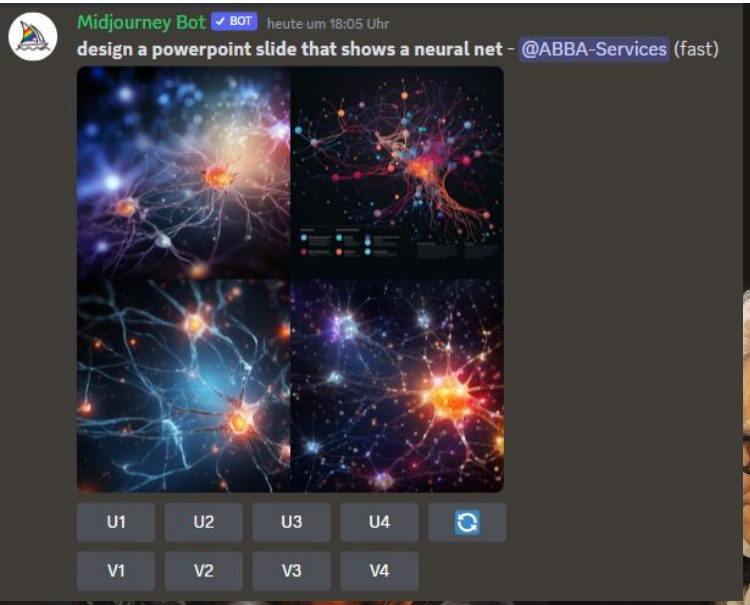
# Generative AI – Legal aspects for users

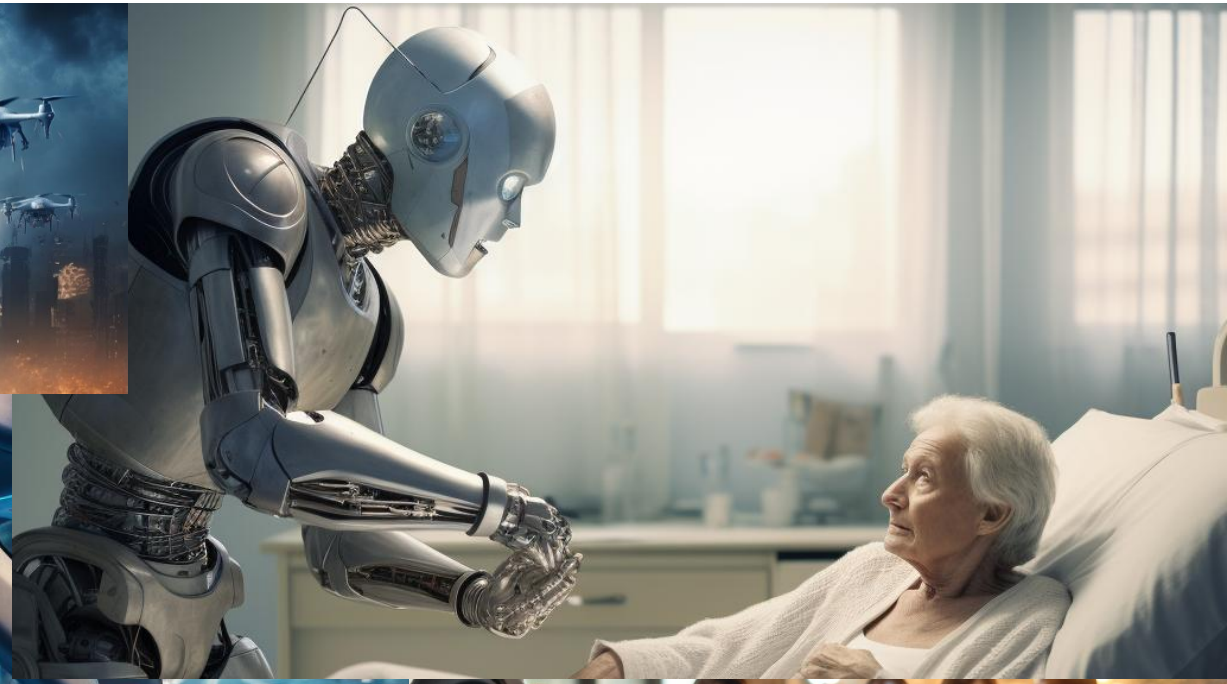
| Type of generative AI | Text-generative AI | Text-to-image generative AI |
|---|---|---|
| Example | ChatGPT | Midjourney |
| Legal implications | **Can I use my generated text as I want?**<br>➢ Yes, you own all rights on the generated output<br>➢ Not clearly defined if it is limitless for commercial use<br>➢ BUT: You should not claim it as the work of humans<br>➢ Use of ChatGPT for scientific work (bachelor, master thesis, …):<br>  ➢ Follow the guideline of your university | **Can I use my generated pictures as I want?**<br>➢ Yes, there are no limitations for the non-commercial use<br>➢ The fee-based abonnement gives you a license<br>  ➢ You don't even have to cite<br>➢ BUT: Midjourney also has a license to use the pictures<br>➢ There are limitations for commercial use, like pictures generated by others are not allowed to be used commercially |

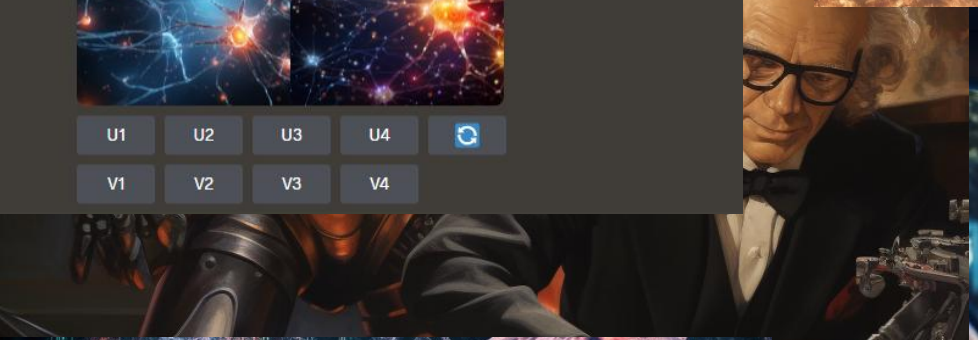
**The generated outcomes of AIs are most of the times yours if you own a license for the service but the operator itself always keeps a right to use it as well**

*(1) Stahl & Stahl, 2021*

# Many pictures in this course were generated with Midjourney…



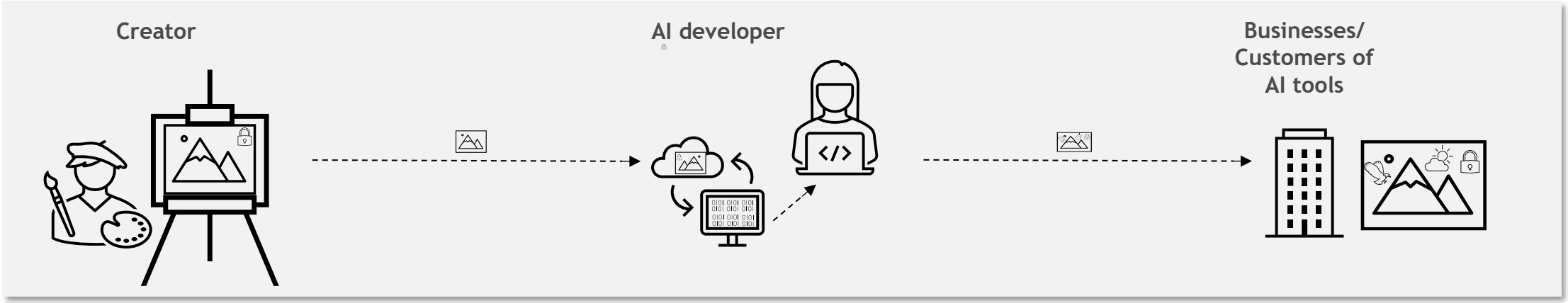






# GenAI has an intellectual property issue

**Process of intellectual property issues and actions to mitigate risks by different stakeholders**




| | Creator | AI developer | Businesses/ Customers of AI tools |
|---|---|---|---|
| Process | ▪ Creator designs new art | ▪ Developer uses databases to train the models of GenAI and generate new art based on existing | ▪ Businesses use generated art |
| Actions | ▪ Proactively protect their work by looking for their work in datasets by using search tools (including visual elements, textual elements, …) | ▪ Ensure to be in compliance with the law regarding the data used for the model training<br>➢ License or compensate owners of IP (by licensing or revenue sharing) | ▪ Ask providers if models were trained with unlicensed data, read terms of service and avoid AI tools that do not confirm properly licensed data |

**There is the risk of infringement (direct or unintentional) because data might consist of unlicensed work**

*(1) Appel et al., 2023*

# Today’s lecture at a glance

- We have experienced ethical decision dilemmas that AIs need to solve
- We have gained knowledge about the different aspects of trustworthiness and how to achieve it across the AI system lifecycle
- We understood why value-based engineering approach leads to more ethical, technical product innovation
- We know how to handle GenAI outputs and how to decrease the risk of infringement

# Questions, comments, observations

# Scientific references


- Asimov, I. (1950). Runaround. In I, Robot (The Isaac Asimov Collection Ed.). New York City: Doubleday.
- Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. MIS quarterly, 45(3), 1433-1450.
- Boddington, P. (2021). AI and moral thinking: how can we live well with machines to enhance our moral agency?. AI and Ethics, 1, 109-111.
- Formosa, P., & Ryan, M. (2021). Making moral machines: why we need artificial moral agents. AI & society, 36, 839-851.
- Himma, K. E. (2009). Artificial agency, consciousness, and the criteria for moral agency: What properties must an artificial agent have to be a moral agent?. Ethics and Information Technology, 11, 19-29.
- Kaur Davinder, Uslu Suleyman, and Durresi Arjan. 2020. Requirements for trustworthy artificial intelligence—A review. In Proceedings of the International Conference on Network-Based Information Systems. 105–115.
- Kumar Abhishek, Braud Tristan, Tarkoma Sasu, and Hui Pan. 2020. Trustworthy AI in the age of pervasive computing and big data. In Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops'20). IEEE, Los Alamitos, CA, 1–6.
- Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., ... & Zhou, B. (2023). Trustworthy AI: From principles to practices. ACM Computing Surveys, 55(9), 1-46. https://dl.acm.org/doi/pdf/10.1145/3555803
- Ramchurn, S. D., Stein, S., & Jennings, N. R. (2021). Trustworthy human-AI partnerships. Iscience, 24(8).

# Scientific references


- Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: ethics of AI and ethical AI. Journal of Database Management (JDM), 31(2), 74-87.
- Sioli, L. (2021) A European Approach to Artificial Intelligence. https://cor.europa.eu/en/events/Documents/SEDEC/FINAL%20PDF%20AI%20Presentatiofor%20COR%20Sedec%20Committee%20meeting%2023%2006%2021.pdf
- Stahl, B. C., & Stahl, B. C. (2021). Ethical issues of AI. Artificial Intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies, 35-53.
- Spiekermann, S. (2021, October). From value-lists to value-based engineering with IEEE 7000™. In 2021 IEEE International Symposium on Technology and Society (ISTAS) (pp. 1-6). IEEE.
- Spiekermann, S., & Winkler, T. (2022). Value-Based Engineering With IEEE 7000. IEEE Technology and Society Magazine, 41(3), 71-80.

# Non-scientific references


- moralmachine.mit.edu/
- https://docs.midjourney.com/docs/terms-of-service
- https://www.urheberrecht.de/midjourney/
- https://haerting.de/wissen/ki-und-copyright-wie-haelt-es-chatgpt-mit-dem-urheberrecht/
- https://openai.com/policies/terms-of-use
- https://commons.wikimedia.org/wiki/File:Midjourney_Emblem_%E2%80%94_June.png
- https://commons.wikimedia.org/wiki/File:Midjourney_Emblem_%E2%80%94_June.png
- https://hbr.org/2023/04/generative-ai-has-an-intellectual-property-problem
- https://medium.datadriveninvestor.com/you-will-be-arrested-by-data-this-is-predictive-policing-623d9e14afec
- https://www.technologyreview.com/2020/07/17/1005396/predictive-policing-algorithms-racist-dismantled-machine-learning-bias-criminal-justice/