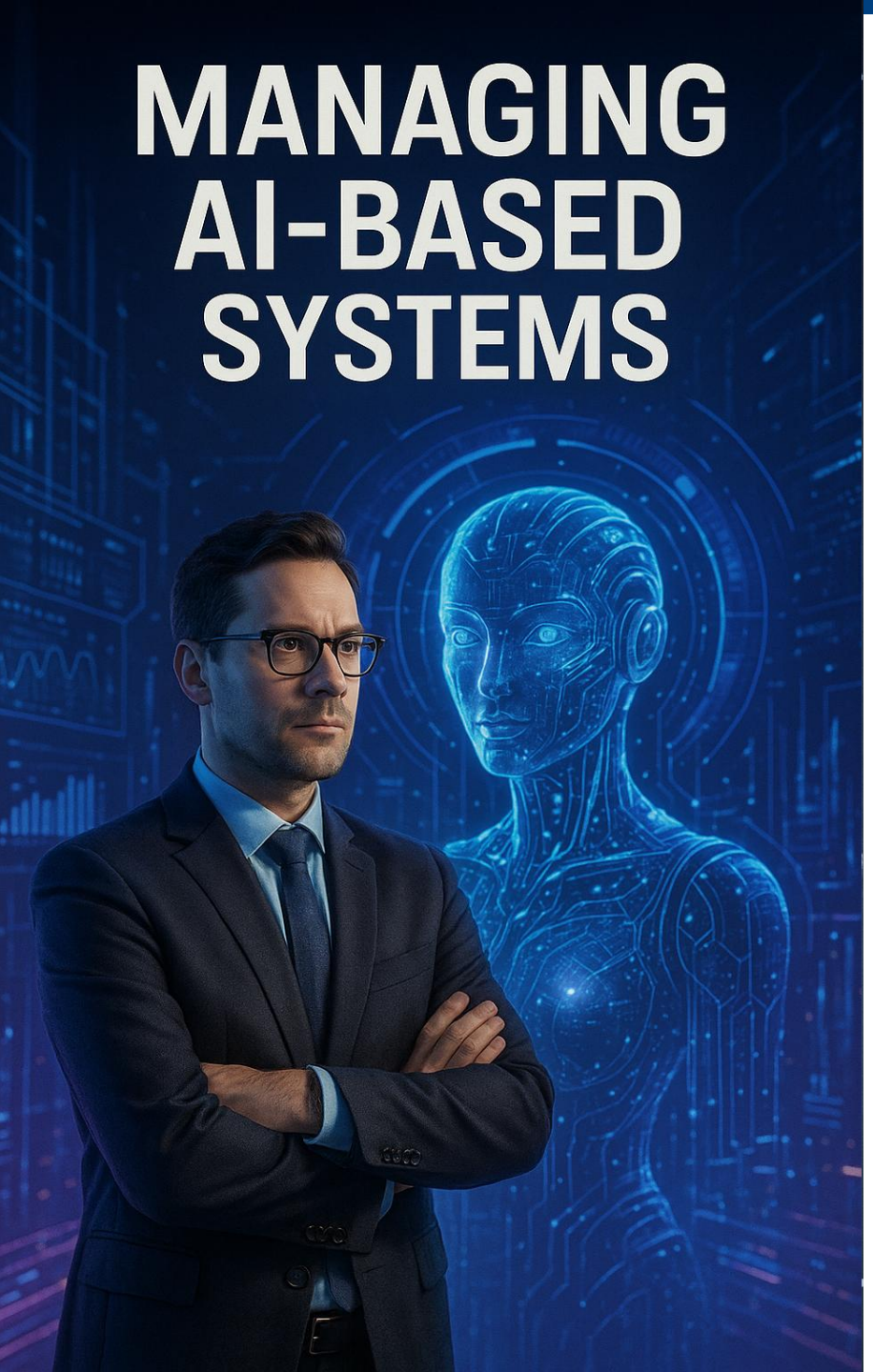


Session 11: Human-AI collaboration

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

This work is licensed under CC BY-NC-SA 4.0. To view a copy of this license, visit:
https://creativecommons.org/licenses/by-nc-sa/4.0/

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

9 Architectures of AI applications

10 Data Management and Model Transparency

11 **Design of human-AI interactions**

AI Operations at Scale

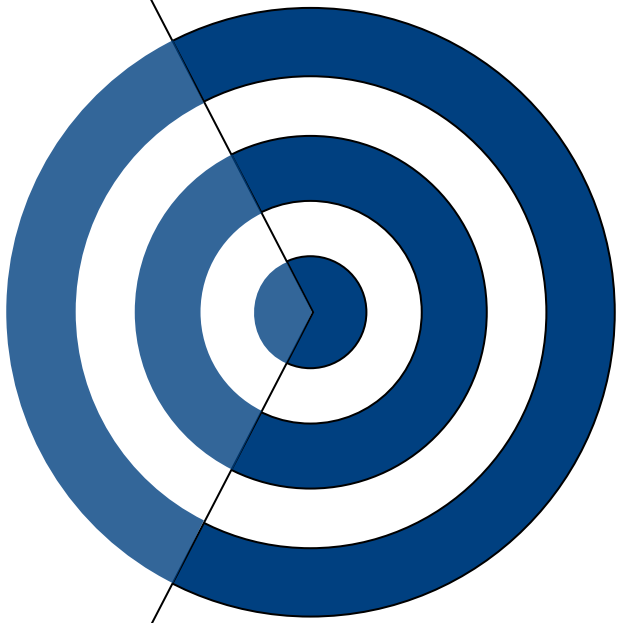
12 Monitoring and KPI-based control

13 Ethical, legal and social implications of AI

Implementation

Objectives of today's lecture

1. Understand the concept of human-AI collaboration and its practical relevance
2. Comprehend the different factors enhancing acceptance of AI
3. Discover the issues arising from human-AI collaboration and possible solutions



Will AI replace humans in their jobs?

(1) Goldman Sachs (2023), (2) Forbes (2023), (3) McKinsey (2023), (4) Continental (2023)

Types of intelligence

Mechanical

- Minimal degree of learning or adaption
- Precise, consistent, and efficient
- For example, self-service technologies and service robots
- Rely on observations to act and react repetitively
- Skills that require limited training or education
- Call center agents, retail salespersons, waiters/ waitress, and taxi drivers

Analytical

- Learns and adapts systematically based on data
- Logical, analytical, and rule-based learning
- For example, IBM's chess player Deep Blue
- Rational decision-making
- Technical skills requiring training and expertise on data and analysis
- Technology-related workers, data scientists, accountants, financial analyst, auto service technicians, and engineers

Intuitive

- Learns and adapts intuitively based on understanding
- Artificial neural networks-based or statistical-based deep learning
- For example, Watson's Jeopardy, Google's DeepMind AlphaGo, and AI poker player Libratus
- Hard thinking professionals that require creative thinking for problem-solving skills
- Marketing managers, management consultants, lawyers, doctors, sales managers, and senior travel agents

Empathic

- Learn and adapt empathetically based on experience
- Emotion recognition, affective computing, and communication style learning
- For example, Hanson's humanoid robot Sophia and chatbot Replika
- Soft empathetic professionals that require social, communication, and relationship building skills
- Thinking jobs requiring people skill, for example, politicians and negotiators or feeling jobs, for example, psychiatrists

Huang and Rust (2018)

Will AI replace humans in their jobs?

Huang and Rust (2018)

Wait a minute… what about emotional/ social intelligence?

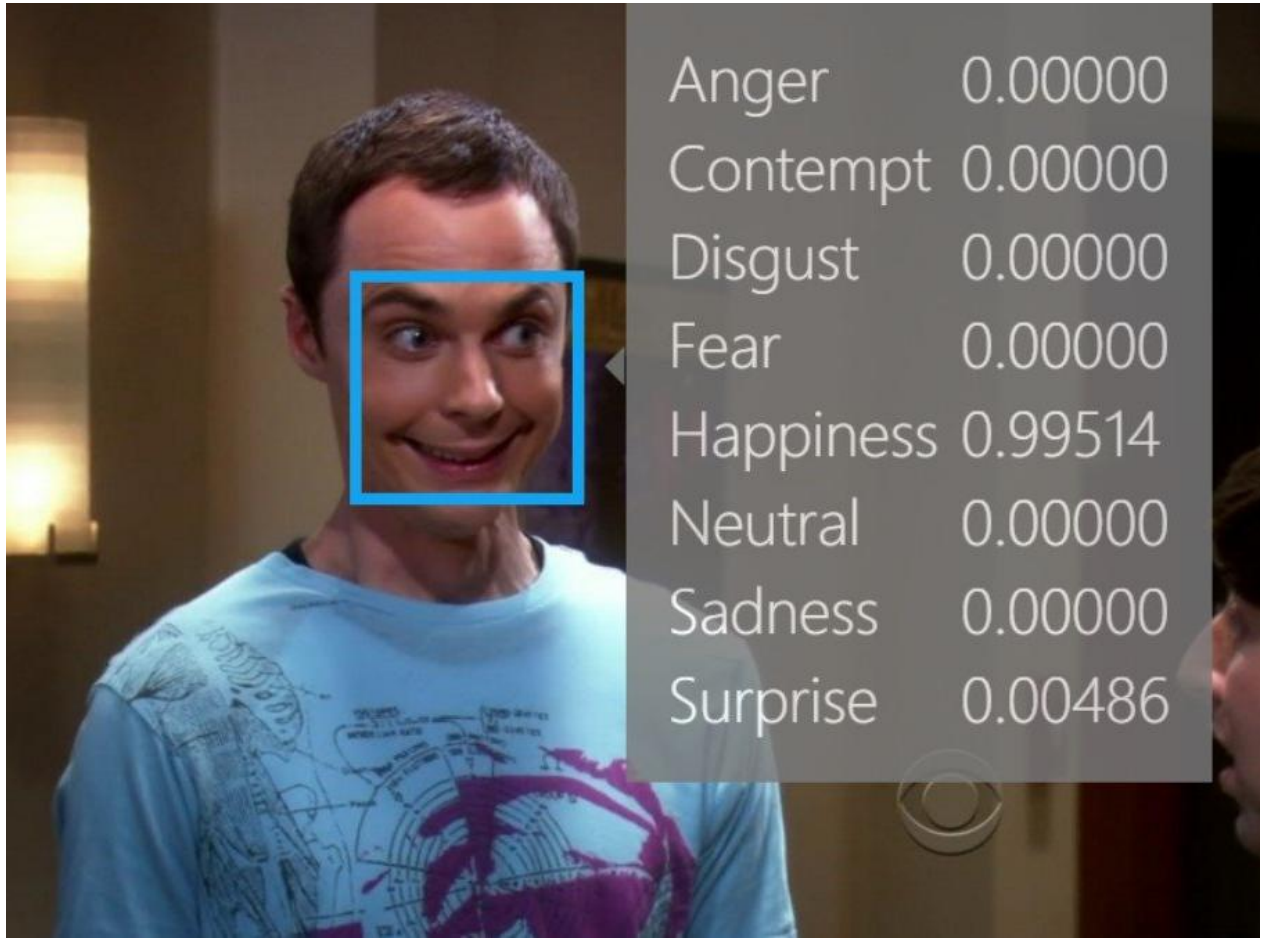
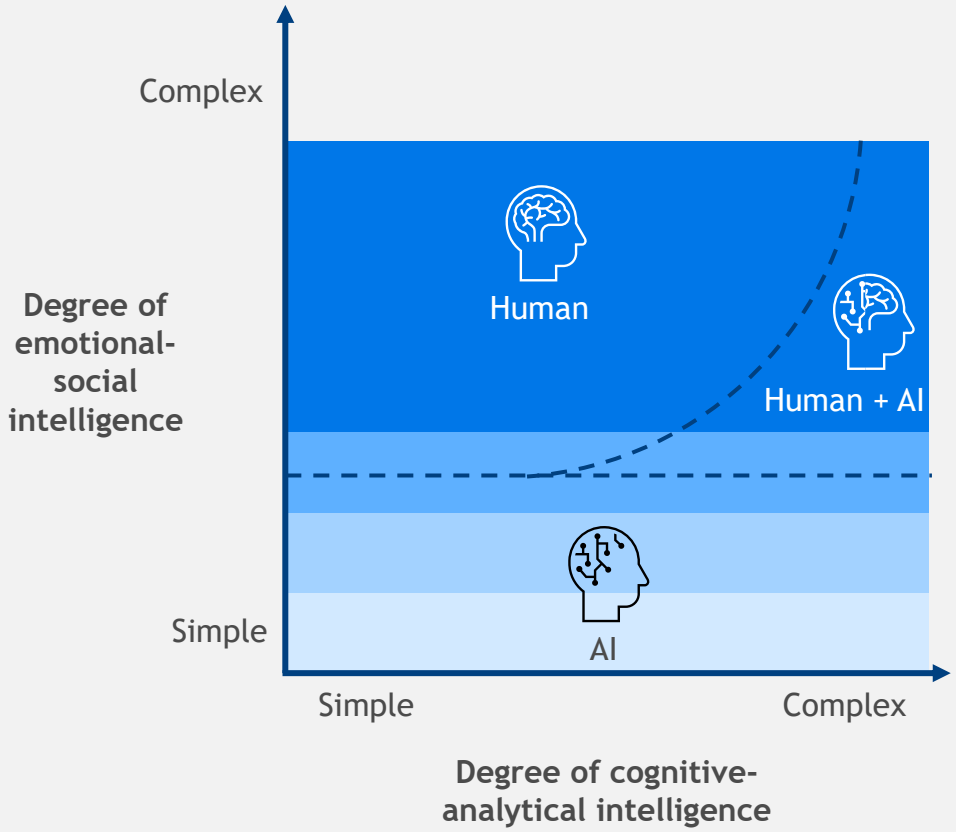


Image: knowyourmeme.com

A coexistence through collaboration is more realistic

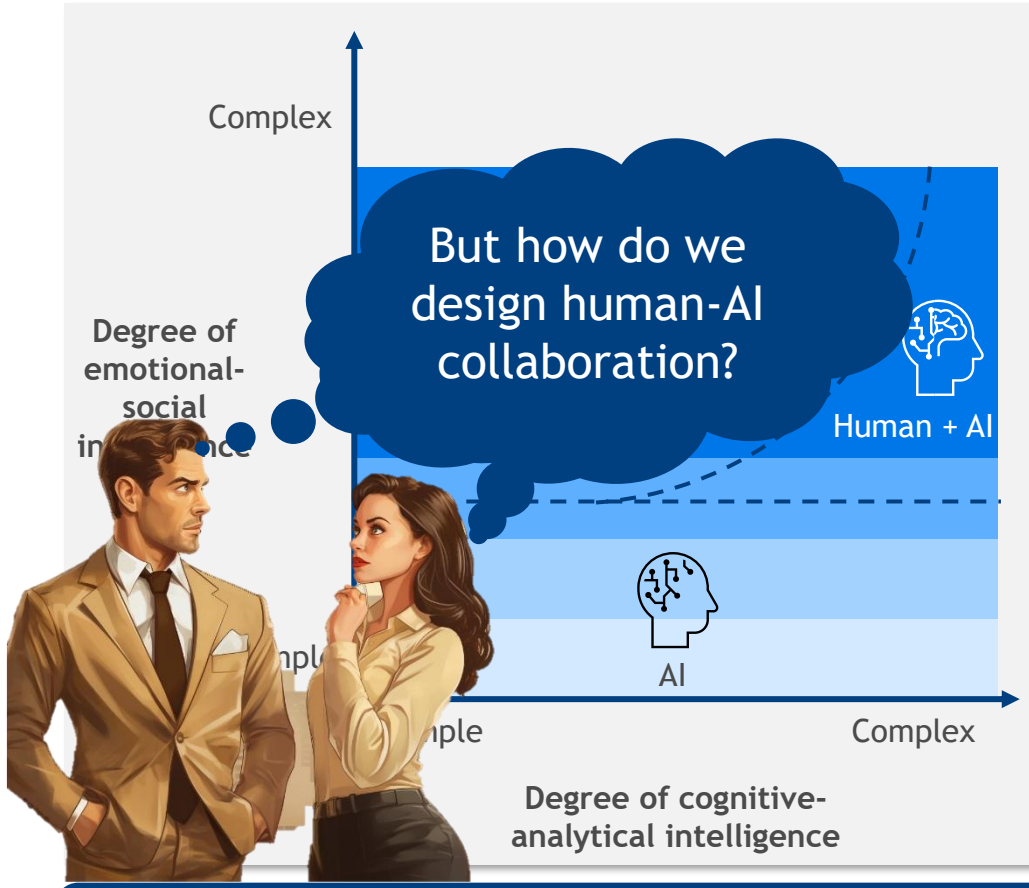


- **Empathetic intelligence:** understand and interact with emotions of humans
- **Intuitive intelligence:** creative thinking in new situations with required understanding (e.g., generative AIs like Midjourney or ChatGPT are weak forms of that)
- **Analytical intelligence:** relies on algorithmic learning from data (e.g., accounting)
- **Mechanical intelligence:** routine tasks (e.g., call center)

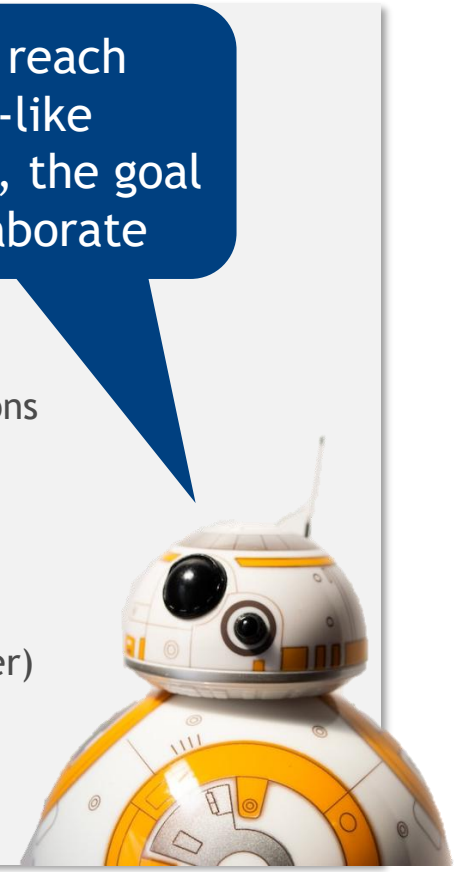
AI already has/ will probably surpass human-like intelligence in the first three levels, but it is still debatable if AI will ever achieve human-like empathetic intelligence

Wirtz et al. (2018), Huang & Rust (2018)

A coexistence through collaboration is more realistic



- **Empathetic intelligence:** understand and interact with emotions of humans
- **Intuitive intelligence:** creative thinking in new situations with required understanding (e.g. generative AIs like Midjourney or ChatGPT are weak forms of that)
- **Analytical intelligence:** relies on algorithmic learning from data (e.g., accounting)
- **Mechanical intelligence:** routine tasks (e.g., call center)



AI already has/ will probably surpass human-like intelligence in the first three levels, but it is still debatable if AI will ever achieve human-like empathetic intelligence

Wirtz et al. (2018), Huang & Rust (2018)

The goal of HAIC is to focus on the human's benefit

- Automating what humans don't want to do (e.g., transcribing interviews).
- Scaling what humans practically can't do (e.g., continuous health monitoring of patients).
- Performing what humans used to do at a "superhuman" level (e.g., in the future: autonomous driving)

Dimensions of HAIC

- Combine complementary capabilities
- Make decision models interpretable and insights explainable
- Ensure legal and ethical compliance

Yeung (2020)

Agenda

01 | Acceptance driving factors of AI

02 | Algorithmic Transparency & Explainability

03 | Complementarity

04 | Fairness

05 | Delegation

06 | Risk of Human-AI Collaboration

Agenda

01 | Acceptance driving factors of AI

02 | Algorithmic Transparency & Explainability

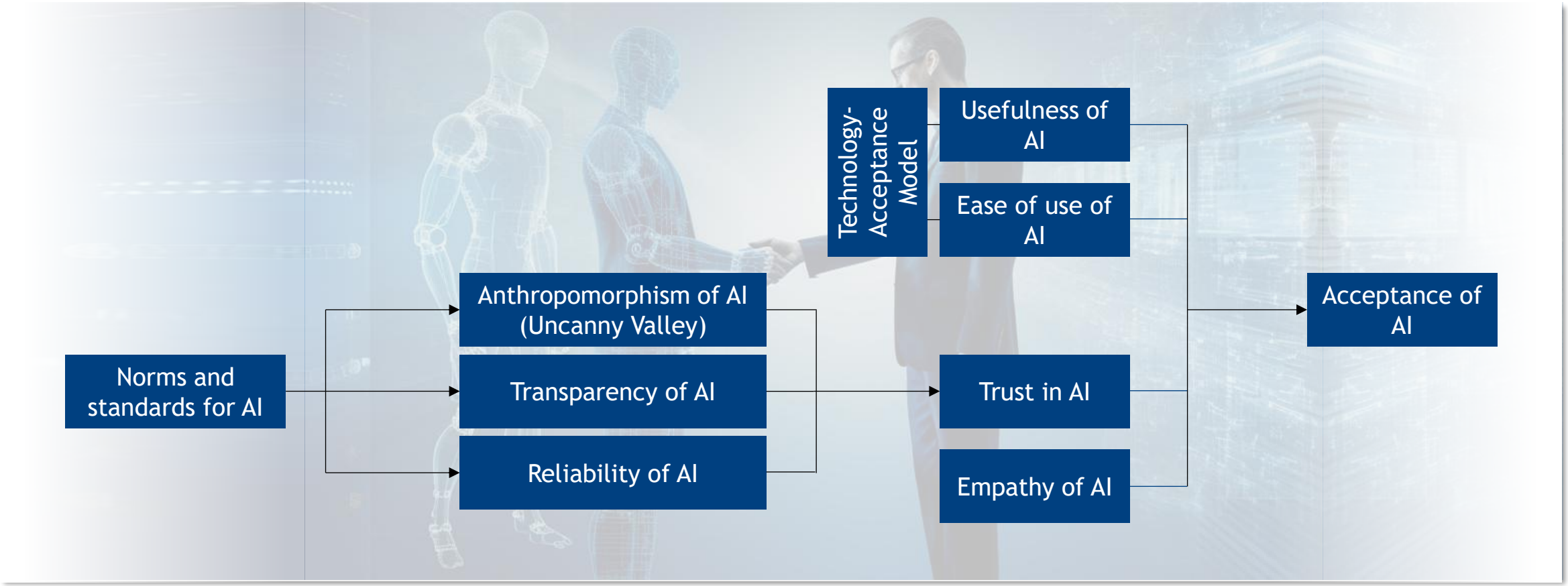
03 | Complementarity

04 | Fairness

05 | Delegation

06 | Risk of Human-AI Collaboration

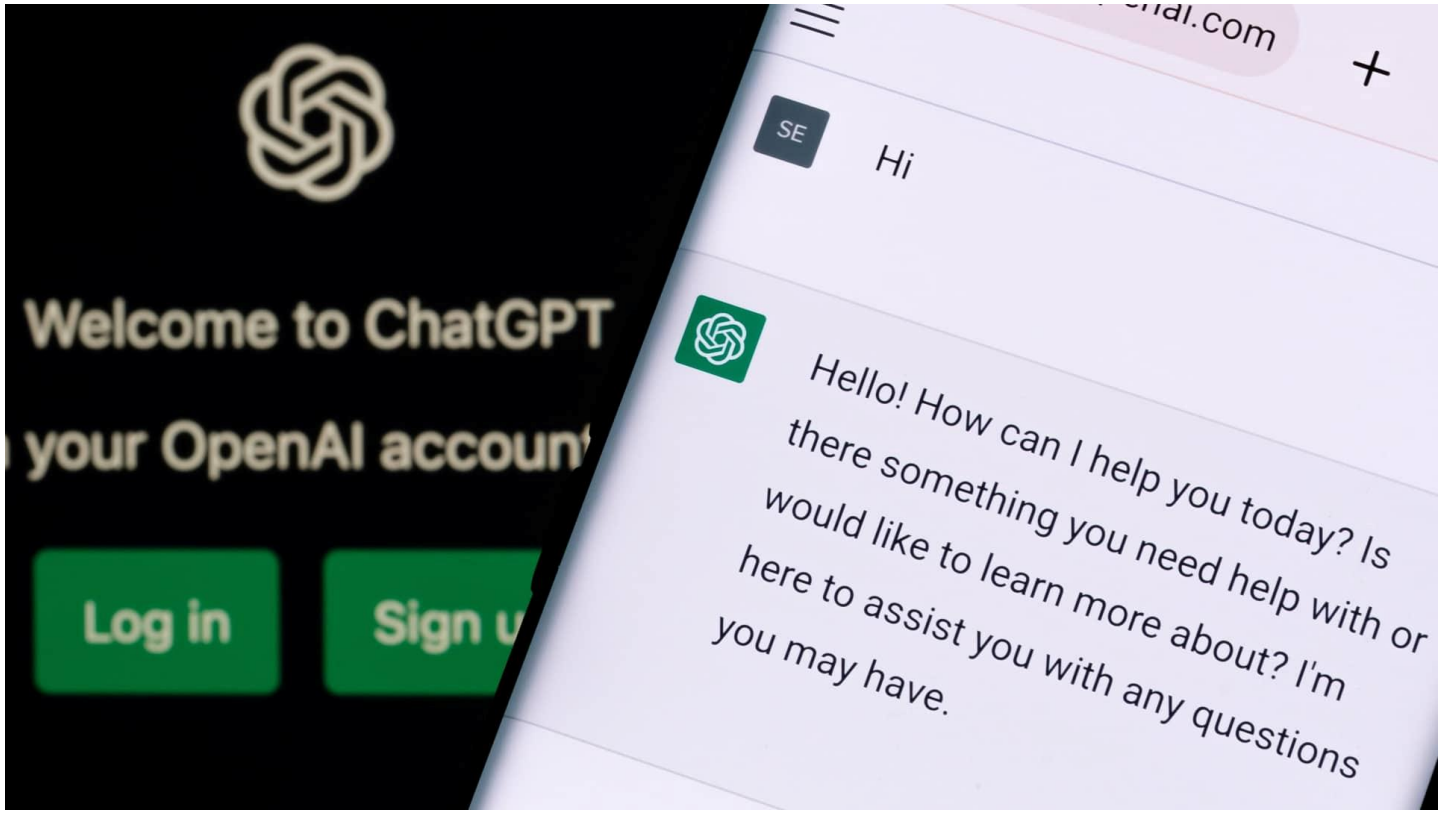
Acceptance driving factors for Human-AI Collaboration



Acceptance driving factors for AI have widely been researched and deliver a solid starting point for the design of Human-AI Collaboration

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

Which factors do we find in current AI-systems, like ChatGPT?



https://www.it-daily.net/it-sicherheit/cloud-security/chatgpt-ki-ist-jetzt-der-natuerlichen-ignoranz-gewachsen

- ✓ Anthropomorphism: ChatGPT is able to chat with you like a human (passes Turing Test)
- ✓ Usefulness: If you understand for which cases ChatGPT delivers value, it can be very useful
- ✓ Ease of use: It is very easy to use in its Chat User Interface
- ~ Transparency: Even if we somehow understand the heuristic (including temperature factor), OpenAI is not fully transparent on how it works, on which data it was trained, etc.
- ~ Reliability: Sometimes it produces non usable output, it is even reported that the system got worse over time

In the case of ChatGPT, the most important factors for its acceptance might be its usefulness, the ease of use and its ability to write human-like.

Papalias, 2023

Agenda

01 | Acceptance driving factors of AI

02 | **Algorithmic Transparency & Explainability**

03 | Complementarity

04 | Fairness

05 | Delegation

06 | Risk of Human-AI Collaboration

Algorithmic transparency

Algorithmic transparency means to disclose information about algorithms to enable monitoring, criticizing or intervening by interested parties (Diakopoulos & Koliska, 2017)

Benefits:

- People, who use, regulate or are somehow affected by AI algorithms, can understand the underlying concepts of the algorithm because of its transparency
- With that, mistakes of an algorithmic system can also be made more traceable
- The accountability of mistakes still belongs to the developer of the technology

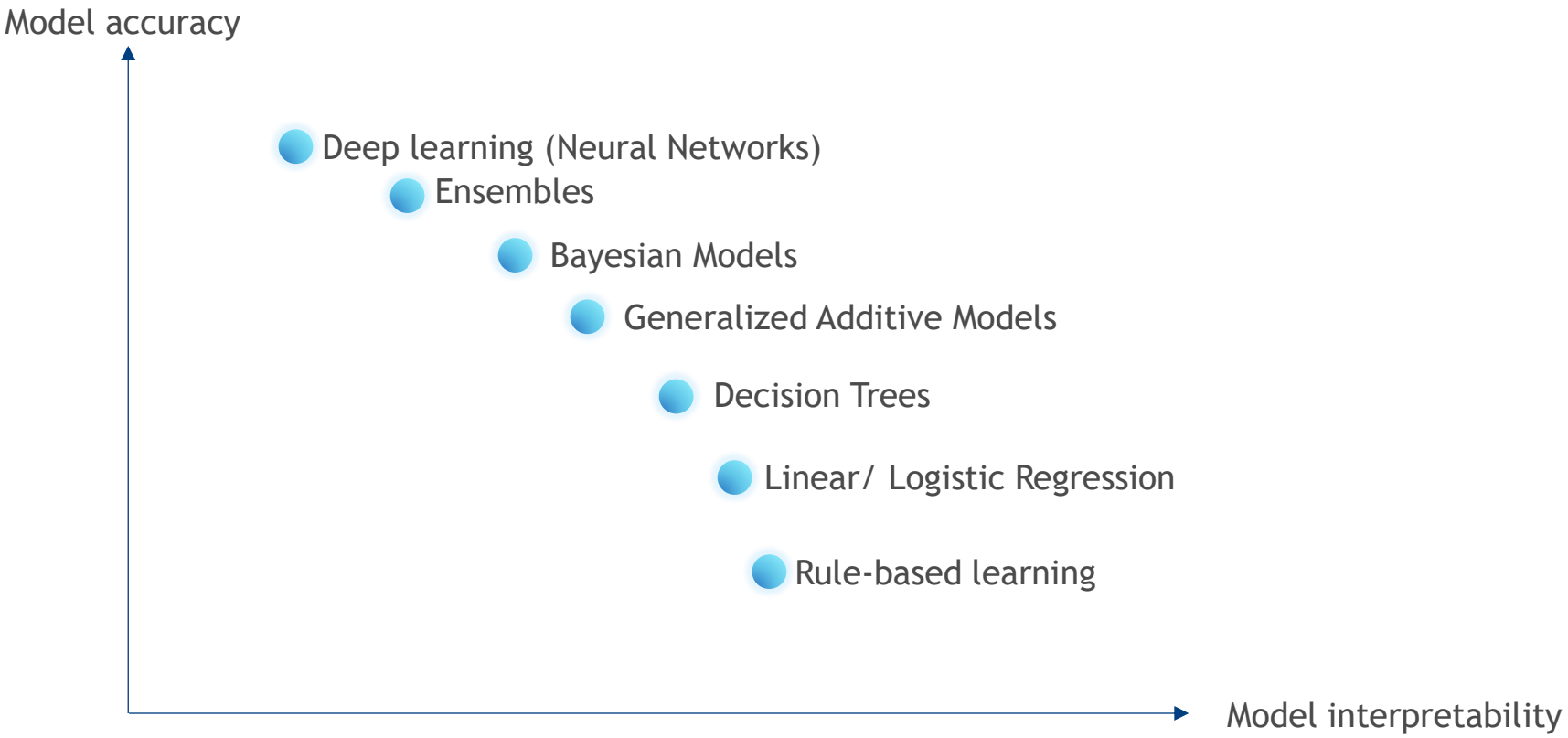
Components of algorithmic transparency:

1. Human involvement (Who is accountable, goal of algorithm, who has control, …)
2. Data (Quality of data, Data (pre)-processing, …)
3. Model (Features and variables used, assumptions behind the model, …)
4. Inference (Accuracy rates, biases, …)
5. Personalization (Types of personal information used, …)

Algorithmic transparency is an important component of Human-AI Collaboration because it sets a base for the explainability of AI

Diakopoulos & Koliska, (2017), Kossow et al. (2021)

Relationship between model interpretability and accuracy



There is a certain relationship between different models' interpretability and their accuracy, which shows why algorithmic transparency and explainability gets more important

Arrieta et al. (2020)

Why the explainability of AI matters?

Explanations of intelligent systems are important to understand

…how and why…

it arrived at a particular decision.

This is especially crucial in scenarios where AI makes decision that affect people directly, like in…

Finance

Health care

Autonomous driving

Legal affairs

Benefits of AI explainability

- Building trust and confidence (e.g., for decision makers)
- Helping to identify bias and error (e.g., for developers)
- Verifying alignment with goals and values (e.g., regulatory agencies)
- Helping to understand decisions to end-users (e.g., loan applicants)

By offering explanations, the black box nature of many AI algorithms becomes more trustworthy, accountable and user-friendly

explainml-tutorial.github.io

Why different groups of people need different explanations?

Decision-makers

- Justifying the reliability of AI model
- Evaluating individual predictions of AI models
- Provide additional information, from which decision maker can learn from

Developers

- Important to understand models from a technical standpoint
- Use for quality assurance purpose
- Support debugging
- Can help resolve failures and improve performance

Regulators

- Crucial component of trustworthy AI
- Helps ensuring that AI involve aspects of fairness, safety, privacy and accountability
- Auditing of AI products

End-users

- Help to understand AI's decisions and how to change the AI's decision
- Means to create trust

The explainability of AI is subjective to the stakeholder that is receiving it and thus fulfills different benefits

Tjoa & Guan (2020), Vokinger & Gasser (2021), Laato et al. (2022)

Agenda

01 | Acceptance driving factors of AI

02 | Algorithmic Transparency & Explainability

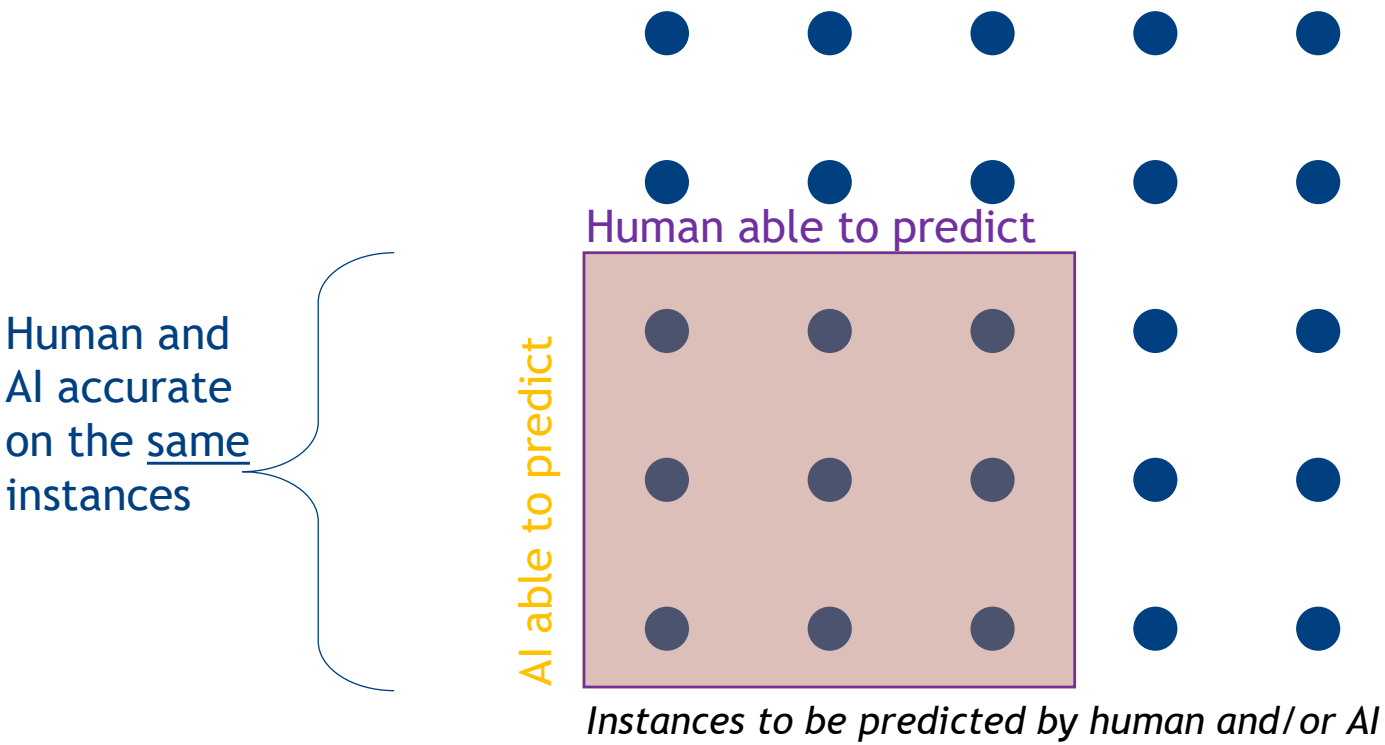
03 | Complementarity

04 | Fairness

05 | Delegation

06 | Risk of Human-AI Collaboration

…as well as the distribution of the individual strengths – the complementary potential (1/3).

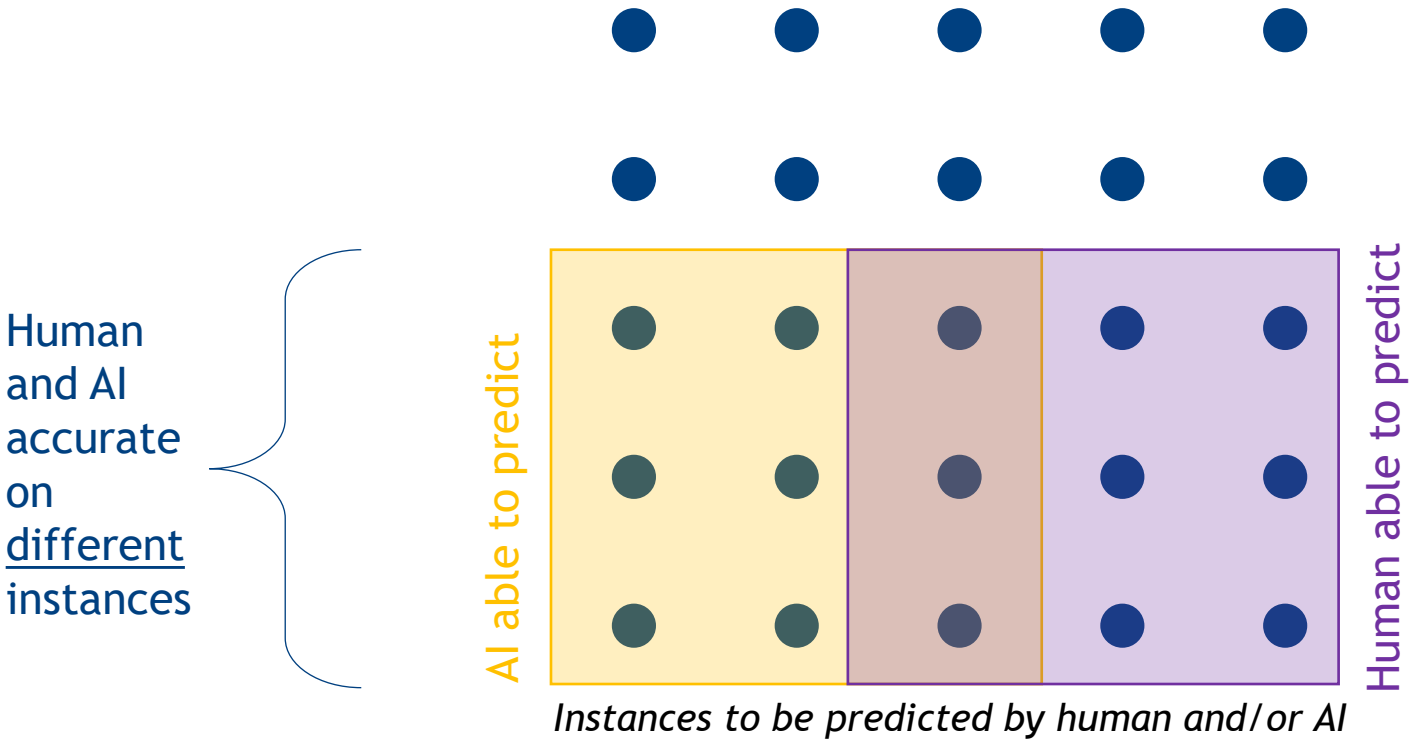


= No Complementary Potential

→ No Complementary Team Performance possible

Hemmer et al. (2022)

…as well as the distribution of the individual strengths – the complementary potential (2/3).

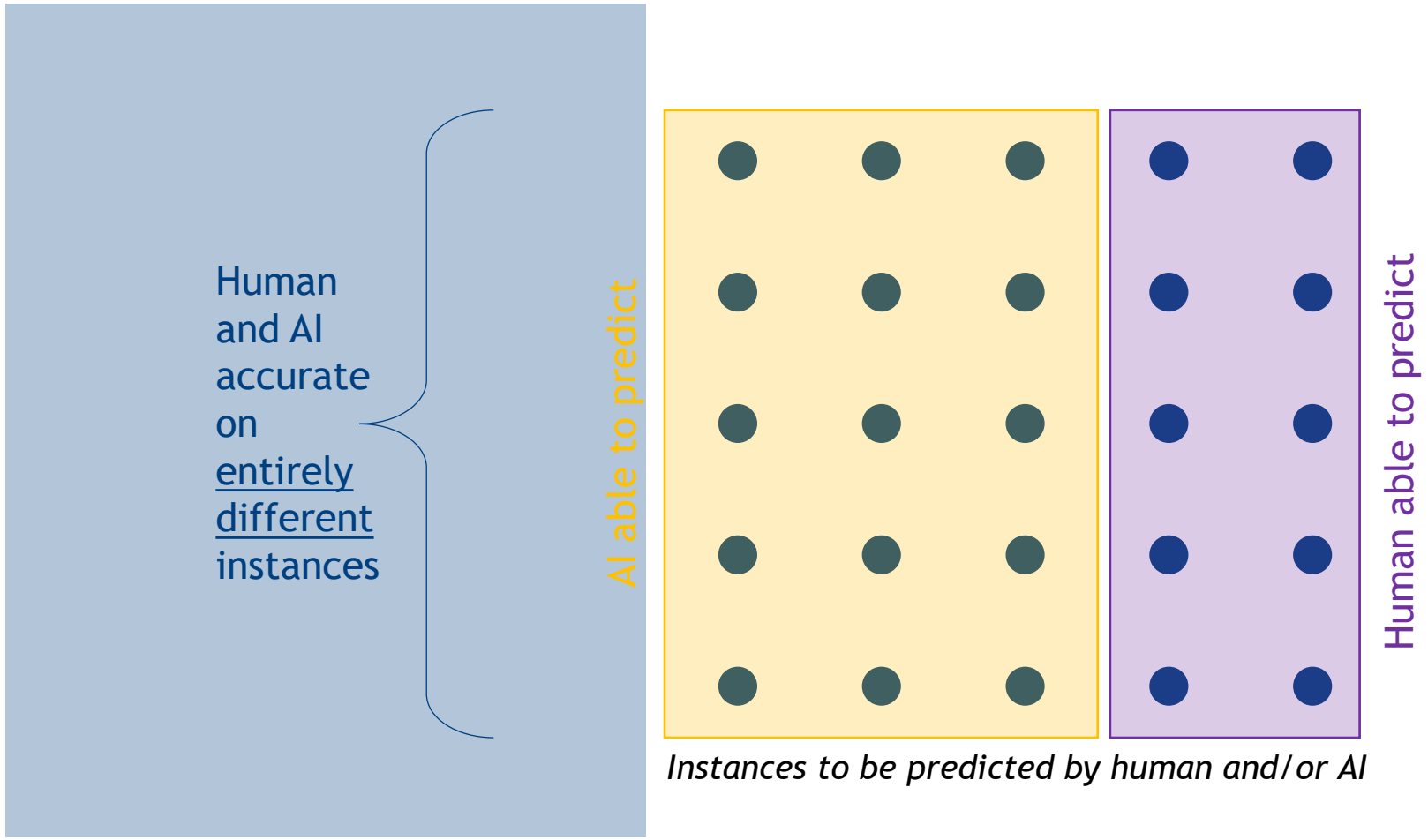


= Complementary Potential

→ Complementary Team Performance possible

Hemmer et al. (2022)

…as well as the distribution of the individual strengths – the complementary potential (3/3).

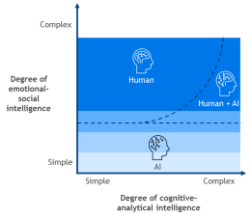


= Maximum Complementary Potential

→ Complementary Team Performance possible

Hemmer et al. (2022)

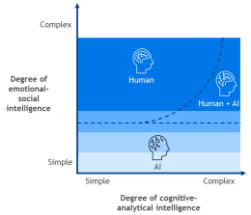
Complementarity in prominent AI use cases

Use case	Autonomous driving	Robo-financial - advisor	Cancer detection	Military services	Conclusion
					
Task AI	▪ Steering & Navigation of car according to **rules, signs** and environmental conditions	▪ Portfolio Allocation ▪ **Automated** Rebalancing ▪ Risk Assessment ▪ Tax Efficiency ▪ Cost Efficiency	▪ Analysis of big databases ▪ **Pattern** Recognition ▪ Early Detection ▪ **Continuous** Monitoring	▪ **Persistent** Monitoring ▪ Terrain Mapping ▪ Early Warning ▪ Military services in **dangerous** areas	**Tasks characteristics:** ▪ Continuous ▪ Analytical task for big database ▪ Pattern Recognition ▪ Routine ▪ Dangerous
Task Human	▪ **Monitoring** of AI ▪ **Intervening** in emergency situations ▪ **Ethical** decisions	▪ **Psychological** components of investing ▪ **Behavioral** Coaching ▪ Personalized advice	▪ Clinical Judgement ▪ **Contextual** Understanding ▪ **Communication** ▪ **Ethical** and **cultural** factors	▪ **Strategic** decision with **psychological** warfare ▪ **Communication** and Coordination of staff and arti. units	**Tasks characteristics:** ▪ Monitoring AI ▪ Intervening in case of specialty ▪ Ethical, cultural and psychological topics ▪ Communicative

Across various industries, the complementary collaboration of human and AI can lead to better overall performance

Images: Midjourney

Complementarity in cancer detection

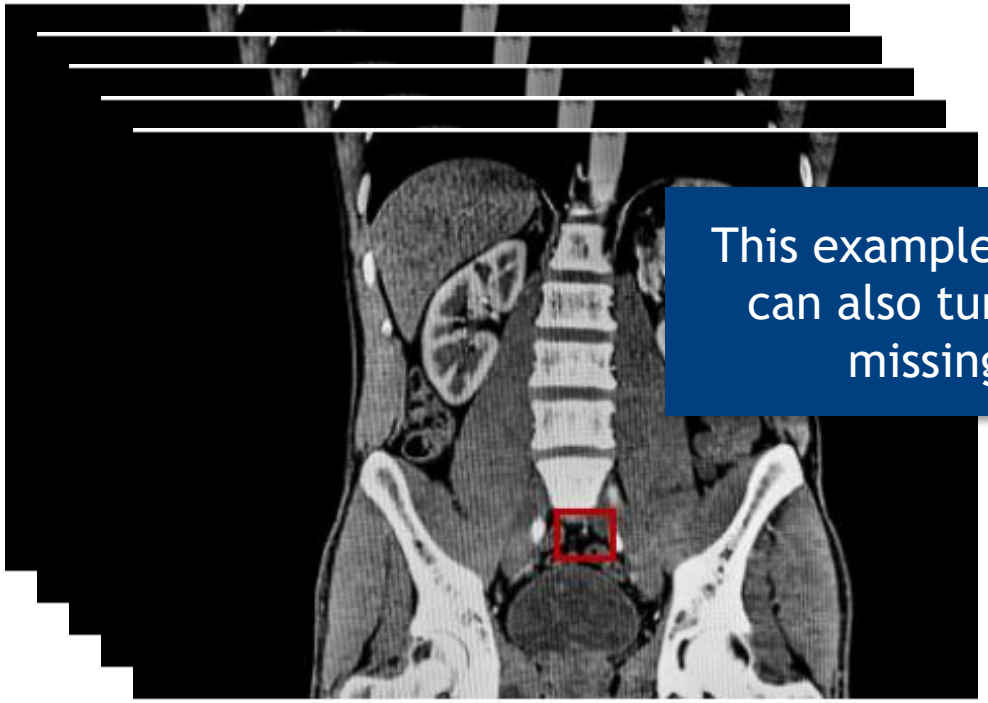
	Autonomous driving	Robo-financial - advisor	Cancer detection	Military services	Conclusion
Use case					
Task AI	▪ Steering & Navigation of car according to **rules, signs** and environmental conditions	▪ Portfolio Allocation ▪ **Automated** Rebalancing ▪ Risk Assessment ▪ Tax Efficiency ▪ Cost Efficiency	▪ Analysis of big databases ▪ **Pat** ▪ …oring	▪ **…ersistent** Monitoring ▪ …rrain Mapping ▪ …ly Warning ▪ Military services in **dangerous** areas	**Tasks characteristics:** ▪ Continuous ▪ Analytical task for big database ▪ Pattern Recognition ▪ Routine ▪ Dangerous
Task Human	▪ **Monitoring** of AI ▪ **Intervening** in emergency situations ▪ **Ethical** decisions	▪ **Psychological** components of investing ▪ **Behavioral** Coaching ▪ Personalized advice	▪ …ical Judgement ▪ **Contextual** Understanding ▪ **Communication** ▪ **Ethical** and **cultural** factors	▪ **Strategic** decision with **psychological** warfare ▪ **Communication** and Coordination of staff and arti. units	**Tasks characteristics:** ▪ Monitoring AI ▪ Intervening in case of specialty ▪ Ethical, cultural and psychological topics ▪ Communicative

Are you sure an AI can't replace a doctor?

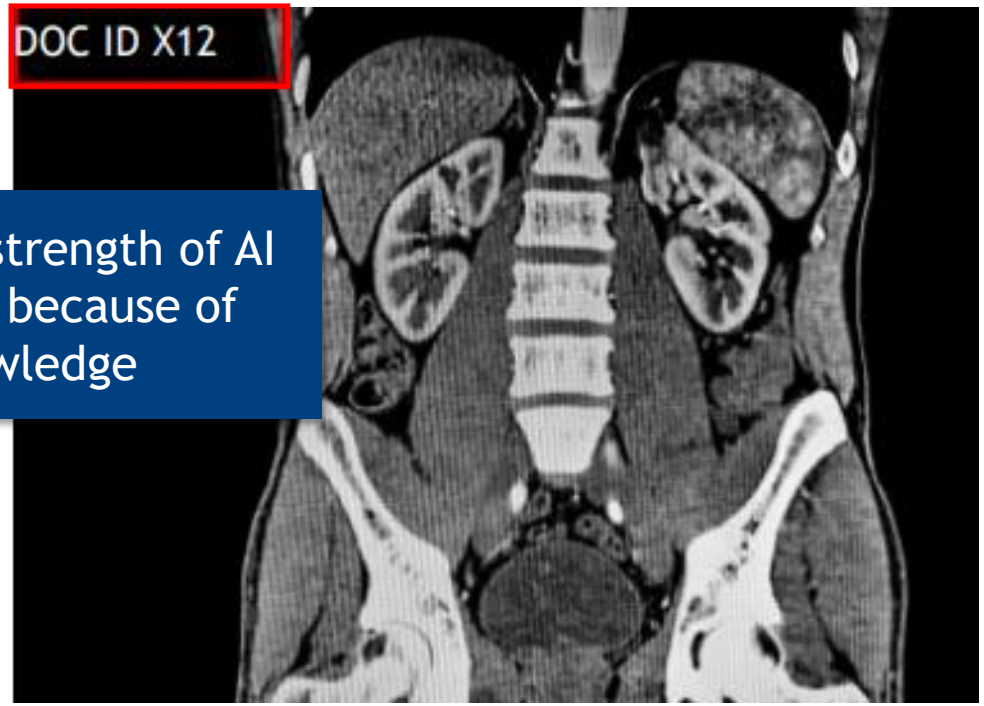
Across various industries, the complementary collaboration of human and AI can lead to better overall performance

Images: Midjounrey

Why Human-AI Collaboration needs to be complementary



This example shows how the strength of AI can also turn into a mistake because of missing contextual knowledge

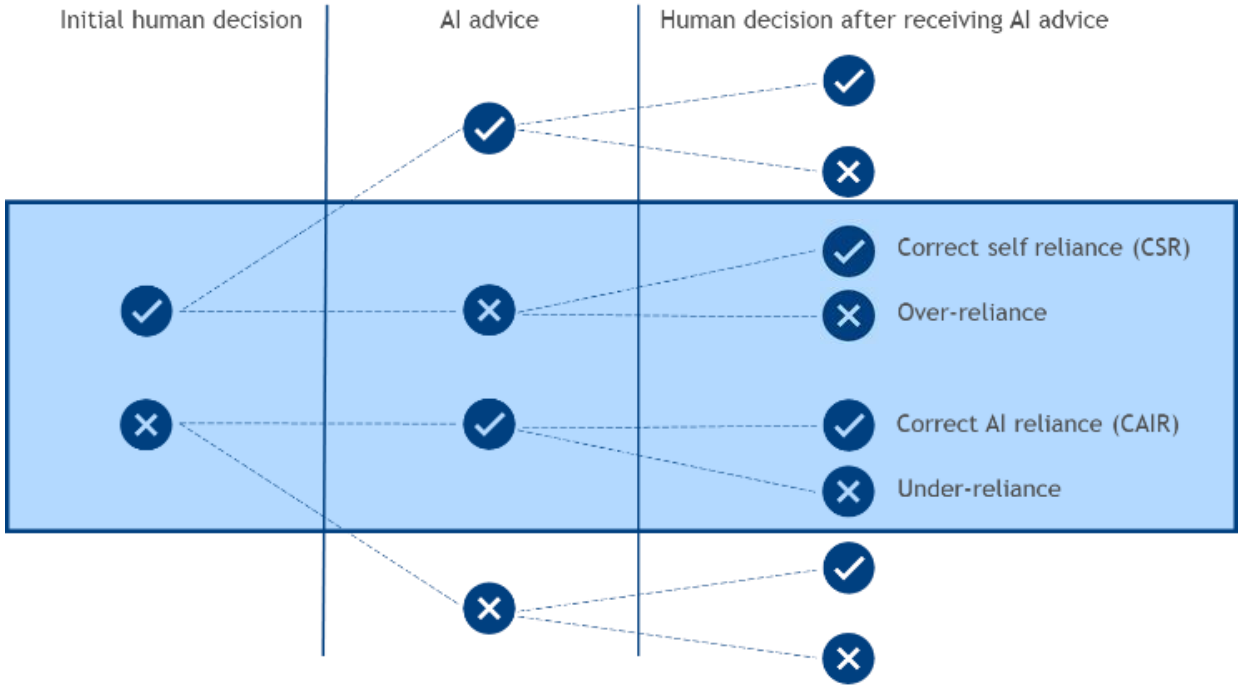


AI system finds anomaly by analyzing big data base of patients.

AI system detects picture numbering as an anomaly and reports it.

Kühl (2022)

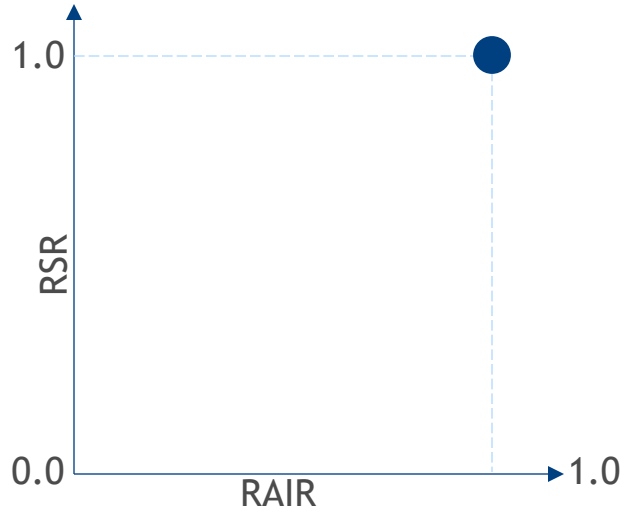
Human-AI Collaboration strives for optimal Appropriateness of Reliance (AoR)



Appropriateness of Reliance (AoR) = (RSR, RAIR)

$$RAIR = \frac{\sum_{i=0}^{N} CAIR_i}{\sum_{i=0}^{N} CA_i} \qquad RSR = \frac{\sum_{i=0}^{N} CSR_i}{\sum_{i=0}^{N} IA_i}$$

CA (Correct advice), IA (Incorrect advice)



When human and AI come to a different conclusion, over- and under-reliance are the two possible types of errors that can occur

Schemmer et al. (2023)

Agenda

01 | Acceptance driving factors of AI

02 | Algorithmic Transparency & Explainability

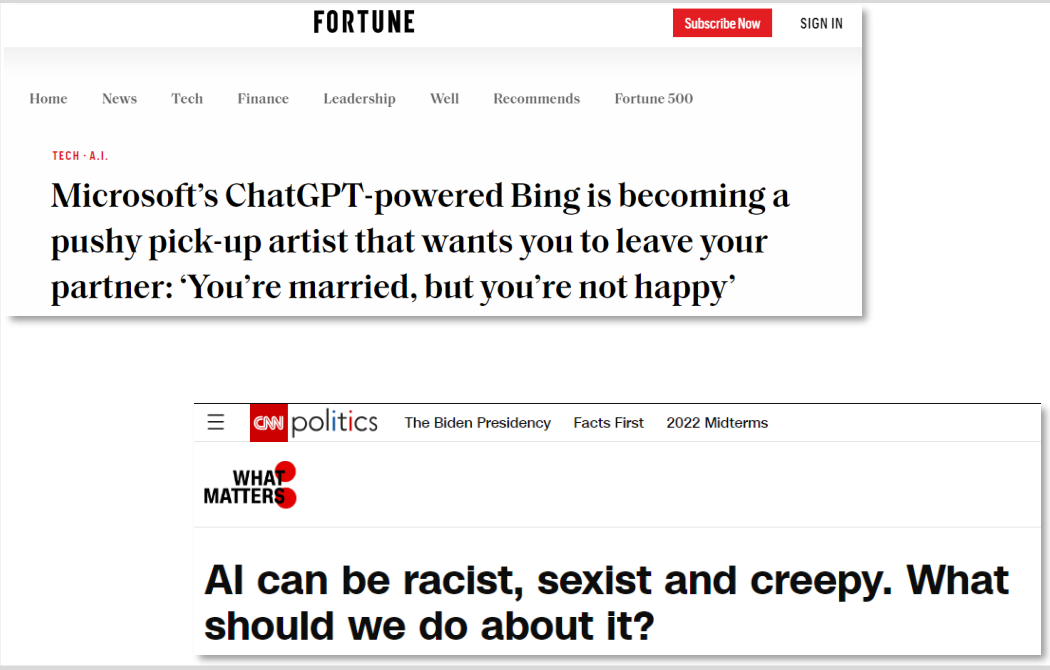
03 | Complementarity

04 | Fairness

05 | Delegation

06 | Risk of Human-AI Collaboration

Some AI systems differentiate based on unjustified data



AI's task is often to find patterns and to differentiate

BUT: sometimes these differentiations are discriminations because they are based on unjustified/ irrelevant information

- AI systems have no conscience and will take the "path of least resistance"

The topic of fairness of AI based decisions also shows the relevance of complementarity to ensure that the differentiating variables are relevant and justified

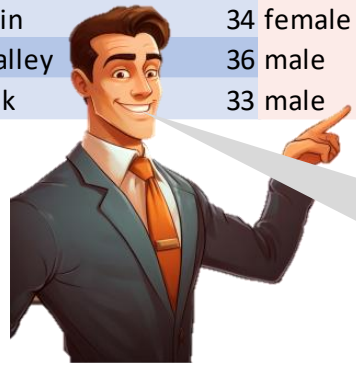
Fortune (2023), CNN (2023), Ribeiro et al. (2016)

Excluding certain data does not always lead to fairness

- Some sensitive information can be reproduced from other features (e.g., gender and name, height, length of hair, etc.)
  - ➢ Excluding gender for a job application does not automatically solve the inequality issue because other features than gender could lead to the same conclusion (e.g., name)
- The decision when to in- and exclude certain sensitive information for AI based decision has to be evaluated by the society for each case

Job applicant for management position

Surname	Last name	Age	Gender	Job last year	University	High school
Anna	Griffin	34	female	Maternity protection	Yale	Girl's school Michigan
Peter	O'Malley	36	male	AI Consultant	TUM	New jersey high school
Brad	Kaslik	33	male	CEO start-up	Cambridge	Alabama high school



We, as an inclusive, diverse company use machine learning for the first selection of our application process but exclude gender data of our applicants to ensure equality!

The yellow highlighted cells could also lead to a similar decision to exclude the woman for the application

Omitting certain sensitive information does not automatically lead to more equality and fairness because systems can conclude similar based on other variables (redundant coding)

Ribeiro et al. (2016)

What exactly does fairness mean to you?

When is an ML algorithm fair and how can we achieve this?

Remember what we just learned about redundant encodings.

There are different approaches to achieve fairness in AI

I. **Statistical parity: "Acceptance rate is equal for all sensitive groups"**
   - Example: 70% females and 30% males apply to university
   - Statistical parity would mean that there are also 70% females and 30% males in the pool of admitted students

   I. Question to you: Does this prevent unfair outcomes?

II. **Equal opportunity: "True positive rates are equal for all sensitive groups"**
   - Example: Out of 40 qualified female applicants, 30 were correctly accepted.
   - Equal opportunity would mean that ¾ of male applicants would also have been accepted correctly

   II. Question to you: Does this prevent unfair outcomes?

Which approach do you think is fairer ? (No right or wrong)

		Ground Truth: Positive	Ground Truth: Negative
Predictions	Positive	True Positive (TP)	False Positive (FP)
Predictions	Negative	False Negative (FN)	True Negative (TN)

Each approach has its own strength and weakness and is thus dependent on the case itself, which again shows the importance of Human-AI collaboration to monitor mistakes

Barocas et al. (2019), Hardt et al. (2016)

How to achieve algorithmic fairness better

a. **Pre-Processing**
   - Edit (training) data through trying to remove underlying discrimination
   - Example: Sampling of representative data with ImageNet

b. **In-Processing**
   - Modifying algorithms to eliminate discrimination during training
   - Either objective function is modified or additional constraints get introduced

c. **Post-Processing**
   - Re-assignment of labels after training based on fairness criteria
   - Mostly only the case if data and ML algorithm cannot be adapted

Even when trying to be fairer, absolute fairness is hard to achieve and likely comes with a cost in accuracy

Yang et al. (2020), Haas (2019), Rodolfa et al. (2020)

7 principles of the European Commission for trustworthy AI

Principles		Definitions in Ethics Guideline	Implication based on lecture topics
1) Human agency and oversight	1	Human agency and oversight: AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights. At the same time, proper oversight mechanisms need to be ensured, which can be achieved through human-in-the-loop, human-on-the-loop, and human-in-command approaches	European commission focuses on Human-AI collaboration
2) Technical robustness and safety	2	Technical Robustness and safety: AI systems need to be resilient and secure. They need to be safe, ensuring a fall back plan in case something goes wrong, as well as being accurate, reliable and reproducible. That is the only way to ensure that also unintentional harm can be minimized and prevented.	Reliability and Reproducibility (connected to transparency) is important
3) Privacy and data governance	3	Privacy and data governance: besides ensuring full respect for privacy and data protection, adequate data governance mechanisms must also be ensured, taking into account the quality and integrity of the data, and ensuring legitimised access to data.	Careful handling of sensitive data (fairness)
4) Transparency	4	Transparency: the data, system and AI business models should be transparent. Traceability mechanisms can help achieving this. Moreover, AI systems and their decisions should be explained in a manner adapted to the stakeholder concerned. Humans need to be aware that they are interacting with an AI system, and must be informed of the system's capabilities and limitations.	Explainability and algorithmic transparency for every stakeholder
5) Diversity, non-discrimination and fairness	5	Diversity, non-discrimination and fairness: Unfair bias must be avoided, as it could could have multiple negative implications, from the marginalization of vulnerable groups, to the exacerbation of prejudice and discrimination. Fostering diversity, AI systems should be accessible to all, regardless of any disability, and involve relevant stakeholders throughout their entire life circle.	Fairness to every vulnerable group to ensure non-discrimination
6) Societal and environmental well-being	6	Societal and environmental well-being: AI systems should benefit all human beings, including future generations. It must hence be ensured that they are sustainable and environmentally friendly. Moreover, they should take into account the environment, including other living beings, and their social and societal impact should be carefully considered.	Design of Human-AI Collaboration should be sustainable & environmentally friendly
7) Accountability	7	Accountability: Mechanisms should be put in place to ensure responsibility and accountability for AI systems and their outcomes. Auditability, which enables the assessment of algorithms, data and design processes plays a key role therein, especially in critical applications. Moreover, adequate an accessible redress should be ensured.	Responsibility and accountability are crucial topics for HAIC

https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Agenda

01 | Acceptance driving factors of AI

02 | Algorithmic Transparency & Explainability

03 | Complementarity

04 | Fairness

05 | Delegation

06 | Risk of Human-AI Collaboration

Different types of integration of Human and AI predictions

Integrator	Timing of integration: Ex-ante (i.e., before tasks are conducted)	Timing of integration: Ex-post (i.e., after tasks are conducted)
Human	Human-led delegation	Human integration
AI	AI-led delegation	AI integration

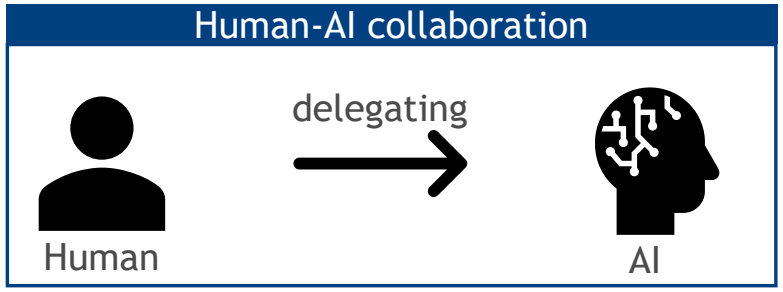
Depending on the integrator and the timing, human and AI predictions can be integrated differently:

- Human-led delegation: Human delegates task to either human or an AI
- AI-led delegation: AI delegates task to either human or AI
- Human integration: Human accepts/ rejects AI advice
- AI integration: AI combines predictions of humans and AI

The delegation of tasks in Human-AI Collaboration settings can not only come from humans, but also from AI

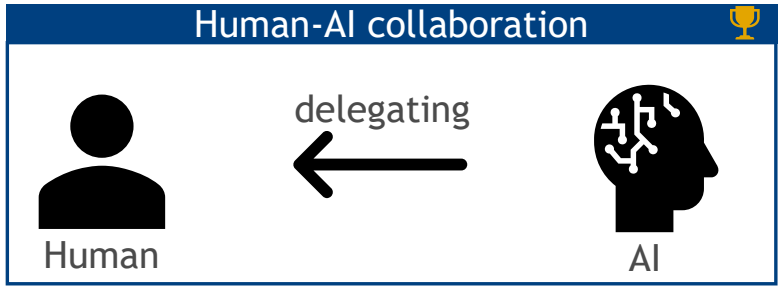
Kühl (2022), Leyer & Schneider (2019), Hemmer et al. (2022), Goodwin (2000)

Who is better at delegating – Human or AI? (in image classification)



Observations:

- Humans did not delegate well, even though they acted rationally and internally consistent and followed a clear delegation strategy
- BUT: Humans were not able to assess their own capabilities well, which then led to poor delegation decisions



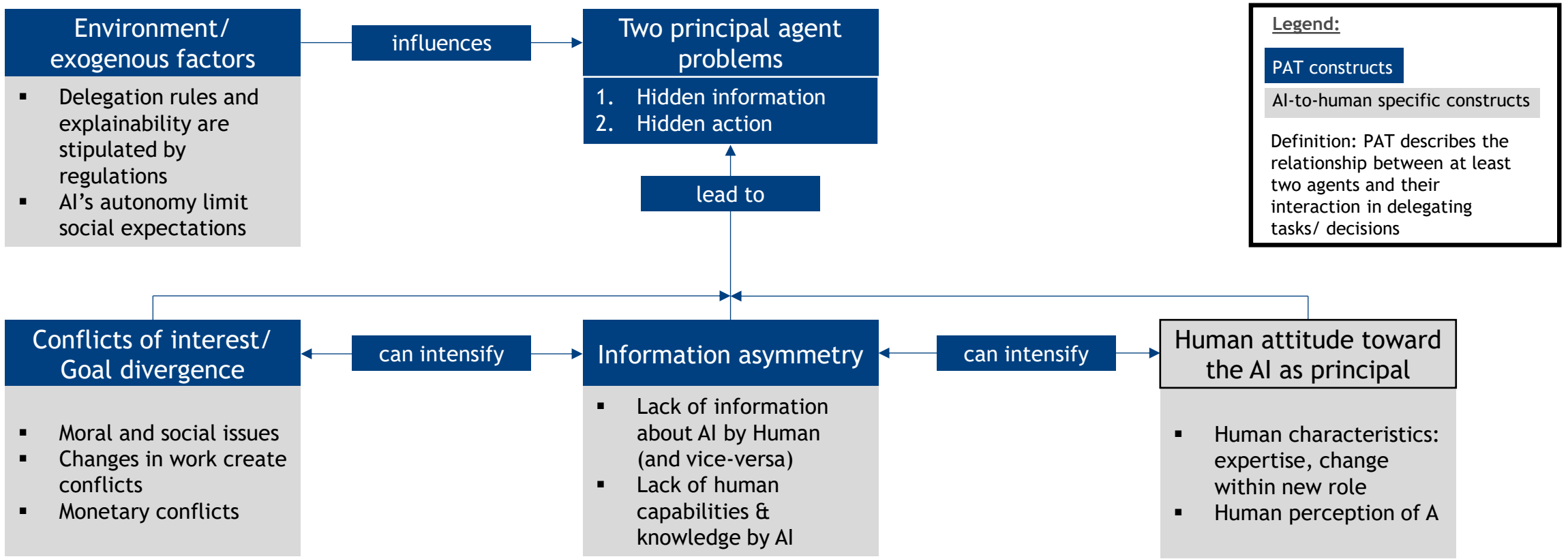
Observations:

- Collaboration with delegating AI has highest performance, which makes it economically desirable
- Humans could focus on the tougher classification examples, while AI did the mundane tasks by classifying easy picture
- AI would still not be a human's boss, rather an assistant
- BUT: This comes with a loss of control since AI only asks when it needs help

Current research has figured out that for image classification Human-AI collaboration outperforms AI (only) and Human (only) at work, but only when AI is delegating tasks

Fügener et al. (2022)

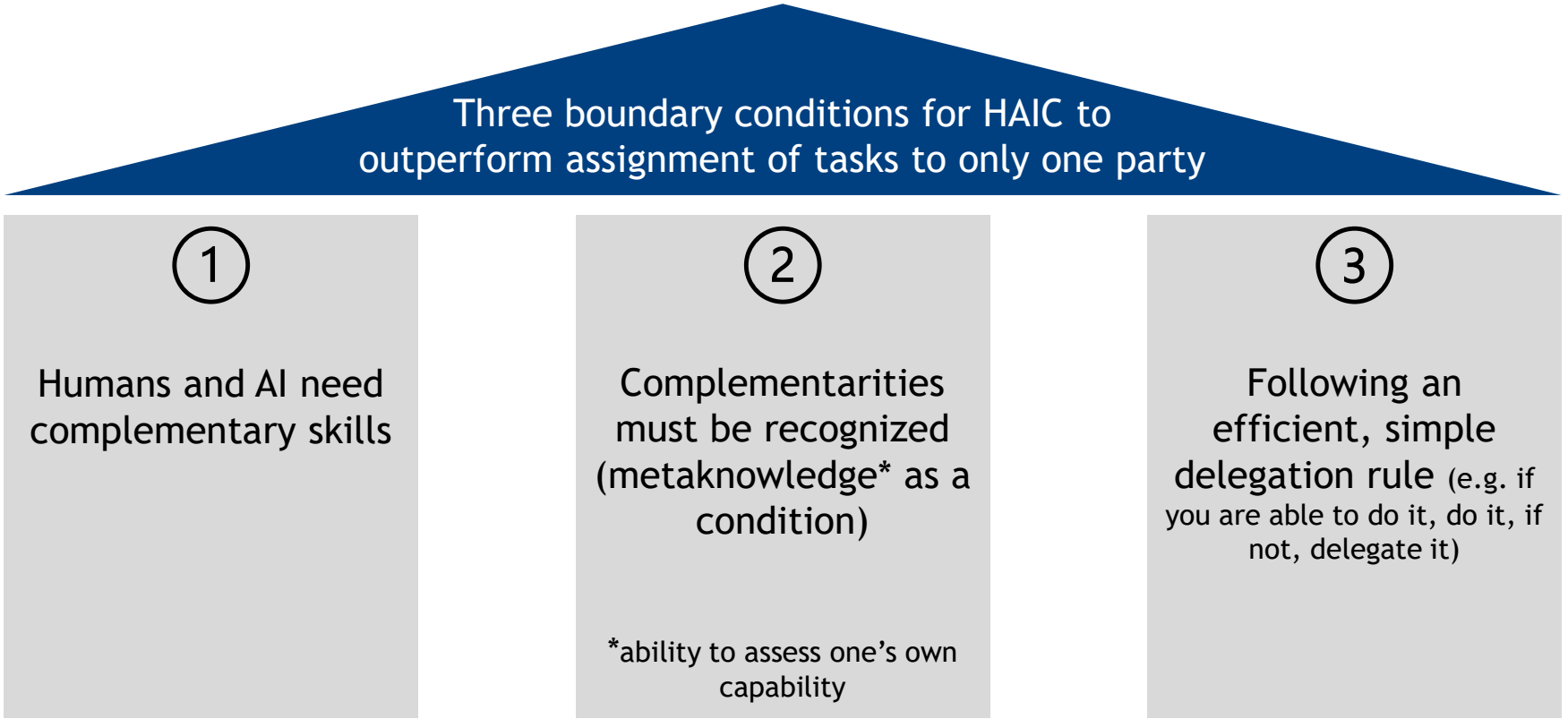
AI-led delegation from a principal-agent perspective



The principal agent theory applied to the collaboration of AI and human shows the two potential conflicts that can arise, which are hidden action and hidden information in HAIC

Guggenberger et al. (2023)

Three boundary conditions of outperforming HAIC



The three layers are important to consider for the design of Human-AI collaboration settings

Fügener et al. (2022)

Agenda

01 | Acceptance driving factors of AI

02 | Algorithmic Transparency & Explainability

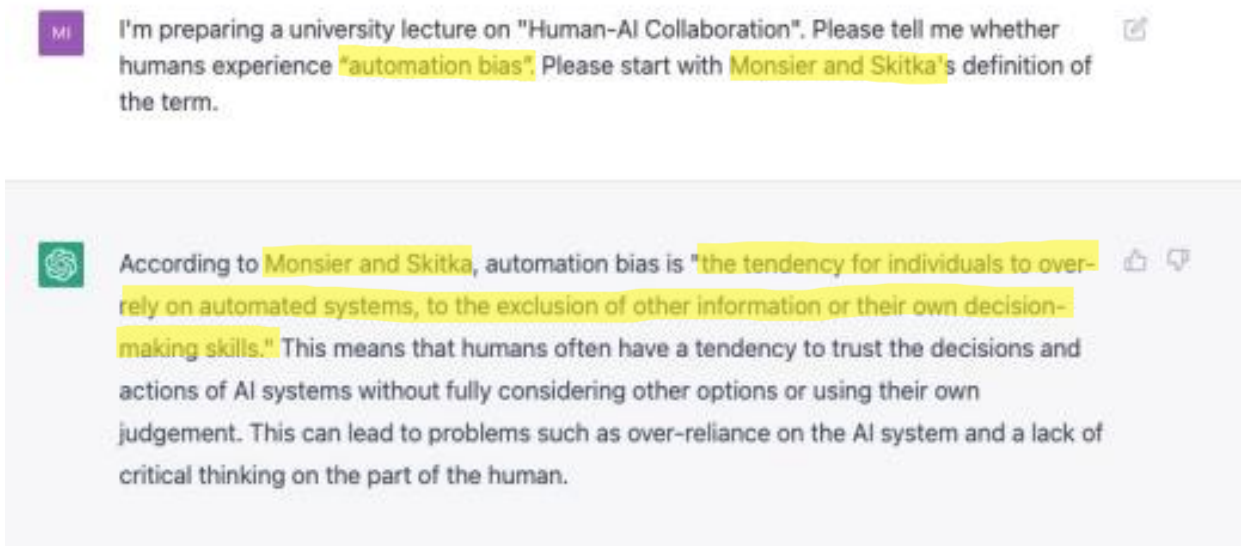
03 | Complementarity

04 | Fairness

05 | Delegation

06 | Risk of Human-AI Collaboration

Automation bias as a risk of Human-AI Collaboration



Prompt was to define “automation bias” according to Mosier and Skitka (funnily with a typo)

Here is what ChatGPT did:

- x Adopts typo of author (Mosier ≠ Monsier)
- x Gives us a definition that sounds correct but is not from Mosier and Skitka

This example shows the importance of humans staying the last insuring instance before a decision is taken with an artificial intelligence and do not blindly trust

ChatGPT (2022)

Automation bias as a risk of Human-AI Collaboration

Skitka & Mosier: Automation Bias is “the tendency to use automated cues as a heuristic replacement for vigilant information seeking and processing”. This means that humans sometimes trust the decisions and actions of AI systems without (a) considering other options or (b) using their own judgement.

Reasons for automation bias:

- Perceived reliability and accuracy of an “infallible” AI system
- Assumption that AI systems are unbiased and objective
- Insufficient clarity and comprehension regarding the decision-making processes of the AI system

The phenomenon of automation bias can lead to overreliance on AI systems and a lacking critical thinking by humans

Mosier & Skitka (1996)

Today’s lecture at a glance

- We learned how algorithmic transparency, explainability and other factors influence the acceptance of AI systems by humans
- We understood why omitting certain variables will not prevent AI models to come up with biased conclusions
- We discovered potential issues in a human AI collaboration based on the principal agent theory

Questions, comments, observations

Scientific references

- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information fusion, 58, 82-115.
- Barocas, Solon; Hardt, Moritz; Narayanan, Arvind. Fairness and Machine Learning: Limitations and Opportunities. fairmlbook.org, 2019
- Diakopoulos, N., & Koliska, M. (2017). Algorithmic transparency in the news media. Digital journalism, 5(7), 809-828.
- Fügener, A., Grahl, J., Gupta, A., and Ketter, W. 2022. “Cognitive Challenges in Human–Artificial Intelligence Collaboration: Investigating the Path Toward Productive Delegation,” Information Systems Research (33:2), pp. 678-696.
- Goodwin, P. (2000). Correct or combine? Mechanically integrating judgmental forecasts with statistical methods. International Journal of Forecasting, 16(2), 261-275.
- Guggenberger, T., Lämmermann, L., Urbach, N., Walter, A. M., & Hofmann, P. (2023). Task delegation from AI to humans: A principal-agent perspective.
- Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2018). When will AI exceed human performance? Evidence from AI experts. Journal of Artificial Intelligence Research, 62, 729-754.
- Haas, Christian. "The price of fairness-A framework to explore trade-offs in algorithmic fairness.” International Conference on Information Systems, 2019.
- Hardt, Moritz; Price, Eric; Srebro, Nathan. Equality of Opportunity in Supervised Learning. 30th Conference on Neural Information Processing Systems 2016
- Hemmer, P., Schellhammer, S., Vössing, M., Jakubik, J., & Satzger, G. (2022). Forming Effective Human-AI Teams: Building Machine Learning Models that Complement the Capabilities of Multiple Experts.
- Huang, M-H. and Rust, R. T.(2018),“Artificial intelligence in service”, Journal of Service Research, Vol.21No.2,pp.155-172.

Scientific references

- Kossow, N., Windwehr, S., & Jenkins, M. (2021). Algorithmic transparency and accountability. Transparency International.
- Laato, S., Tiainen, M., Najmul Islam, A. K. M., & Mäntymäki, M. (2022). How to explain AI systems to end users: a systematic literature review and research agenda. Internet Research, 32(7), 1-31.
- Leopold, H., van Der Aa, H., & Reijers, H. A. (2018). Identifying candidate tasks for robotic process automation in textual process descriptions. In Enterprise, Business-Process and Information Systems Modeling: 19th International Conference, BPMDS 2018, 23rd International Conference, EMMSAD 2018, Held at CAiSE 2018, Tallinn, Estonia, June 11-12, 2018, Proceedings 19 (pp. 67-81). Springer International Publishing.
- Leyer, M., & Schneider, S. (2019). Me, you or AI? How do we feel about delegation.
- Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., ... & Sanghvi, S. (2017). Jobs lost, jobs gained: Workforce transitions in a time of automation. McKinsey Global Institute, 150(1), 1-148.
- Mosier, K. L., & Skitka. L. J. (in press). Human Decision Makers and Automated Decision Ads: Made for Each Other? In R. Parasuraman & M. Mouloua (Eds.), Automation and Human Performance: Theory and Applications. NJ: Lawrence Erlbaum Associates, lnc.
- Rodolfa, Kit T., Hemank Lamba, and Rayid Ghani. "Machine learning for public policy: Do we need to sacrifice accuracy to make models fair?." Preprint, 2020
- Schemmer, M., Kuehl, N., Benz, C., Bartos, A., & Satzger, G. (2023, March). Appropriate Reliance on AI Advice: Conceptualization and the Effect of Explanations. In Proceedings of the 28th International Conference on Intelligent User Interfaces
- Scheuer, D. (2020). Akzeptanz von Künstlicher Intelligenz. Springer. https://doi. org/10.1007/978-3-658-29526-4_5.
- Siau, K., & Wang, W. (2018). Building Trust in Artificial Intelligence, Machine Learning, and Robotics. Cutter Business Technology Journal, 31(2). www.cutter.com

Scientific references

- Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (xai): Toward medical xai. IEEE transactions on neural networks and learning systems, 32(11), 4793-4813. Hemmer, P., Schemmer, M., Riefle, L., Rosellen, N., Vössing, M., & Kühl, N. (2022). Factors that influence the adoption of human-AI collaboration in clinical decision-making. arXiv preprint arXiv:2204.09082
- Vokinger, K. N., & Gasser, U. (2021). Regulating AI in medicine in the United States and Europe. Nature machine intelligence, 3(9), 738-739., https://digitalstrategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-a, https://artificialintelligenceact.eu/the-act/
- Wirtz, J., Patterson, P. G., Kunz, W. H., Gruber, T., Lu, V. N., Paluch, S., & Martins, A. (2018). Brave new world: service robots in the frontline. Journal of Service Management, 29(5), 907-931.
- Yang, K., Qinami, K., Fei-Fei, L., Deng, J., & Russakovsky, O. (2020, January). Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy. In Proceedings of the 2020 conference on fairness, accountability, and transparency (pp. 547-558).
- Zhang, T., Kaber, D. B., Zhu, B., Swangnetr, M., Mosaly, P., & Hodge, L. (2010). Service robot feature design effects on user perceptions and emotional responses. Intelligent Service Robotics, 3(2), 73–88. https://doi.org/10.1007/s11370-010- 0060-9

Non-scientific references

- https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html
- https://www.forbes.com/sites/ariannajohnson/2023/03/30/which-jobs-will-ai-replace-these-4-industries-will-be-heavily-impacted/
- https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier?utm_source=ai-automation-life-daily.beehiiv.com&utm_medium=referral&utm_campaign=the-future-of-work-ai-automation-to-replace-half-of-human-jobs-by-2045#work-and-productivity
- https://www.continental.com/en/press/press-releases/germans-fear-job-losses-due-to-artificial-intelligence/
- Donner, S. (2021, May 7). Standards für KI gesucht. VDI Nachrichten, 12.
- Kühl, Niklas (2022). HAIC Complementarity Vorlesung. Uni Bayreuth
- https://fortune.com/2023/02/17/microsoft-chatgpt-bing-romantic-love/
- https://edition.cnn.com/2023/03/18/politics/ai-chatgpt-racist-what-matters/index.html
- https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Pictures

- https://knowyourmeme.com/memes/ai-emotion-detection
- https://coacherocg.fishing-park.co.il/enx/wait-a-minute-who-are-you/
- https://www.it-daily.net/it-sicherheit/cloud-security/chatgpt-ki-ist-jetzt-der-natuerlichen-ignoranz-gewachsen