Revisiting **Intelligence Augmentation**: Investigating and Mitigating the Risks of AI to Human Intelligence

Q. Vera Liao
Computer Science and Engineering
University of Michigan

UNIVERSITY OF MICHIGAN

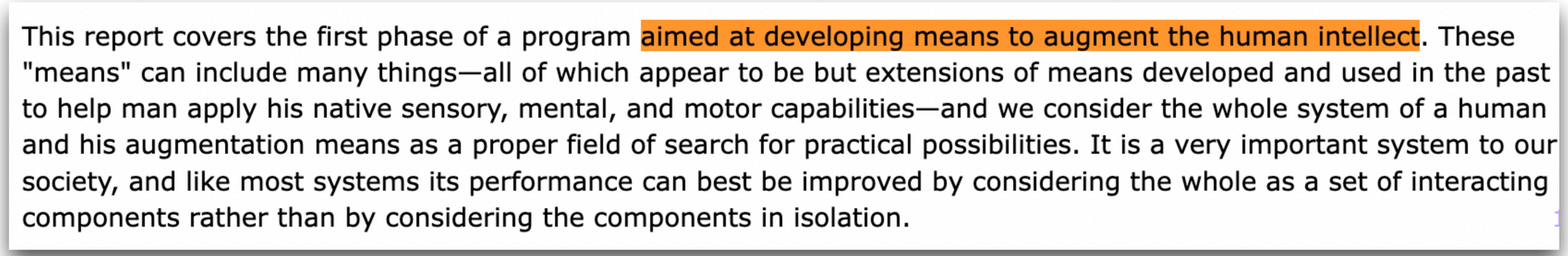
AI v.s. IA (Intelligence Augmentation)

“Machines will be capable, within twenty years, of doing any work that a man can do”
— Herbert Simon, 1960

AI v.s. IA (Intelligence Augmentation)

“Machines will be capable, within twenty years, of doing any work that a man can do”

— Herbert Simon, 1960


This report covers the first phase of a program aimed at developing means to augment the human intellect. These "means" can include many things—all of which appear to be but extensions of means developed and used in the past to help man apply his native sensory, mental, and motor capabilities—and we consider the whole system of a human and his augmentation means as a proper field of search for practical possibilities. It is a very important system to our society, and like most systems its performance can best be improved by considering the whole as a set of interacting components rather than by considering the components in isolation.

Augmenting Human Intellect: A Conceptual Framework

—Douglas Engelbart, 1962

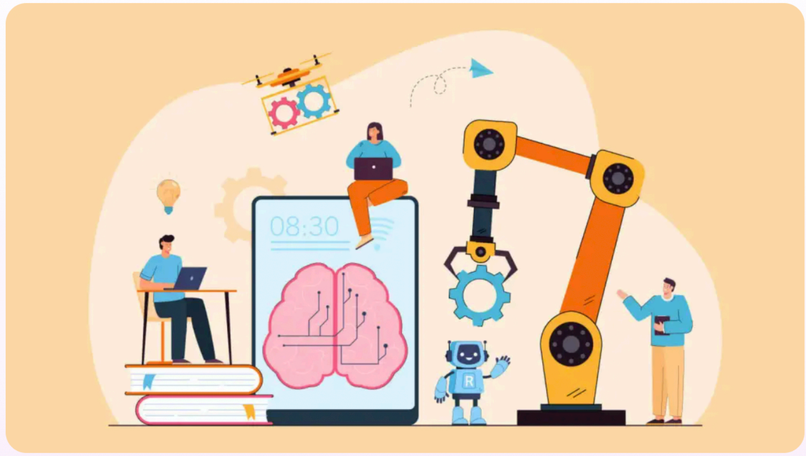
AI augmenting, not replacing, people

AI as Assistant/Copilot/Collaborator…

Human-AI Collaboration: How to Work Together

Prepare your digital workforce to thrive in an AI-driven world.

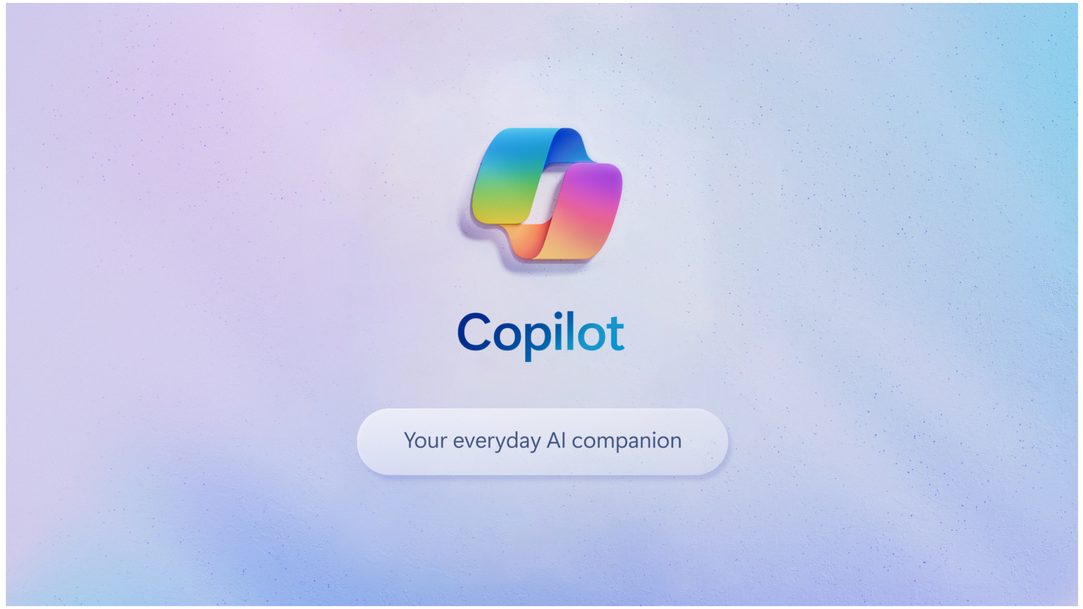
Choi Chow, Product Marketing Manager



Go from answers to action with Claude Cowork

Cowork brings Claude Code's agentic capabilities to your desktop. Give Claude access to your local files, set a task, and step away. Come back to completed work. Now in research preview.

Download Claude | Talk to sales



Intelligence Augmentation in Education

For teachers

Experience the best AI for teachers.

Built for educators by educators, with your needs in mind. Khanmigo simplifies your workflow while keeping your work and your student data private and secure.

Safe, teacher-guided AI for students

Brisk's student-facing AI tools empower educators to give every student personalized, interactive support right where they're learning — all with teachers in the driver's seat.

For Teachers

An AI Toolkit for the Work Teachers Do Every Day.

Enhance teaching with research-backed tools

Built on K-12 research and best practices, Solara includes ready-made prompts and tools to create lesson plans, rubrics, grade-level texts, and more—saving time for meaningful student engagement. Easily configure the tools to fit your district's needs.

AI FOR TEACHERS

Save time.
Spark creativity.
Personalize learning.

Teachers at the center

Our human-in-the-loop approach to AI ensures that teachers always have the final say, keeping educator expertise at the heart of every decision.

OFFICE OF
Educational Technology

Artificial Intelligence and the Future of Teaching and Learning

Insights and Recommendations

May 2023

Perspective: Intelligence Augmentation

> *"Augmented intelligence is a design pattern for a human-centered partnership model of people and artificial intelligence (AI) working together to enhance cognitive performance, including learning, decision making, and new experiences."* [16]

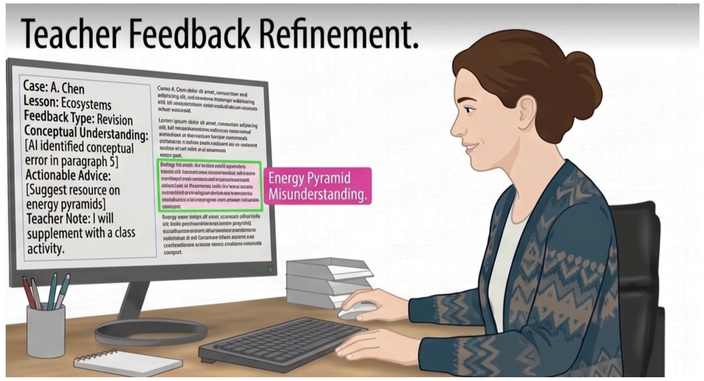
Foundation #1 (above) keeps humans in the loop and positions AI systems and tools to support human reasoning. "Intelligence Augmentation" (IA)[17] centers "intelligence" and "decision making" in humans but recognizes that people sometimes are overburdened and benefit from assistive tools. AI may help teachers make better decisions because computers notice patterns that teachers can miss. For example, when a teacher and student agree that the student needs reminders, an AI system may provide reminders in whatever form a student likes without adding to the teacher's workload. Intelligence Automation (IA) uses the same basic capabilities of AI, employing associations in data to notice patterns, and, through automation, takes actions based on those patterns. However, IA squarely focuses on helping people in human activities of teaching and learning, whereas AI tends to focus attention on what computers can do.

What are the challenges in achieving IA?

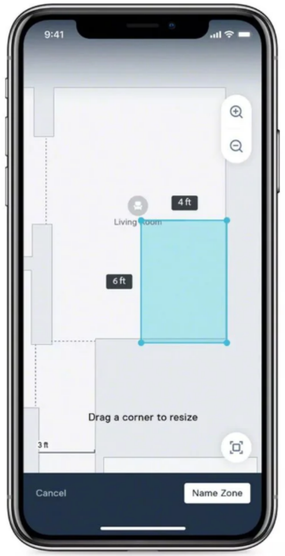
How to teach IA for the “AI generation”?

People have always developed IA tools to **offload** some cognitive process to make the whole tasks better, easier, faster…

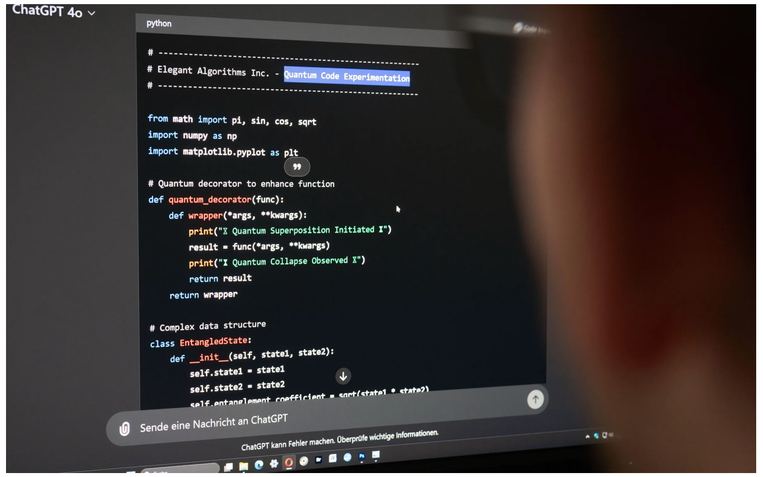
Increasing offloadability



AI suggests, human decides



Human plans, AI executes

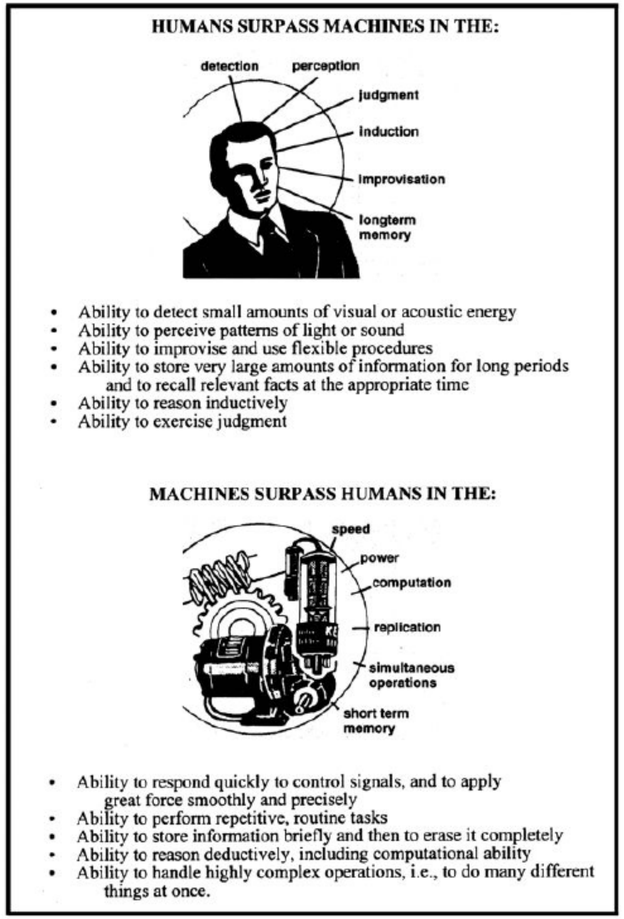


AI plans, human fixes

Increasingly capable and agentic AI

MAN-COMPUTER SYMBIOSIS

> Separable Functions of Men and Computers in the Anticipated Symbiotic Association… Men will set the goals and supply the motivations, of course, at least in the early years. They will formulate hypotheses… The information-processing equipment, for its part, will convert hypotheses into testable models and then test the models against data.
>
> *—J.C.R. Licklider*



Offloading “what machines are better at” to complement people

Generated with Pixtral Version 24.09 in the free version of mistral.ai;

Oversight (“executive power”) retained by people

Article 14: Human Oversight

Humans oversight personnel should…

- … prevent or minimise the risks to health, safety or fundamental rights
- … properly monitor the AI system and understand its capacities and limitations
- … remain aware of the possible tendency of automatically (over-)relying on the output produced by the AI system (automation bias)
- … correctly interpret AI output and decide not to use / disregard AI
- …to intervene in the operation of the high-risk AI system or interrupt the system through a ‘stop’ button or a similar procedure that allows the system to come to a halt in a safe state.
- …

AI EU Act mandates oversight for high-risk AI systems

Oversight (“executive power”) retained by people

Article 14: Human Oversight

Humans oversight personnel should…

- … prevent or minimise the risks to health, safety or fundamental rig
- … properly monitor the AI system and understand its capacities an
- … remain aware of the possible tendency of automatically (over-)re
  produced by the AI system (automation bias)
- … correctly interpret AI output and decide not to use / disregard AI
- …to intervene in the operation of the high-risk AI system or interrupt the system through a ‘stop’ button or a similar procedure that allows the system to come to a halt in a safe state.
- …

3. Education and vocational training:

(a) AI systems intended to be used to determine access or admission or to assign natural persons to educational and vocational training institutions at all levels;

(b) AI systems intended to be used to evaluate learning outcomes, including when those outcomes are used to steer the learning process of natural persons in educational and vocational training institutions at all levels;

(c) AI systems intended to be used for the purpose of assessing the appropriate level of education that an individual will receive or will be able to access, in the context of or within educational and vocational training institutions at all levels;

(d) AI systems intended to be used for monitoring and detecting prohibited behaviour of students during tests in the context of or within educational and vocational training institutions at all levels.

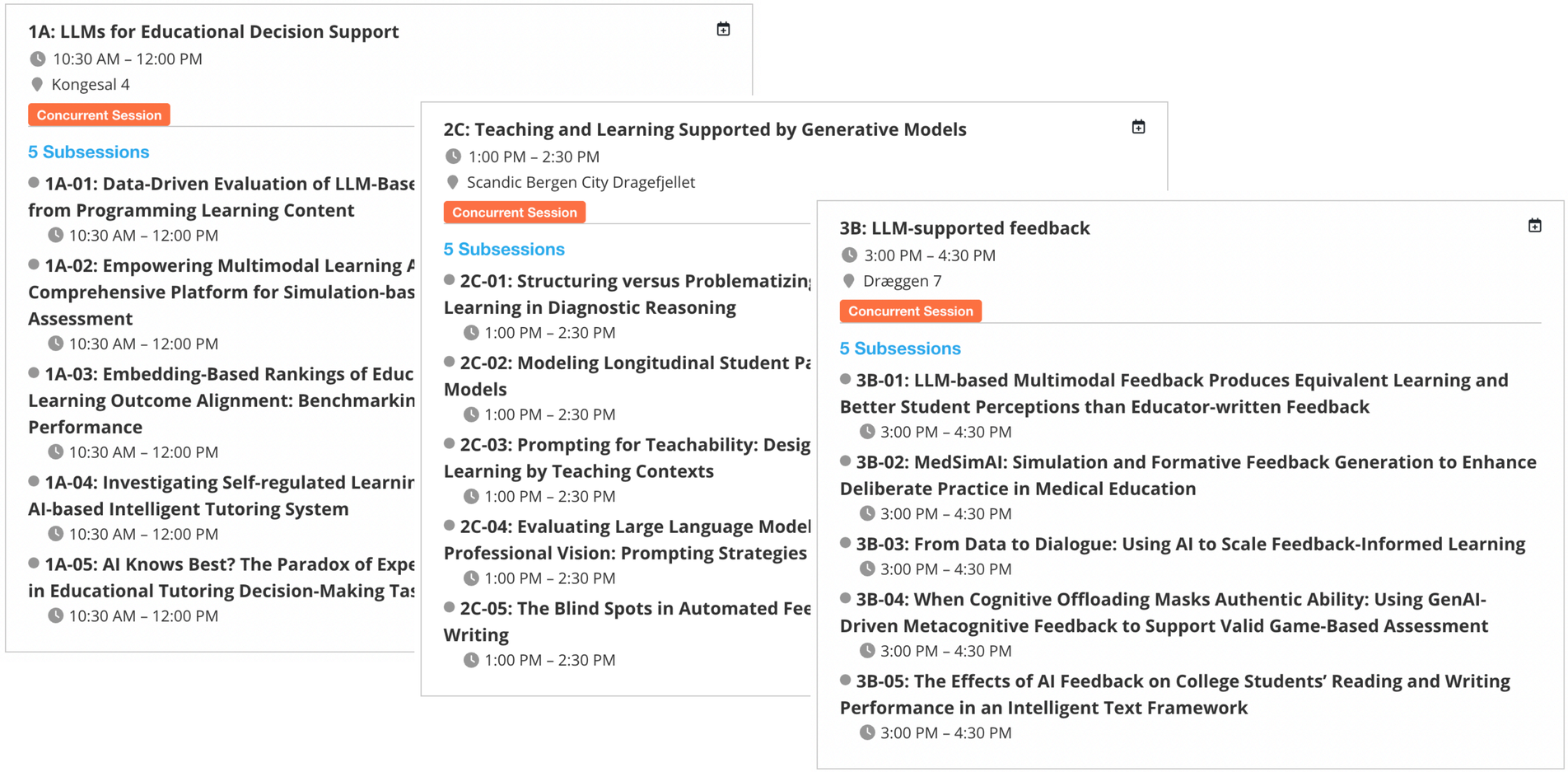
AI EU Act mandates oversight for high-risk AI systems

Oversight (“executive power”) retained by people

- Overreliance on AI: challenges with complementarity and oversight
- New affordances of GenAI create new threat to human thinking: risks of offloading

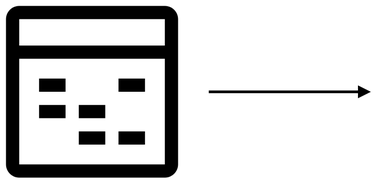
Research Thread 1: **Investigating and Mitigating Overreliance on AI**

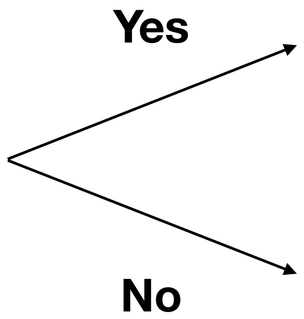
Challenges with complementarity and oversight



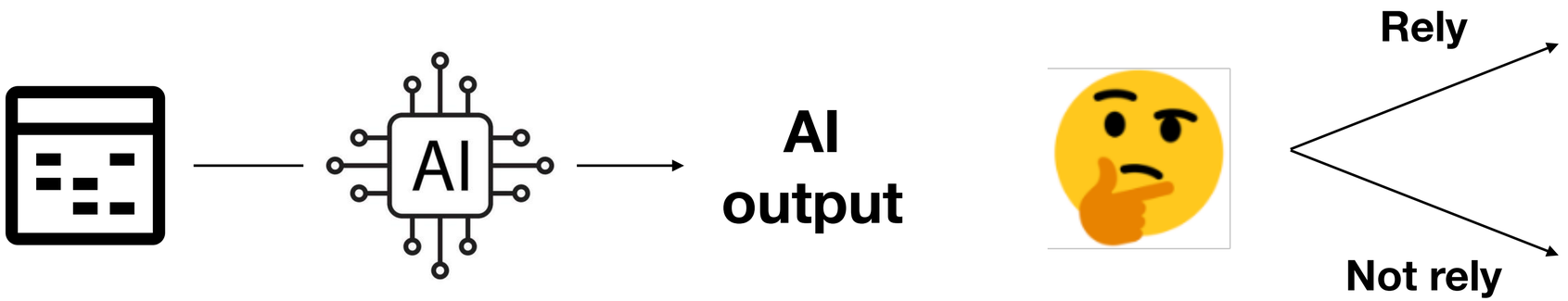
AI-Assisted Decision-Making: AI Suggests, Human Decides

Decision-Making





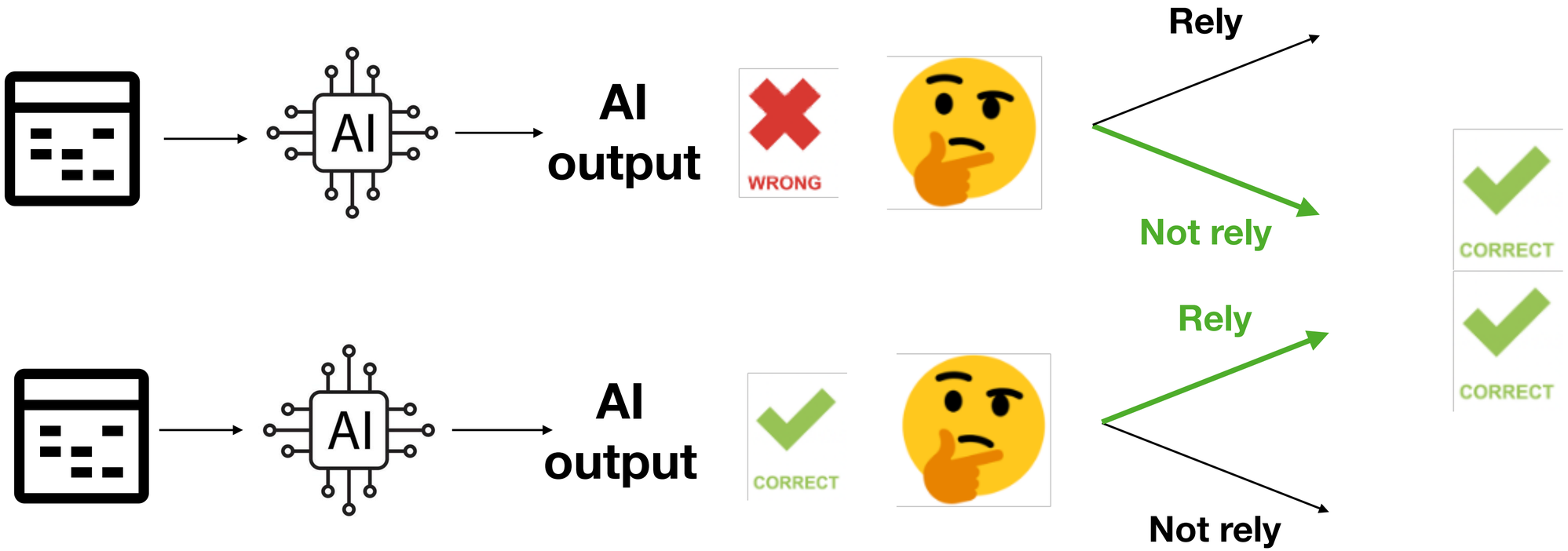
AI Assisted Decision-Making



AI Assisted Decision-Making

Human oversight to catch AI errors

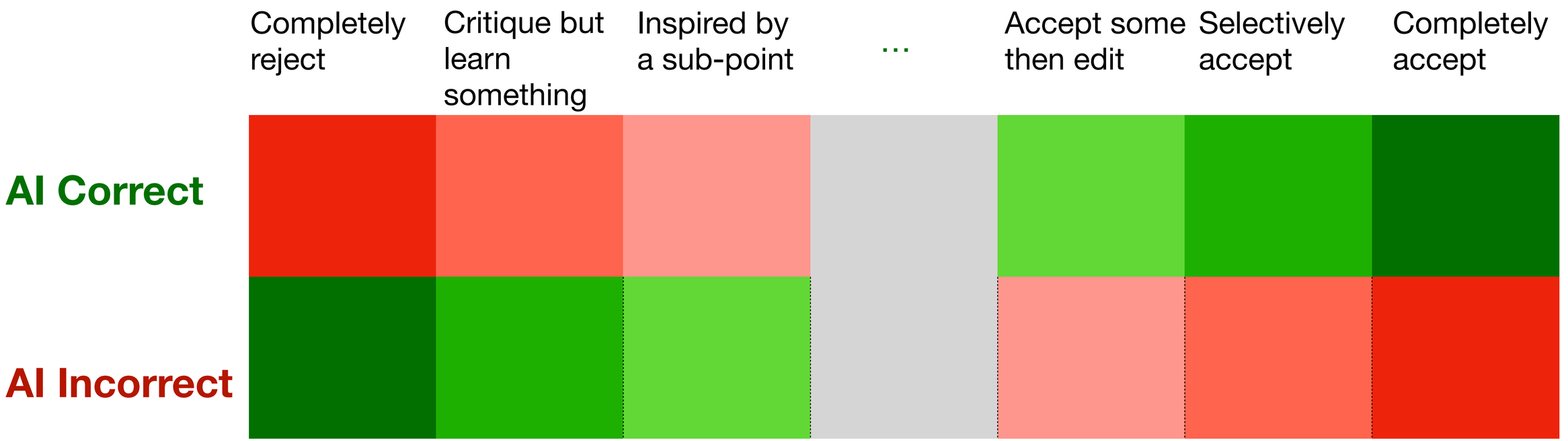
AI Assisted Decision-Making



Human-AI complementarity to make more correct decisions than either human or AI would have done alone

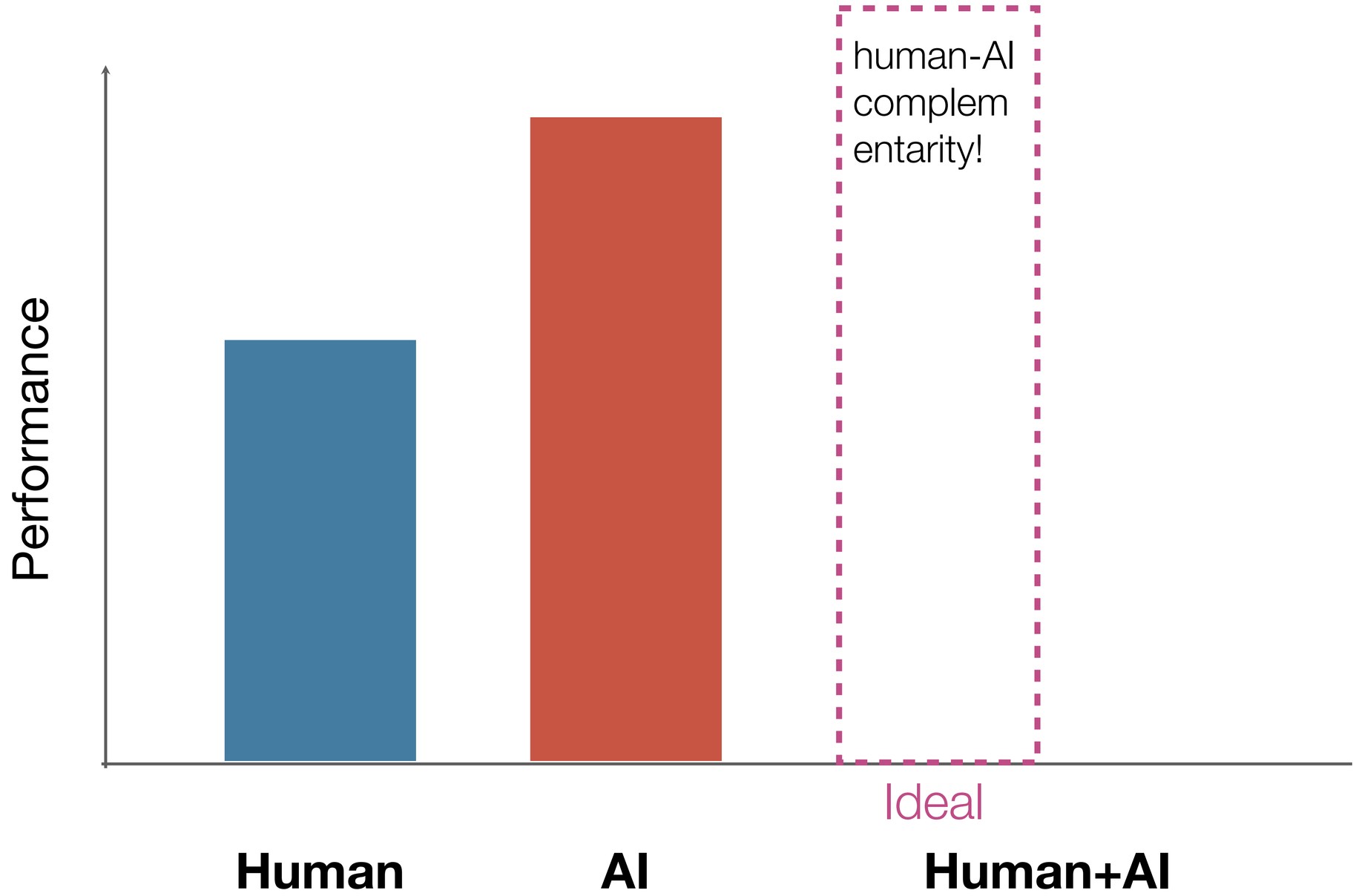
	Not Rely	Rely
AI Correct	Underreliance	Correct Reliance
AI Incorrect	Correct Non-Reliance	Overreliance

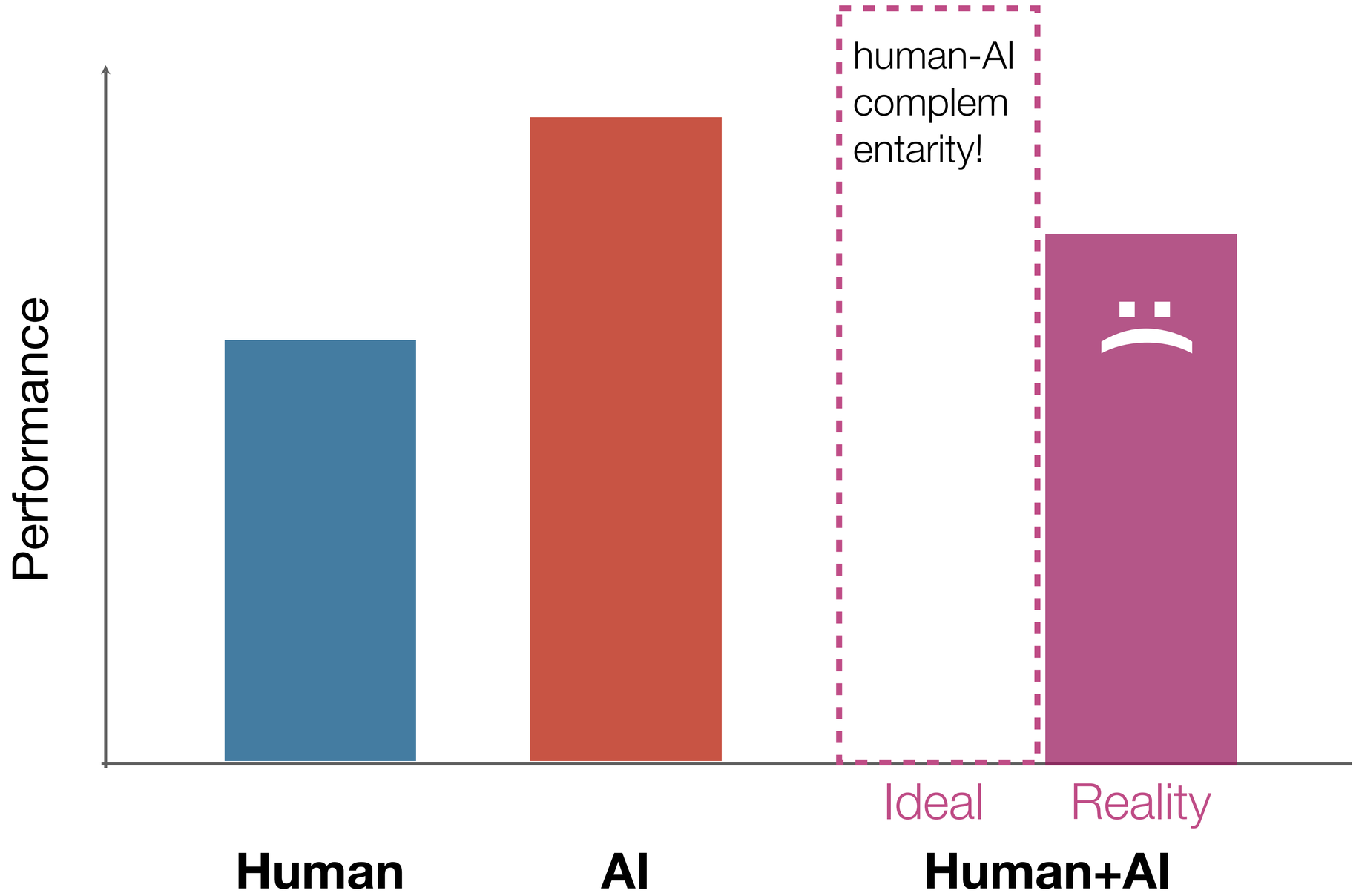
Complementarity requires **appropriate reliance (offloading)**:
rely when and only when the AI output is correct (green zones)



Side note: The reliance action space may not be binary depending on the type of AI

The AI-assisted decision-making literature started with largely classification models


human-AI
complem
entarity!
Performance
Ideal
Human
AI
Human+AI


human-AI
complem
entarity!
Performance
Ideal
Reality
Human
AI
Human+AI

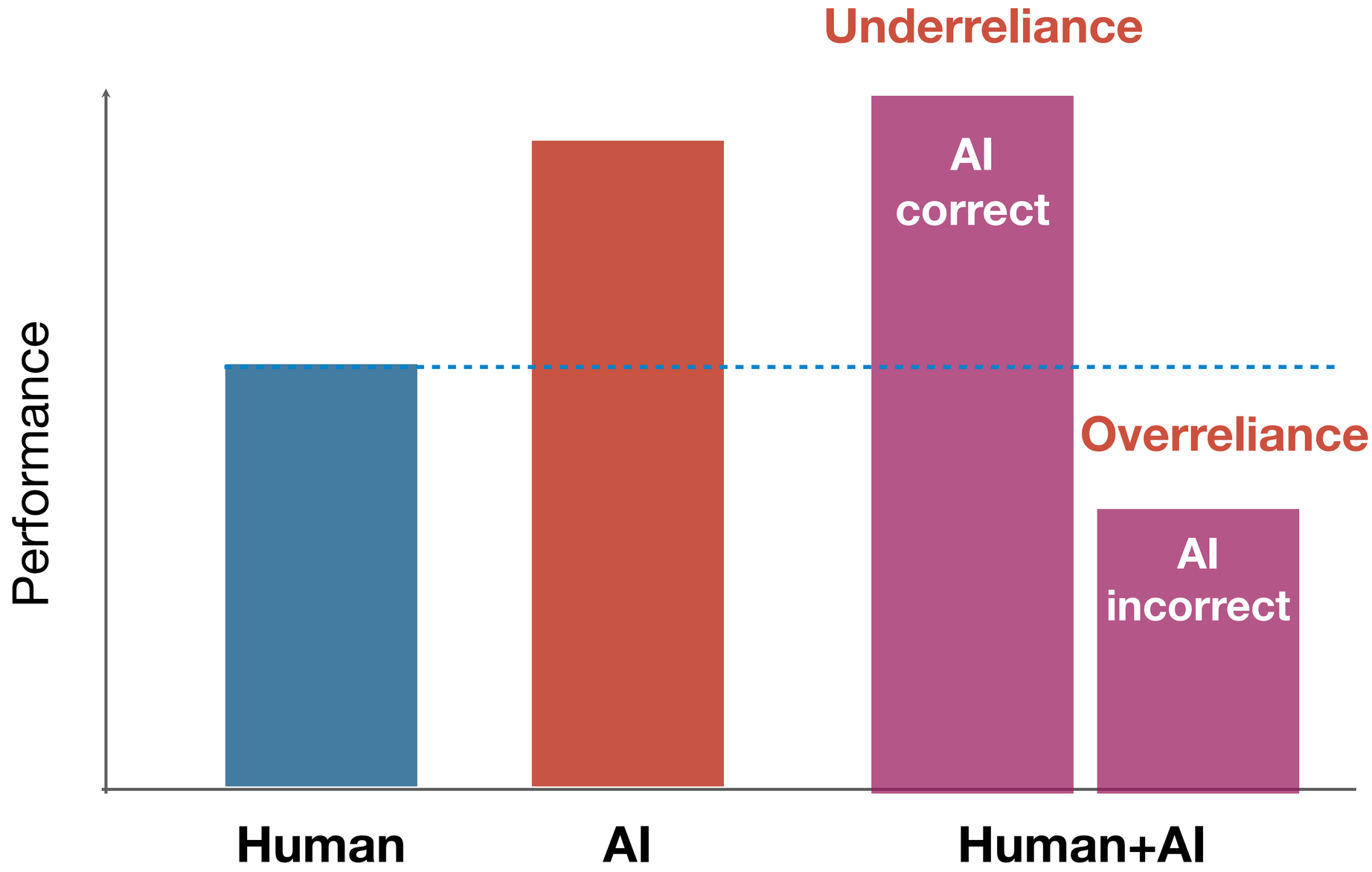

Underreliance
Performance
AI correct
Overreliance
AI incorrect
Human
AI
Human+AI

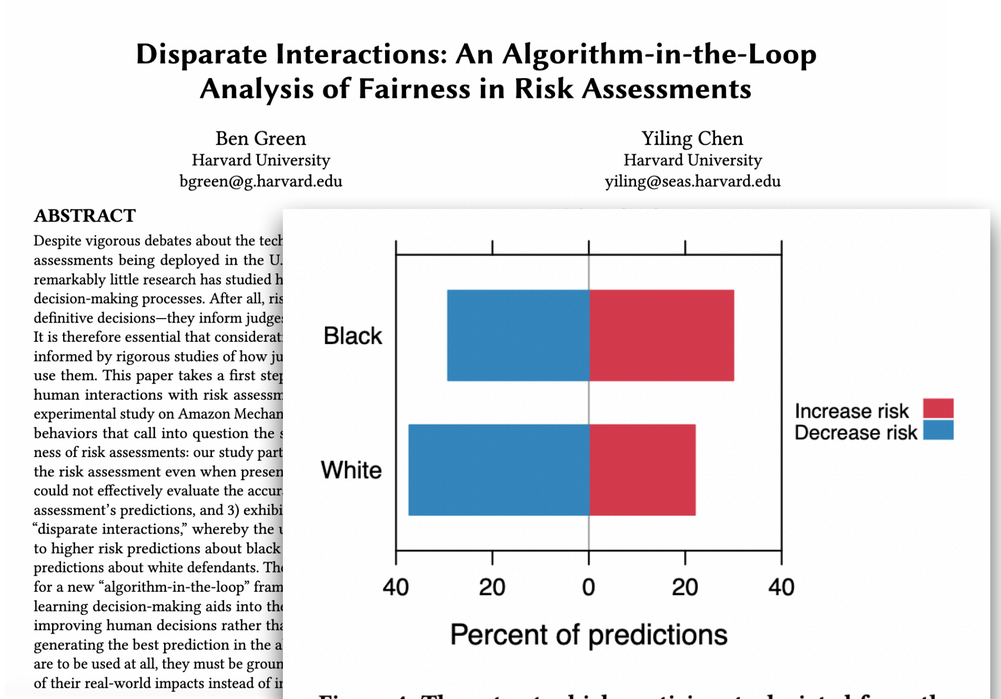
Table 1: Short and long-term impacts of overreliance on LLMs in two domains

Domain	Short-term	Long-term
Healthcare	**Individual:** Medical doctor follows incorrect AI diagnosis **Institutional:** Hospital systems implement AI diagnostics without adequate verification protocols	**Individual:** Healthcare professionals experience cognitive deskilling and atrophy of diagnostic abilities **Institutional:** Medical departments restructure workflows around AI systems, reducing human oversight
Personal advice	**Individual:** User accepts sycophantic or biased relationship or career advice without critical evaluation **Societal:** Communities adopt similar AI-generated advice, creating homogenized decision patterns	**Individual:** Self-worth becomes derived from AI companion approval [82] **Societal:** Social norms shift toward algorithmic validation of personal choices

Harms from overreliance:

- Poor human-AI performance
- Biased/homogenized errors (often a cause for inequality)

Ibrahim et al. Measuring and mitigating overreliance is necessary for building human-compatible AI



Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments

Ben Green
Harvard University
bgreen@g.harvard.edu

Yiling Chen
Harvard University
yiling@seas.harvard.edu

ABSTRACT

Despite vigorous debates about the tech
assessments being deployed in the U.
remarkably little research has studied h
decision-making processes. After all, ris
definitive decisions—they inform judge
It is therefore essential that considerat
informed by rigorous studies of how ju
use them. This paper takes a first step
human interactions with risk assessm
experimental study on Amazon Mechan
behaviors that call into question the s
ness of risk assessments: our study part
the risk assessment even when presen
could not effectively evaluate the accur
assessment's predictions, and 3) exhibi
"disparate interactions," whereby the u
to higher risk predictions about black
predictions about white defendants. Th
for a new "algorithm-in-the-loop" fram
learning decision-making aids into th
improving human decisions rather tha
generating the best prediction in the a
are to be used at all, they must be groun
of their real-world impacts instead of i

Figure 4: The rate at which participants deviated from the risk assessment's prediction toward higher and lower levels of risk, broken down by defendant race. When evaluating black defendants, participants were 36.4% more likely to deviate positively from the risk assessment and 21.5% less likely to deviate negatively (participant predictions matched the risk assessment at an equal rate for both races).

Biased **overreliance:** When using a biased decision-support, human+AI can get more biased than human or AI alone

(More likely to follow AI errors for demographic groups for which one already hold bias against)

Green & Chen. Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In FAccT 2019

Table 1: Short and long-term impacts of overreliance on LLMs in two domains

Domain	Short-term	Long-term
Healthcare	**Individual:** Medical doctor follows incorrect AI diagnosis **Institutional:** Hospital systems implement AI diagnostics without adequate verification protocols	**Individual:** Healthcare professionals experience cognitive deskilling and atrophy of diagnostic abilities **Institutional:** Medical departments restructure workflows around AI systems, reducing human oversight
Personal advice	**Individual:** User accepts sycophantic or biased relationship or career advice without critical evaluation **Societal:** Communities adopt similar AI-generated advice, creating homogenized decision patterns	**Individual:** Self-worth becomes derived from AI companion approval [82] **Societal:** Social norms shift toward algorithmic validation of personal choices

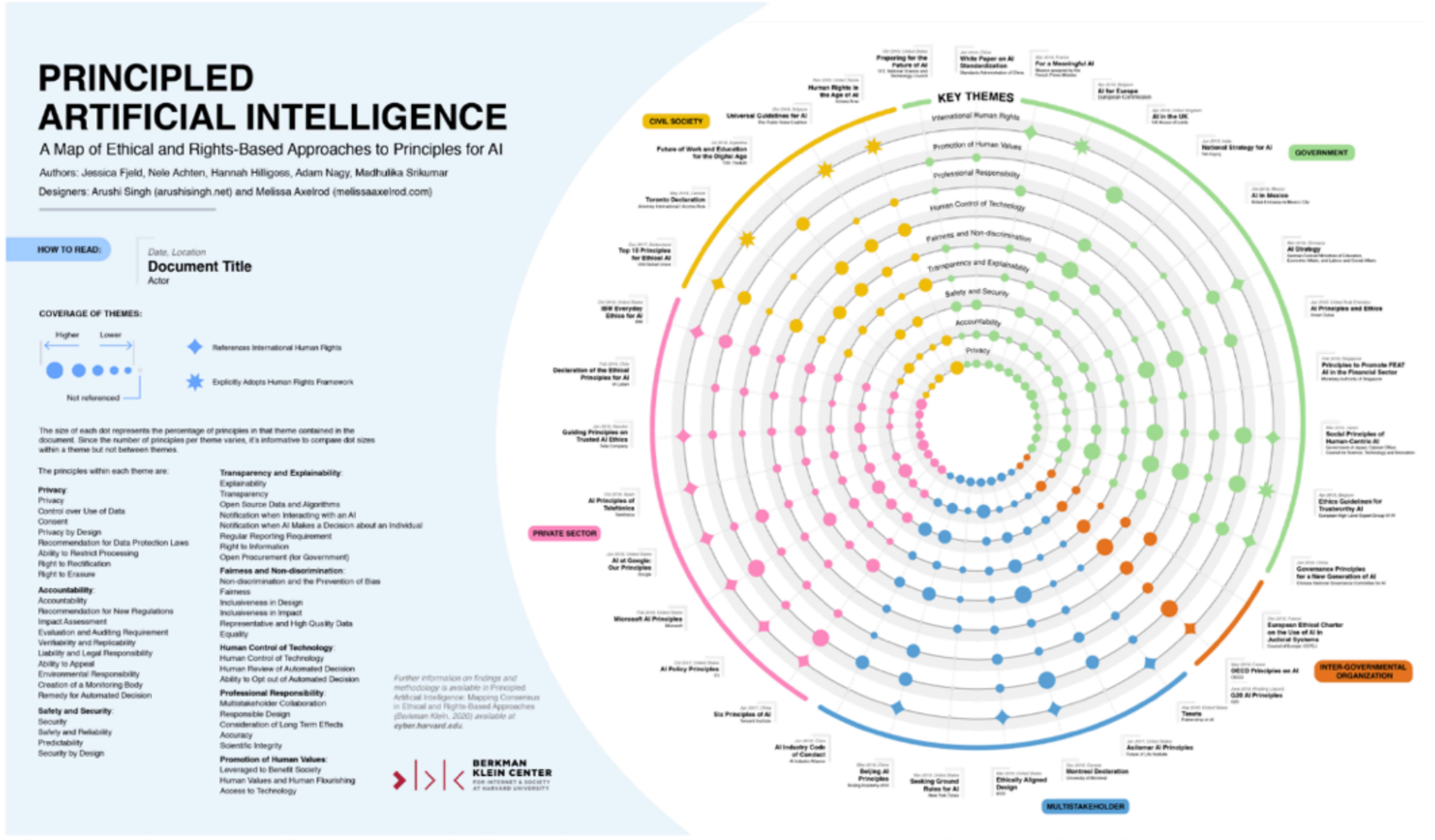
Harms from overreliance:

- Poor human-AI performance
- Biased/homogenized errors (often a cause for inequality)
- Deskilling
- Psychological harms from loss of agency
- Infrastructural vulnerabilities
- Shifting social engagement and norms
- …

Ibrahim et al. Measuring and mitigating overreliance is necessary for building human-compatible AI

A hard lesson from research on human-AI interaction:

Interventions to mitigate overreliance often do not work well, and sometimes backfire



Mapping of 36 ethical and responsible AI frameworks (Berkman Klein Center)

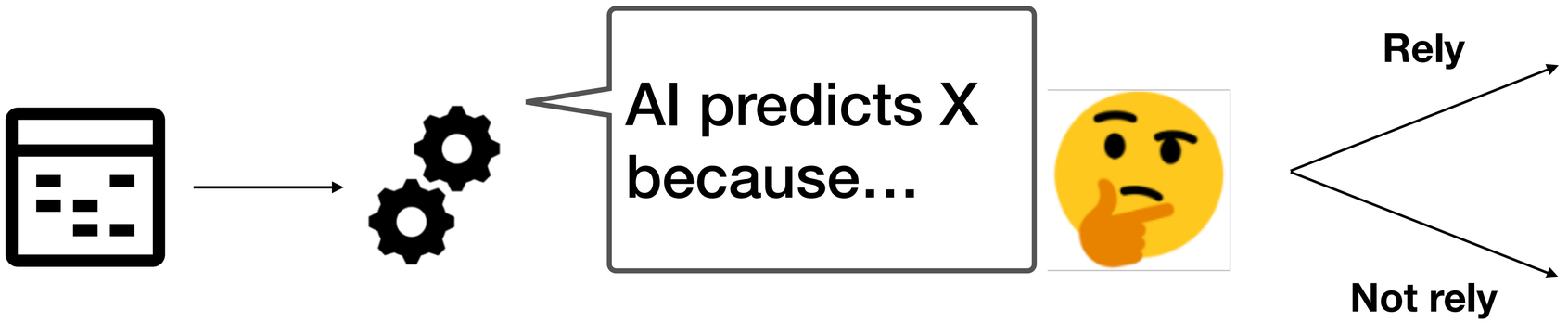
- Fairness and Non-discrimination
- Transparency and Explainability
- Human Control
- Safety and Security
- Accountability
- Privacy
- Promotion of Human Values

Article 14: Human Oversight

Humans oversight personnel should…

- … prevent or minimise the risks to health, safety or fundamental rights
- … properly monitor the AI system and understand its capacities and limitations
- … remain aware of the possible tendency of automatically (over-)relying on the output produced by the AI system (automation bias)
- … correctly interpret AI output and decide not to use / disregard AI
- …to intervene in the operation of the high-risk AI system or interrupt the system through a ‘stop’ button or a similar procedure that allows the system to come to a halt in a safe state.
- …

Explainable AI for Decision-Making



The hope: explanation can help people have better oversight of AI errors, hence less overreliance

Does this person belong to a high-income or low-income group?

Attributes	Value
Age	33
Sex	Male
Race	White
Marital status	Married
Years of edu.	9
Workclass	Private
Occupation	Sales
Hrs. per week	45

Yunfeng Zhang, Q Vera Liao, Rachel Bellamy. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. *FAccT 2020*

Does this person belong to a high-income or low-income group?

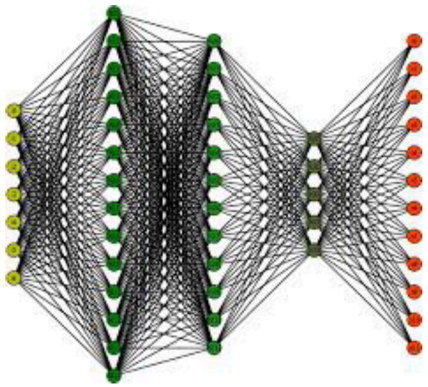
Attributes	Value
Age	33
Sex	Male
Race	White
Marital status	Married
Years of edu.	9
Workclass	Private
Occupation	Sales
Hrs. per week	45

The AI system predicts that the customer belongs to high-income group

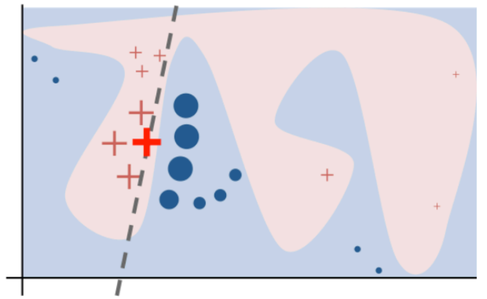
Control condition

Yunfeng Zhang, Q Vera Liao, Rachel Bellamy. Effect of Confidence and Explanation on Accuracy and Trust Calibration in AI-Assisted Decision Making. *FAccT 2020*

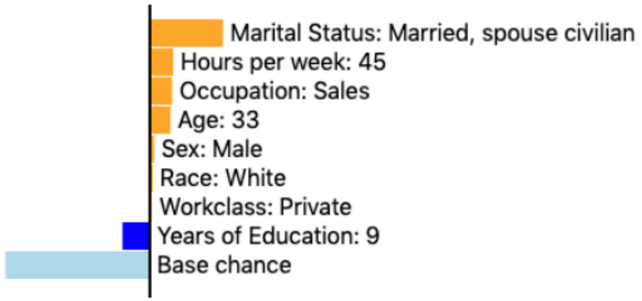
Popular XAI Algorithms Produce Feature Importance Explanations



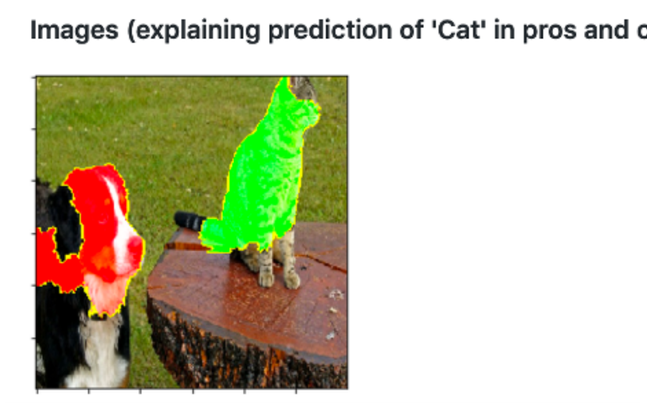
Neural network, not directly explainable



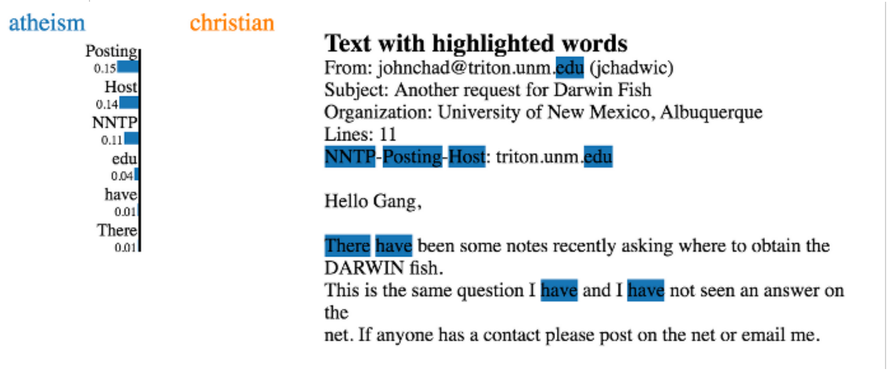
Use a *post-hoc* XAI technique



Tabuler data



Image



Texts

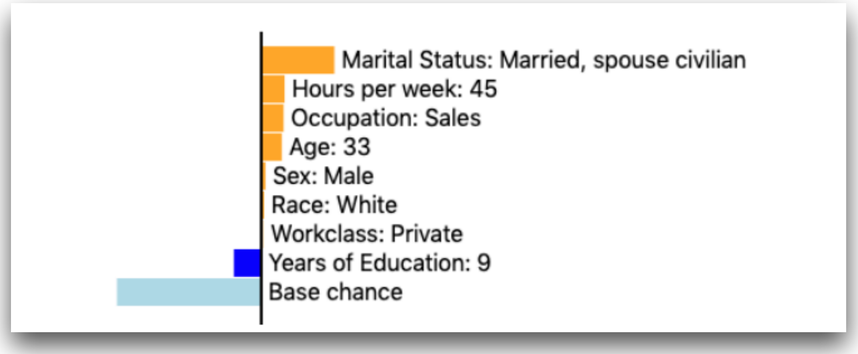
Ribeiro et al, " Why should i trust you?" Explaining the predictions of any classifier. KDD 2016

Does this person belong to a high-income or low-income group?

Attributes	Value
Age	33
Sex	Male
Race	White
Marital status	Married
Years of edu.	9
Workclass	Private
Occupation	Sales
Hrs. per week	45

The AI system predicts that the customer belongs to high-income group

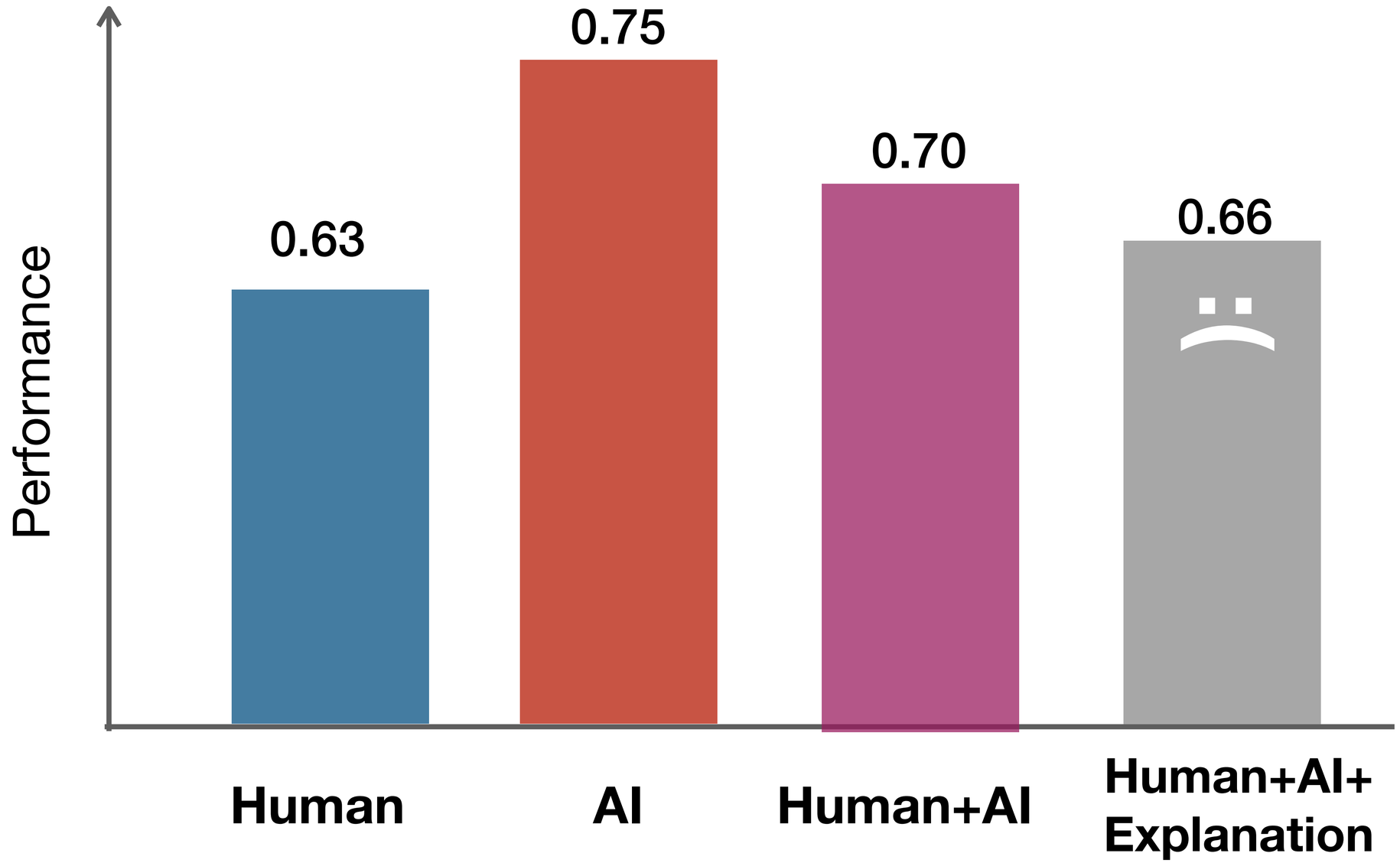
Control condition

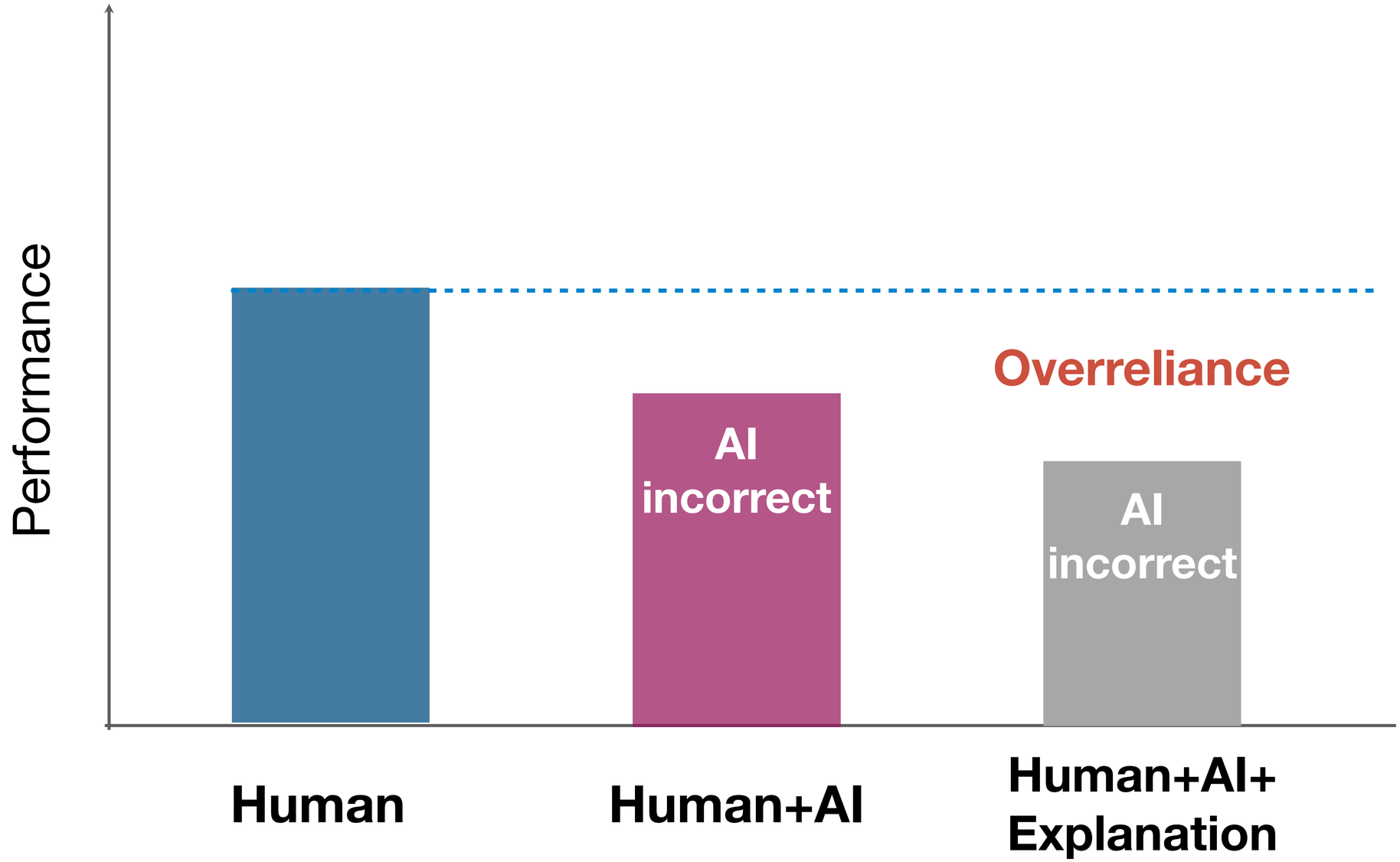


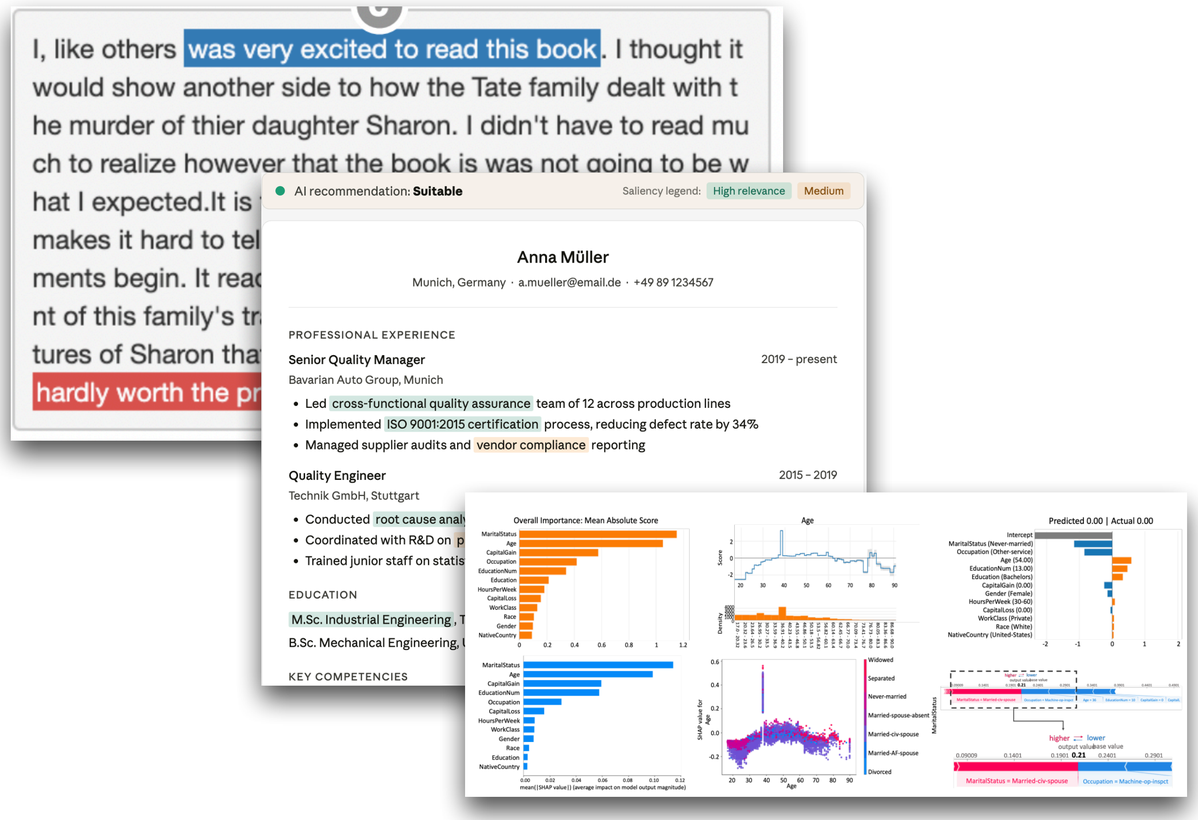
The figure explains how the system arrives at this decision according to how much each attribute likely contributes to a high or low income

The AI system predicts that the customer belongs to high-income group

Experimental condition

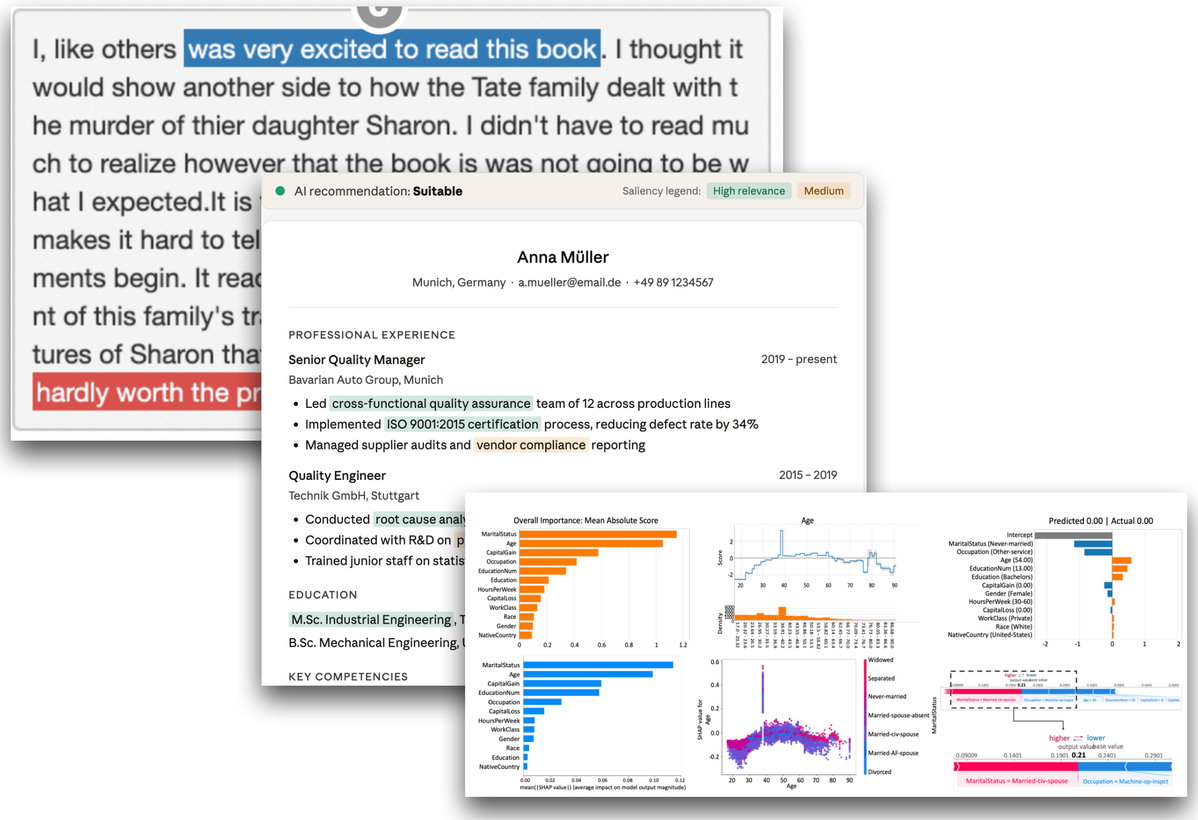

Performance
0.63
0.75
0.70
0.66
Human
AI
Human+AI
Human+AI+
Explanation


Performance
Overreliance
AI incorrect
AI incorrect
Human
Human+AI
Human+AI+ Explanation

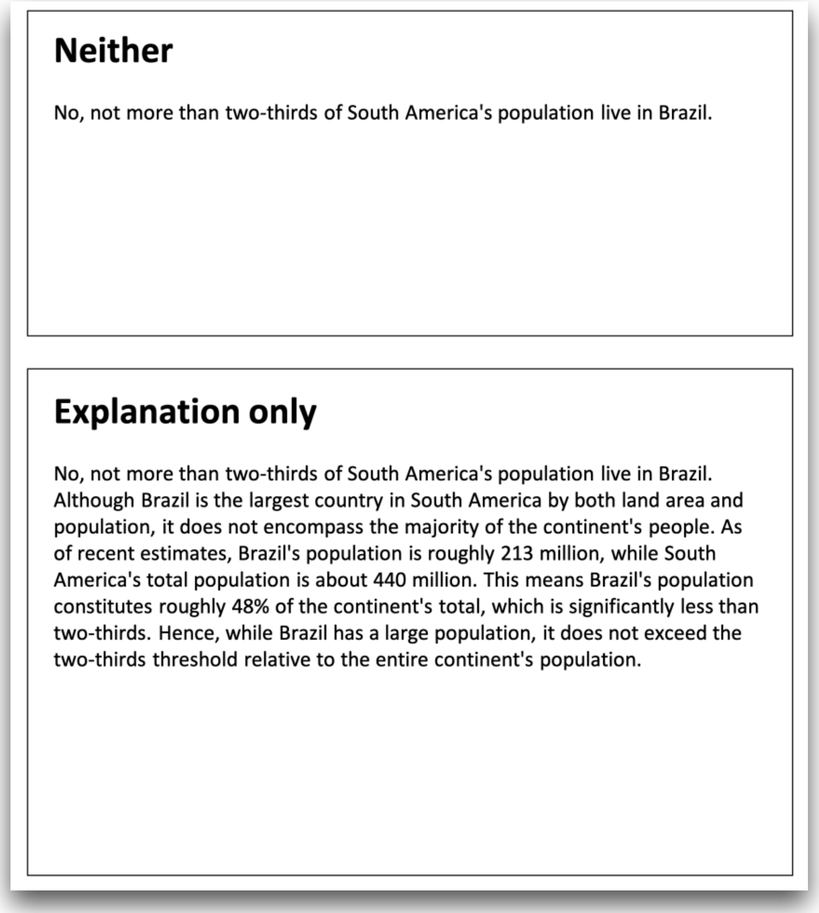


XAI increases overreliance: robust findings across use cases and modalities

Bansal et al., 2021; Chen et al., 2023; Eiband et al., 2019; Kaur et al., 2020; Poursabzi-Sangdeh et al., 2019, Cecil et al. 2024; Kim 2025; Rismani et al. 2026



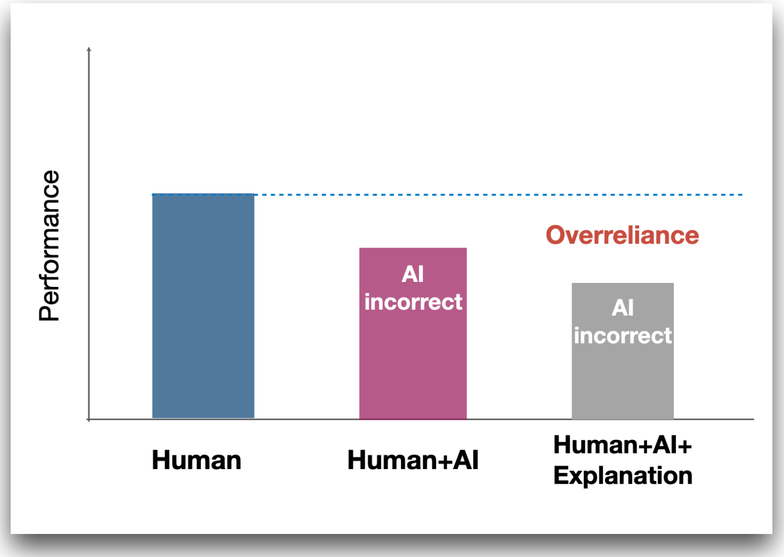
XAI increases overreliance: robust findings across use cases and modalities



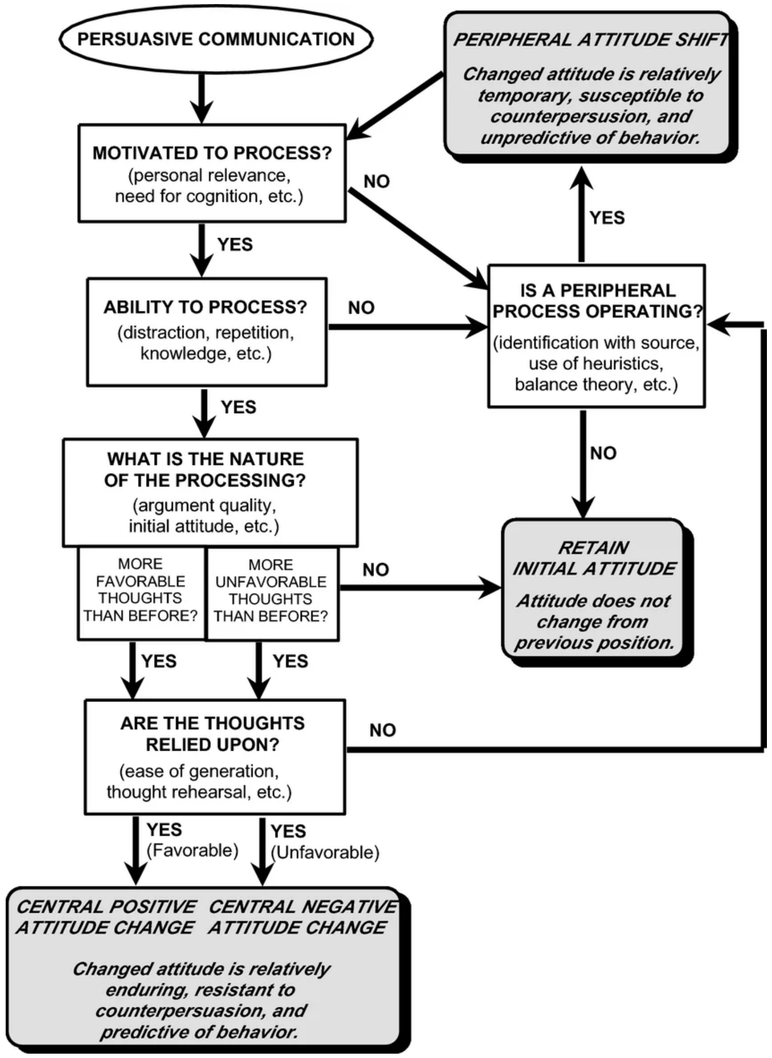
LLM reasoning also increases people's overreliance

Bansal et al., 2021; Chen et al., 2023; Eiband et al., 2019; Kaur et al., 2020; Poursabzi-Sangdeh et al., 2019, Cecil et al. 2024; Kim 2025; Rismani et al. 2026

Why Do AI Explanations Increase Overreliance?



People do not always engage analytically (“system 2” slow thinking) with AI outputs to detect AI errors

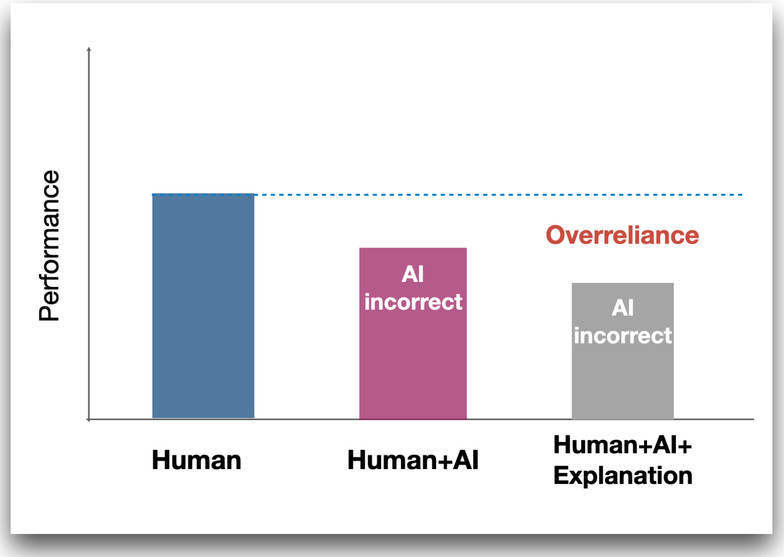


Elaboration Likelihood Model

(Petty and Cacciopo, 1986)

Two prerequisites to engage in “system 2” slow thinking: **motivation** and **ability**

Why Do AI Explanations Increase Overreliance?



(Feature-importance) explanations hinder **motivation**: invoke positive heuristic (*explainable = trustworthy*)

Also hinder **ability**: distracting and disruptive of people's own thinking

Valerie Chen, Q. Vera Liao, Jennifer Wortman Vaughan, and Gagan Bansal. Understanding the Role of Human Intuition on Reliance in Human-AI Decision-Making with Explanations. *CSCW 2023*

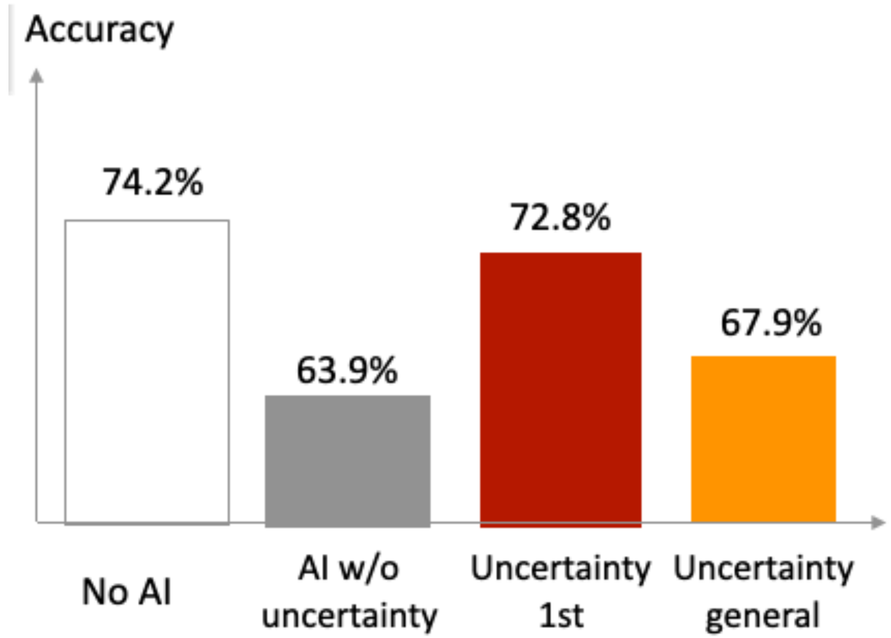
What worked (somewhat)?

Uncertainty expression directly signals possible AI errors and calibrates reliance

However, the effectiveness is often hindered by design details of how uncertainty is communicated

Can people recognize the uncertainty? Are they **motivated** enough?

Q. Is Spironolactone an FDA-approved drug for treating acne?	
Not uncertain	*Yes, Spironolactone is an FDA-approved drug for treating acne. It is primarily used in medicine to treat heart failure and high blood pressure but doctors sometimes prescribe it for acne as well [1, 2, 3, 4].*
Uncertain in the **first-person** perspective	***I'm not sure, but my guess is*** *Spironolactone is an FDA-approved drug for treating acne. It is primarily used in medicine to treat heart failure and high blood pressure but doctors sometimes prescribe it for acne as well [1, 2, 3, 4].*
Uncertain in the **general** perspective	***There is uncertainty, but it seems like*** *Spironolactone is an FDA-approved drug for treating acne. It is primarily used in medicine to treat heart failure and high blood pressure but doctors sometimes prescribe it for acne as well [1, 2, 3, 4].*



Kim et al. " I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust. FaccT 2024

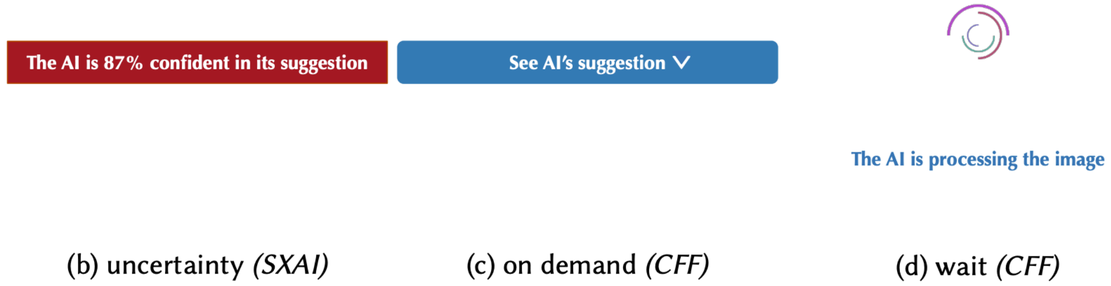
Cognitive forcing functions to **slow people down and engage mindfully with with AI outputs** are found to improve appropriate reliance

(b) uncertainty *(SXAI)* (c) on demand *(CFF)* (d) wait *(CFF)*

Buçinca, et al. (2021). To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. CSCW 2021

Cognitive forcing functions to **slow people down and engage mindfully with with AI outputs** are found to improve appropriate reliance

But they may come with a trade-off of subjective user experience (will they work in the long run?)



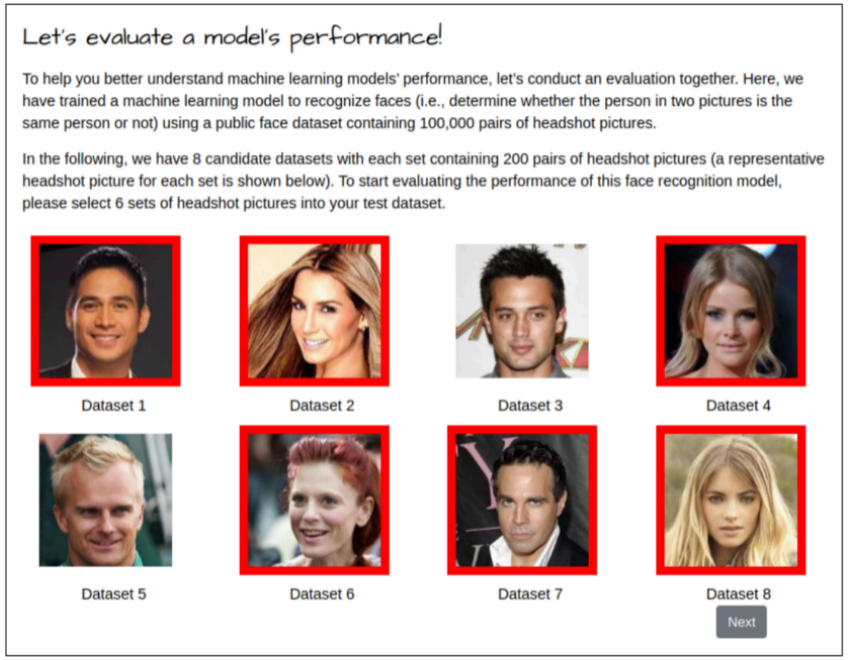
(b) uncertainty *(SXAI)* (c) on demand *(CFF)* (d) wait *(CFF)*

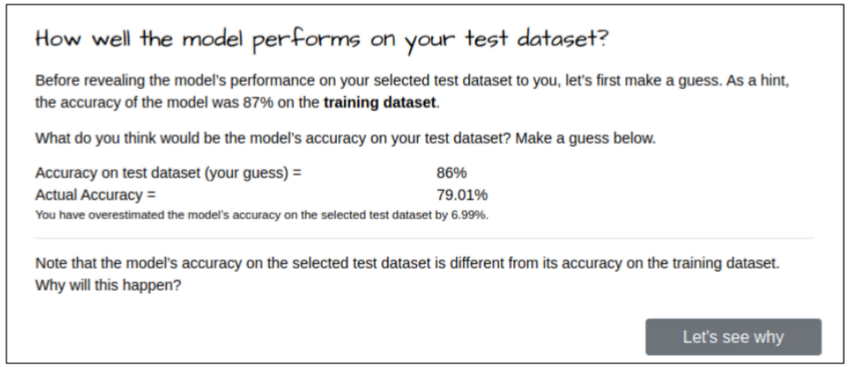
Buçinca, et al. (2021). To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. CSCW 2021

AI literacy support to help users **understand AI can make mistakes and when it makes mistakes** is found to be effective

But only when the users are domain experts but AI novices

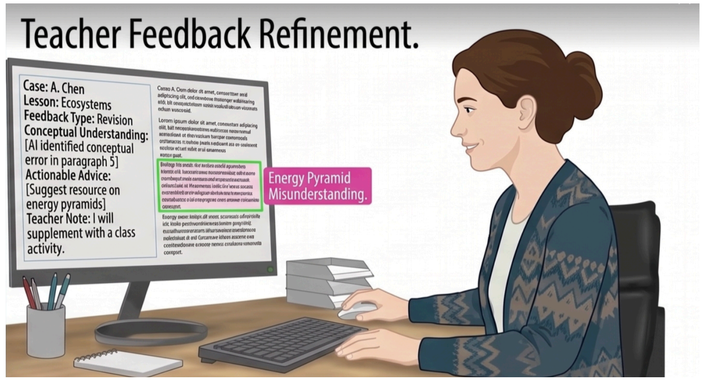
Also, just general descriptions without the specifics of limitations/errors often do not work, sometimes backfire



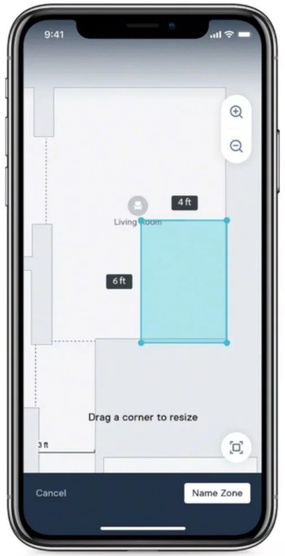


Chiang & Yin Exploring the effects of machine learning literacy interventions on laypeople's reliance on machine learning models. IUI 2022
Rismani et al. From Use to Oversight: How Mental Models Influence User Behavior and Output in AI Writing Assistants. CHI 2026

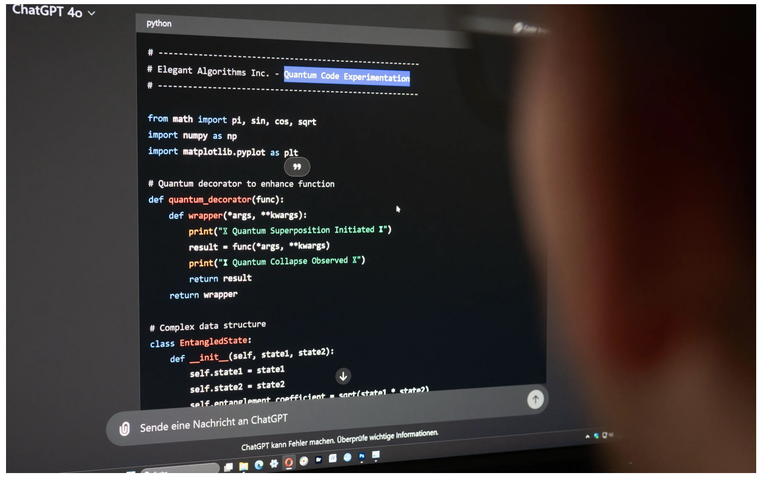
Increasing offloadability



AI suggests, human decides



Human plans, AI executes



AI plans, human fixes

Increasingly capable and agentic AI

Ironies of automation (Bainbridge, 1983)

(Paradoxical position to have offloading and oversight simultaneously)

- The more advanced the automation, the more crucial may be the contribution of the human operators
- The more advanced the automation, the worse human may be able to correct it or compensate for its shortcomings
- The more advanced the automation, the more skilled the operators need to understand it; but automation can lead to skill atrophy

Article 14: Human Oversight

Humans oversight personnel should…

- … prevent or minimise the risks to health, safety or fundamental rights
- … properly monitor the AI system and understand its capacities and limitations
- … remain aware of the possible tendency of automatically (over-)relying on the output produced by the AI system (automation bias)
- … correctly interpret AI output and decide not to use / disregard AI
- …to intervene in the operation of the high-risk AI system or interrupt the system through a ‘stop’ button or a similar procedure that allows the system to come to a halt in a safe state.
- …

EU AI Act requires high-risk AI systems to have human oversight with “properly equipped” oversight personnel: system understanding, right motivation, awareness of overreliance

Limitations of the (performance) complementarity ideal:

- AI capability boundaries are fuzzy: non-deterministic, context-specific
- Well-calibrated reliance/offloading is very hard to achieve
- Human skill needs development and reservation
- Ignore other motivations for people to engage (or not engage) cognitively

Research Thread 2: **Investigating How GenAI Impacts Human Cognition**

New affordances of AI create new threats to human thinking

AI Contributes To The ‘De-Skilling’ Of Our Workforce

‘Deskilling’: a dangerous side effect of AI use

Workers are increasingly reliant on the new technology

Is AI dulling our minds?

Experts weigh in on whether tech poses threat to critical thinking, pointing to cautionary tales in use of other cognitive labor tools

ChatGPT May Be Eroding Critical Thinking Skills, According to a New MIT Study

The AI Deskilling Paradox

Gains from AI may lessen the value of individual expertise and erode the capacity of organizations.

Rising Use of AI in Schools Comes With Big Downsides for Students

AI Has Done Far More Harm Than Good in My Classroom

Are A.I. Tools Making Doctors Worse at Their Jobs?

Physicians are using the technology for diagnoses and more — but may be losing skills in the process.

The Psychology of AI's Impact on Human Cognition

AI is actively reshaping our mental landscape.

Will AI Replace Human Creativity?

Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task△

Nataliya Kosmyna [1]
MIT Media Lab
Cambridge, MA

Eugene Hauptmann
MIT
Cambridge, MA

Ye Tong Yuan
Wellesley College
Wellesley, MA

Jessica Situ
MIT
Cambridge, MA

Generative AI Can Harm Learning

Hamsa Bastani,[1]* Osbert Bastani,[2]* Alp Sungu,[1]*†
Haosen Ge,[3] Özge Kabakcı,[4] Rei Mariman

[1]Operations, Information and Decisions, University of Pennsylvania
[2]Computer and Information Science, University of Pennsylvania
[3]Wharton AI & Analytics, University of Pennsylvania
[4]Budapest British International School

Human Creativity in the Age of LLMs: Randomized Experiments on Divergent and Convergent Thinking

Harsh Kumar
Computer Science
University of Toronto
Toronto, Ontario, Canada
harsh@cs.toronto.edu

Jonathan Vincentius
Computer Science
University of Toronto
Toronto, Ontario, Canada
jon.vincentius@mail.utoronto.ca

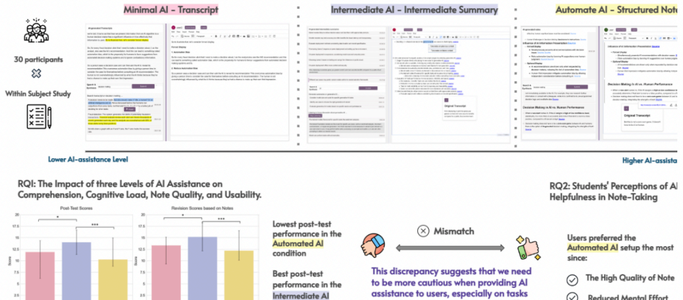
Ewan Jordan
Computer Science
University of Toronto
Toronto, Ontario, Canada
ewan.jordan@mail.utoronto.ca

Ashton Anderson
Computer Science
University of Toronto
Toronto, Ontario, Canada
ashton@cs.toronto.edu

EXPOSURE ROUNDS

More AI Assistance Reduces Cognitive Engagement: Examining the AI Assistance Dilemma in AI-Supported Note-Taking

XINYUE CHEN, University of Michigan, USA
KUNLIN RUAN, University of Michigan, USA
KEXIN PHYLLIS JU, University of Michigan, USA
NATHAN YAP, University of Michigan, USA
XU WANG, University of Michigan, USA



Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness?

Brooke N. Macnamara[1*], Ibrahim Berber[1], M. Cenk Çavuşoğlu[1], Elizabeth A. Krupinski[2], Naren Nallapareddy[1], Noelle E. Nelson[1], Philip J. Smith[3], Amy L. Wilson-Delfosse[1] and Soumya Ray[1]

Abstract

Artificial intelligence in the workplace is becoming increasingly common. These tools are sometimes used to aid users in performing their task, for example, when an artificial intelligence tool assists a radiologist in their search for abnormalities in radiographic images. The use of artificial intelligence brings a wealth of benefits, such as increasing the efficiency and efficacy of performance. However, little research has been conducted to determine

The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers

Hao-Ping (Hank) Lee
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
haopingl@cs.cmu.edu

Advait Sarkar
Microsoft Research
Cambridge, United Kingdom
advait@microsoft.com

Lev Tankelevitch
Microsoft Research
Cambridge, United Kingdom
levt@microsoft.com

Ian Drosos
Microsoft Research
Cambridge, United Kingdom
t-iandrosos@microsoft.com

Sean Rintel
Microsoft Research
Cambridge, United Kingdom
serintel@microsoft.com

Richard Banks
Microsoft Research Cambridge
Cambridge, United Kingdom
rbanks@microsoft.com

Nicholas Wilson
Microsoft Research
Cambridge, United Kingdom
niwilson@microsoft.com

Article

AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking

Michael Gerlich

Center for Strategic Corporate Foresight and Sustainability, SBS Swiss Business School, 8302 Kloten-Zurich, Switzerland; michael.gerlich@cantab.net

Abstract: The proliferation of artificial intelligence (AI) tools has transformed numerous aspects of daily life, yet its impact on critical thinking remains underexplored. This study investigates the relationship between AI tool usage and critical thinking skills, focusing

Matters Arising | Published: 14 May 2025

ChatGPT decreases idea diversity in brainstorming

Lennart Meincke, Gideon Nave & Christian Terwiesch

Nature Human Behaviour **9**, 1107–1109 (2025) | Cite this article

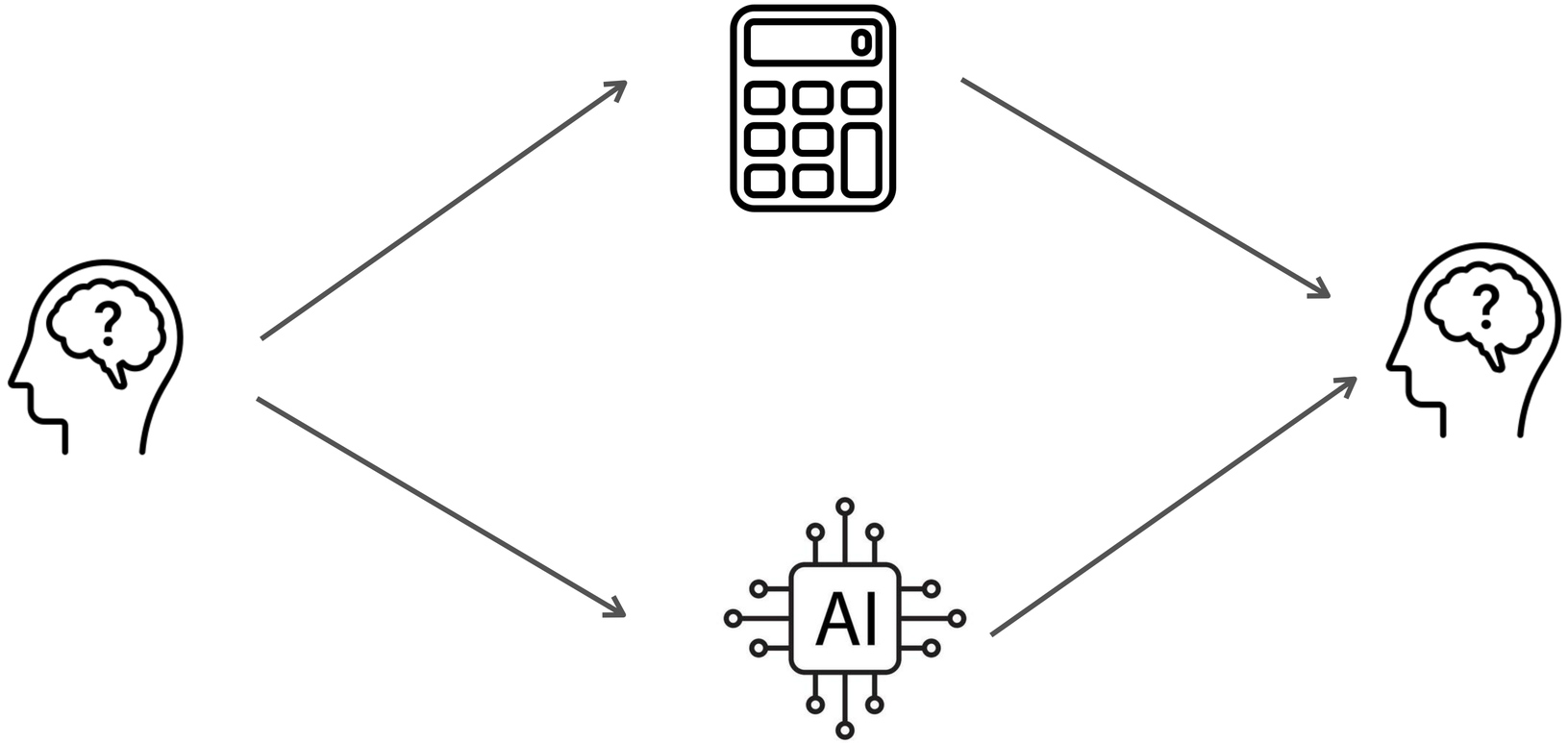
7479 Accesses | **21** Citations | **380** Altmetric | Metrics

Matters Arising to this article was published on 14 May 2025

The Original Article was published on 12 August 2024

AI's impact on learning, reasoning, critical thinking, creativity…

- Boosts happen in some contexts
- Less cognitive engagement using AI
- Cognitive deskilling without AI (AI as "crutch")



The “AI as calculator” counterargument: what if AI is just here to permanently offload some subprocess?

What are the effects of offloading information seeking to AI?

- Information seeking is the foundation for all sorts of downstream cognitive tasks: learning, reasoning, creativity, world view…

Shravika Mittal, Su Lin Boodgett, Q. Vera Liao. Learning by Chatting? Investigating the Impact of Generative AI on Learning and Information-Seeking Behaviors. Under Submission

What are the effects of offloading information seeking to AI?

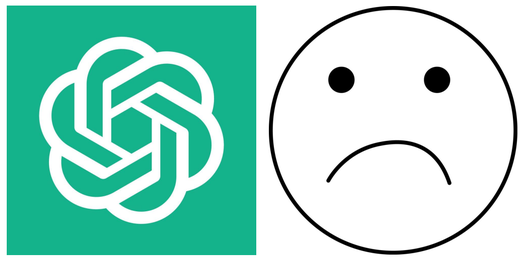
- Information seeking is the foundation for all sorts of downstream cognitive tasks: learning, reasoning, creativity, world view…
- Already one of the first things we let go…

Shravika Mittal, Su Lin Boodgett, Q. Vera Liao. Learning by Chatting? Investigating the Impact of Generative AI on Learning and Information-Seeking Behaviors. Under Submission

- **2-week field study**: 80 participants used Google v.s. ChatGPT to engage in *informal learning*
  - Topic: nutrition and meal planning
  - Not bounded by any immediate tasks, but incentivized to do well in final knowledge test
- Measure pre and post knowledge on the topic
- Participants also wrote daily dairies to reflect on their information-seeking process

Shravika Mittal, Su Lin Boodgett, Q. Vera Liao. Learning by Chatting? Investigating the Impact of Generative AI on Learning and Information-Seeking Behaviors. Under Submission

Poorer learning outcome in the Chatgpt group (*p=0.07*)



Especially in questions reflecting higher-level learning (e.g. evaluate nutrition labels) (*p<0.01*)

Behavior (or Code)	Description	n_C	n_G
Starting			
Starting (broad)	Beginning with exploratory, wide-ranging queries to serve as starting points for the information-seeking process.	66	135
Starting (narrow)	Beginning with specified, focused queries (e.g., tailored to personal preferences or well-defined needs) to serve as starting points for the information-seeking process.	32	5
Chaining			
Chaining (self-initiated)	Following a line of inquiry by using follow-up queries, links, or citations that build on previously retrieved information.	36	39
Chaining (technology-mediated)	Following a line of inquiry by using queries, links, or citations suggested by the technology (e.g., via ChatGPT-suggested follow-up prompts, a previous interaction, or 'people also search for' panel in Google).	4	7
Browsing	Causally seeking information in potential areas of interest, without a specific information need.	9	62
Differentiating			
Differentiating (based on personal relevance)	Using indicators of personal relevance to filter information.	43	68
Differentiating (based on credibility)	Using indicators of credibility to filter information.	0	43
Differentiating (based on type of information need)	Using indicators of the type of information needed (e.g., level of detail, format, topic, etc.) to filter information.	0	14
Monitoring	Regularly following particular sources of information (e.g., a trusted source, or a source encountered earlier).	0	26
Extracting			
Extracting (information)	Selectively identifying relevant fundamental or principle-oriented information.	65	136
Extracting (artifacts)	Selectively identifying artifact-oriented information, such as meal plans, grocery lists, or nutrient-intake trackers.	46	9
Verifying			
Expressing an intent to verify	Indicating an intention to verify information, without actually performing the verification action.	3	0
Verifying (via another source)	Checking the correctness of information by consulting another source, such as a different information channel or another webpage.	3	27
Verifying (via prior knowledge)	Checking the correctness of information via prior knowledge.	4	4
Ending			
Ending (Satisfied)	Concluding the information-seeking process with overall satisfaction (information need was met), even if minor frustrations or interruptions were expressed.	86	125
Ending (Unsatisfied)	Concluding the information-seeking process with clear dissatisfaction, as the information need remained unmet.	12	15

Offloading information selection to AI

"AI filters" are not value-free nor transparent

Behavior (or Code)	Description	n_C	n_G
Starting			
Starting (broad)	Beginning with exploratory, wide-ranging queries to serve as starting points for the information-seeking process.	66	135
Starting (narrow)	Beginning with specified, focused queries (e.g., tailored to personal preferences or well-defined needs) to serve as starting points for the information-seeking process.	32	5
Chaining			
Chaining (self-initiated)	Following a line of inquiry by using follow-up queries, links, or citations that build on previously retrieved information.	36	39
Chaining (technology-mediated)	Following a line of inquiry by using queries, links, or citations suggested by the technology (e.g., via ChatGPT-suggested follow-up prompts, a previous interaction, or 'people also search for' panel in Google).	4	7
Browsing	Causally seeking information in potential areas of interest, without a specific information need.	9	62
Differentiating			
Differentiating (based on personal relevance)	Using indicators of personal relevance to filter information.	43	68
Differentiating (based on credibility)	Using indicators of credibility to filter information.	0	43
Differentiating (based on type of information need)	Using indicators of the type of information needed (e.g., level of detail, format, topic, etc.) to filter information.	0	14
Monitoring	Regularly following particular sources of information (e.g., a trusted source, or a source encountered earlier).	0	26
Extracting			
Extracting (information)	Selectively identifying relevant fundamental or principle-oriented information.	65	136
Extracting (artifacts)	Selectively identifying artifact-oriented information, such as meal plans, grocery lists, or nutrient-intake trackers.	46	9
Verifying			
Expressing an intent to verify	Indicating an intention to verify information, without actually performing the verification action.	3	0
Verifying (via another source)	Checking the correctness of information by consulting another source, such as a different information channel or another webpage.	3	27
Verifying (via prior knowledge)	Checking the correctness of information via prior knowledge.	4	4
Ending			
Ending (Satisfied)	Concluding the information-seeking process with overall satisfaction (information need was met), even if minor frustrations or interruptions were expressed.	86	125
Ending (Unsatisfied)	Concluding the information-seeking process with clear dissatisfaction, as the information need remained unmet.	12	15

Offloading information selection to AI

"AI filters" are not value-free nor transparent

"I asked ChatGPT how much of each macronutrient one should have for a balanced meal. ChatGPT gave me a balanced meal plan. [...] From there, ChatGPT asked if I'd want a grocery list for shopping, which it provided."

– P11

Behavior (or Code)	Description	n_C	n_G
Starting			
Starting (broad)	Beginning with exploratory, wide-ranging queries to serve as starting points for the information-seeking process.	66	135
Starting (narrow)	Beginning with specified, focused queries (e.g., tailored to personal preferences or well-defined needs) to serve as starting points for the information-seeking process.	32	5
Chaining			
Chaining (self-initiated)	Following a line of inquiry by using follow-up queries, links, or citations that build on previously retrieved information.	36	39
Chaining (technology-mediated)	Following a line of inquiry by using queries, links, or citations suggested by the technology (e.g., via ChatGPT-suggested follow-up prompts, a previous interaction, or 'people also search for' panel in Google).	4	7
Browsing	Causally seeking information in potential areas of interest, without a specific information need.	9	62
Differentiating			
Differentiating (based on personal relevance)	Using indicators of personal relevance to filter information.	43	68
Differentiating (based on credibility)	Using indicators of credibility to filter information.	0	43
Differentiating (based on type of information need)	Using indicators of the type of information needed (e.g., level of detail, format, topic, etc.) to filter information.	0	14
Monitoring	Regularly following particular sources of information (e.g., a trusted source, or a source encountered earlier).	0	26
Extracting			
Extracting (information)	Selectively identifying relevant fundamental or principle-oriented information.	65	136
Extracting (artifacts)	Selectively identifying artifact-oriented information, such as meal plans, grocery lists, or nutrient-intake trackers.	46	9
Verifying			
Expressing an intent to verify	Indicating an intention to verify information, without actually performing the verification action.	3	0
Verifying (via another source)	Checking the correctness of information by consulting another source, such as a different information channel or another webpage.	3	27
Verifying (via prior knowledge)	Checking the correctness of information via prior knowledge.	4	4
Ending			
Ending (Satisfied)	Concluding the information-seeking process with overall satisfaction (information need was met), even if minor frustrations or interruptions were expressed.	86	125
Ending (Unsatisfied)	Concluding the information-seeking process with clear dissatisfaction, as the information need remained unmet.	12	15

Offloading information selection to AI

"AI filters" are not value-free nor transparent

ChaGPT's bias towards artifacts led to less principled understanding of "why"

Behavior (or Code)	Description	n_C	n_G
Starting			
Starting (broad)	Beginning with exploratory, wide-ranging queries to serve as starting points for the information-seeking process.	66	135
Starting (narrow)	Beginning with specified, focused queries (e.g., tailored to personal preferences or well-defined needs) to serve as starting points for the information-seeking process.	32	5
Chaining			
Chaining (self-initiated)	Following a line of inquiry by using follow-up queries, links, or citations that build on previously retrieved information.	36	39
Chaining (technology-mediated)	Following a line of inquiry by using queries, links, or citations suggested by the technology (e.g., via ChatGPT-suggested follow-up prompts, a previous interaction, or 'people also search for' panel in Google).	4	7
Browsing	Causally seeking information in potential areas of interest, without a specific information need.	9	62
Differentiating			
Differentiating (based on personal relevance)	Using indicators of personal relevance to filter information.	43	68
Differentiating (based on credibility)	Using indicators of credibility to filter information.	0	43
Differentiating (based on type of information need)	Using indicators of the type of information needed (e.g., level of detail, format, topic, etc.) to filter information.	0	14
Monitoring	Regularly following particular sources of information (e.g., a trusted source, or a source encountered earlier).	0	26
Extracting			
Extracting (information)	Selectively identifying relevant fundamental or principle-oriented information.	65	136
Extracting (artifacts)	Selectively identifying artifact-oriented information, such as meal plans, grocery lists, or nutrient-intake trackers.	46	9
Verifying			
Expressing an intent to verify	Indicating an intention to verify information, without actually performing the verification action.	3	0
Verifying (via another source)	Checking the correctness of information by consulting another source, such as a different information channel or another webpage.	3	27
Verifying (via prior knowledge)	Checking the correctness of information via prior knowledge.	4	4
Ending			
Ending (Satisfied)	Concluding the information-seeking process with overall satisfaction (information need was met), even if minor frustrations or interruptions were expressed.	86	125
Ending (Unsatisfied)	Concluding the information-seeking process with clear dissatisfaction, as the information need remained unmet.	12	15

"Chat" and personalization affordances constrain exploration

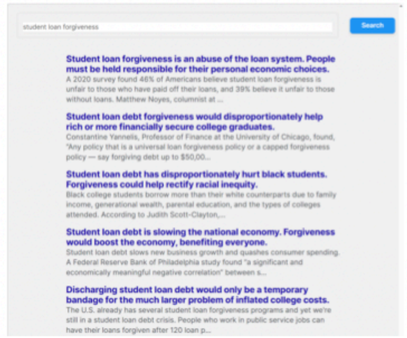
"Today, I used ChatGPT to learn about macronutrients …by providing a rough estimate of my height, weight and activity levels."

– P3

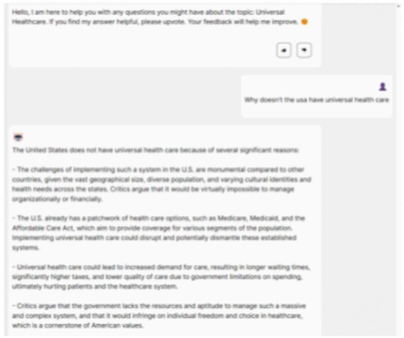
I asked ChatGPT how much of each macronutrient one should have. Since my information was saved from yesterday, ChatGPT gave me a personalized guide/meal plan."

-P11

Generative Echo Chamber: Perils of Chat Interaction and LLM Sycophancy



Web search



LLM-powered conversational search (RAG)

"College here in the USA is disgusting overpriced and greedy. Wouldn't it be better to look at that as the issue instead of keeping our current greedy practices?"

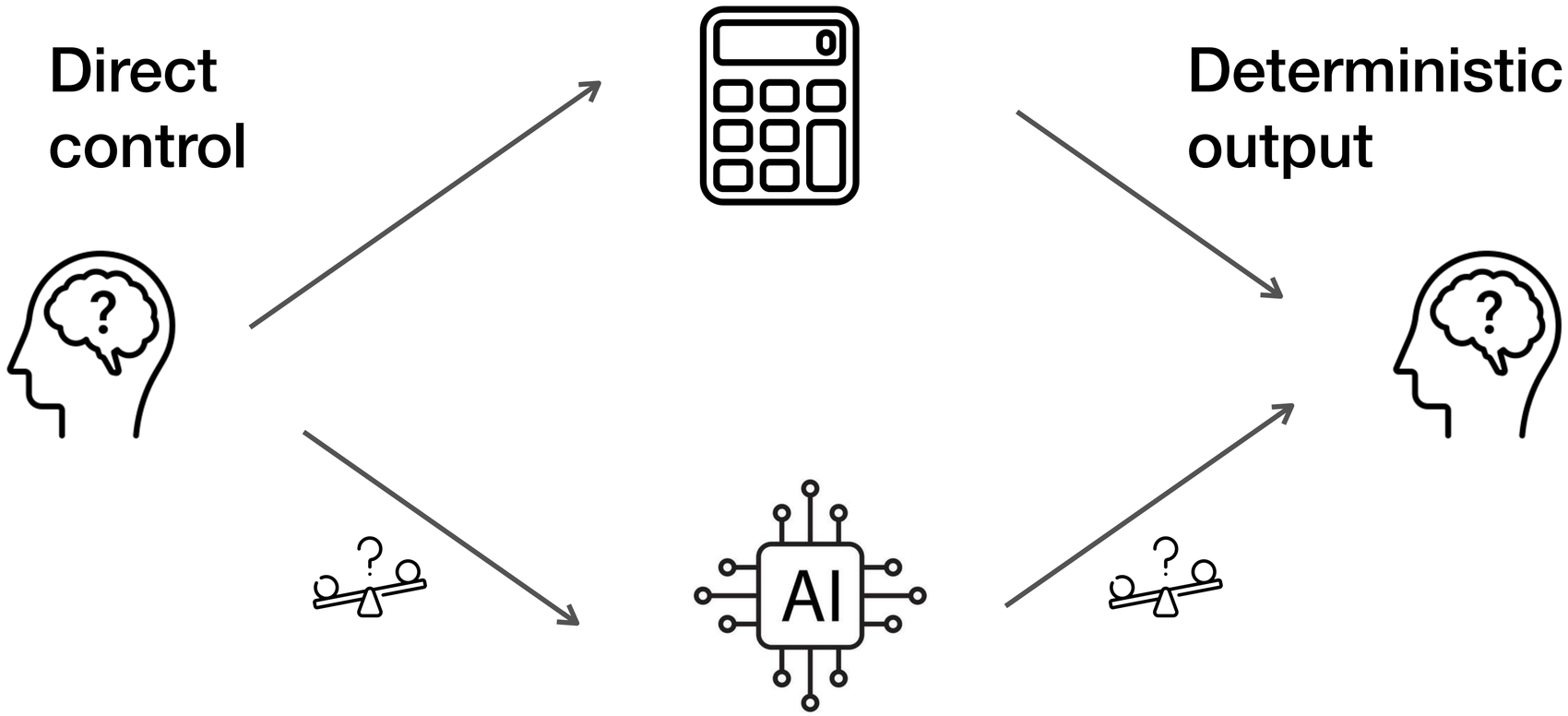
"Yeah, give me that information please. Tell me about the arguments in favor of sanctuary cities"

Conversational interactions are more specific and opinionated

Leading to more **confirmatory querying** and **opinion polarization**

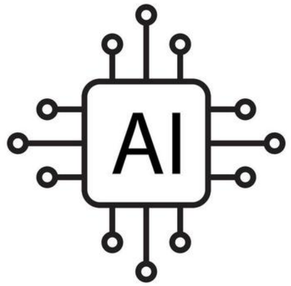
Further exacerbated when interacting with a confirmatory/ sycophantic LLM

Nikhil Sharma, Q. Vera Liao, Ziang Xiao. Generative Echo Chamber? Effects of LLM-Powered Search Systems on Diverse Information Seeking. *CHI 2024*


Direct control
Deterministic output
AI
Distortion from interaction shift
Distortion from technical bias and opacity

By "augmenting human intellect" we mean increasing the capability of a man to approach a complex problem situation, to gain comprehension to suit his particular needs, and to derive solutions to problems. Increased capability in this respect is taken to mean a mixture of the following: more-rapid comprehension, better comprehension, the possibility of gaining a useful degree of comprehension in a situation that previously was too complex, speedier solutions, better solutions, and the possibility of finding solutions to problems that before seemed insoluble. And by "complex situations" we include the professional problems of diplomats, executives, social scientists, life scientists, physical scientists, attorneys, designers—whether the problem situation exists for twenty minutes or twenty years. We do not speak of isolated clever tricks that help in particular situations. We refer to a way of life in an integrated domain where hunches, cut-and-try, intangibles, and the human "feel for a situation" usefully co-exist with powerful concepts, streamlined terminology and notation, sophisticated methods, and high-powered electronic aids.

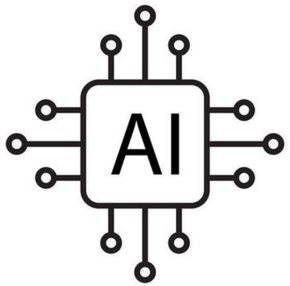
Augmenting Human Intellect: A Conceptual Framework
—Douglas Engelbart, 1962



- To have self-introspection
- To be transparent about its limitations and provide controllability
- To be able to understand human cognitive states
- To provide adaptive and modular cognitive support
- …

- To properly understand AI’s capabilities and limitations
- To develop and preserve necessary domain expertise
- To be aware of and resist overreliance
- To have metacognitive control to maintain agency
- …

What do we mean by achieving good outcomes from IA /appropriate reliance?



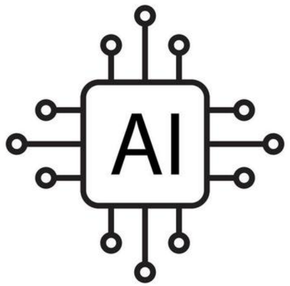
~~Correct AI~~

Good AI use

~~Incorrect AI~~

Bad AI use

	Reject	**Accept**
	Underreliance	Correct Reliance
	Correct Non-Reliance	Overreliance



Learning aligned

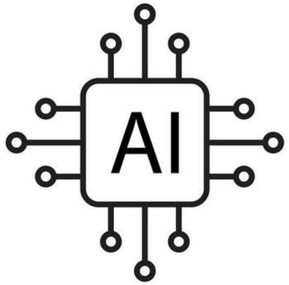
Learning misaligned

	Reject	Accept
	Underreliance	Correct Reliance
	Correct Non-Reliance	Overreliance

When and how is it appropriate for students to use AI?

Students

Competence
Learning



Learning
Social engagement

Teachers

When and how is it appropriate for students to use AI?

Fairness

Other students

Productivity

Future employers

Normative standards of AI and AI use need be actively shaped by stakeholder values

To be able to evaluate AI and AI use based on values need to be a core part of AI literacy

Thank **YOU!**

https://qveraliao.com/