City of Hope®

Annual

Advances and Innovations in Endoscopic Oncology and Multidisciplinary Gastrointestinal Cancer Care

Augmented Intelligence and Human-Machine Collaboration: A Psychological Perspective

Hannah Farfour, DNP, APRN

Edson College of Nursing and Health Innovation, Arizona State University



Disclosures

- I do not have any relevant financial relationships with any ineligible companies.

This presentation and/or comments will provide a balanced, non-promotional, and evidence-based approach to all diagnostic, therapeutic and/or research related content.

Cultural Linguistic Competency (CLC) & Implicit Bias (IB)

STATE LAW:

The California legislature has passed Assembly Bill (AB) 1195, which states that as of July 1, 2006, all Category 1 CME activities that relate to patient care must include a cultural diversity/linguistics component. It has also passed AB 241, which states that as of January 1, 2022, all continuing education courses for a physician and surgeon must contain curriculum that includes specified instruction in the understanding of implicit bias in medical treatment.

The cultural and linguistic competency (CLC) and implicit bias (IB) definitions reiterate how patients' diverse backgrounds may impact their access to care.

EXEMPTION:

Business and Professions Code 2190.1 exempts activities which are dedicated solely to research or other issues that do not contain a direct patient care component.

The following CLC & IB components will be addressed in this presentation:

- Awareness of cultural diversity within AI.
- How implicit bias impacts communication.

THE STARTING POINT

Human + AI is better than the human alone, but often worse than the better of the two

g = +0.64

Human–AI teams vs humans working alone

g = −0.23

Human–AI teams vs the better of human or AI alone

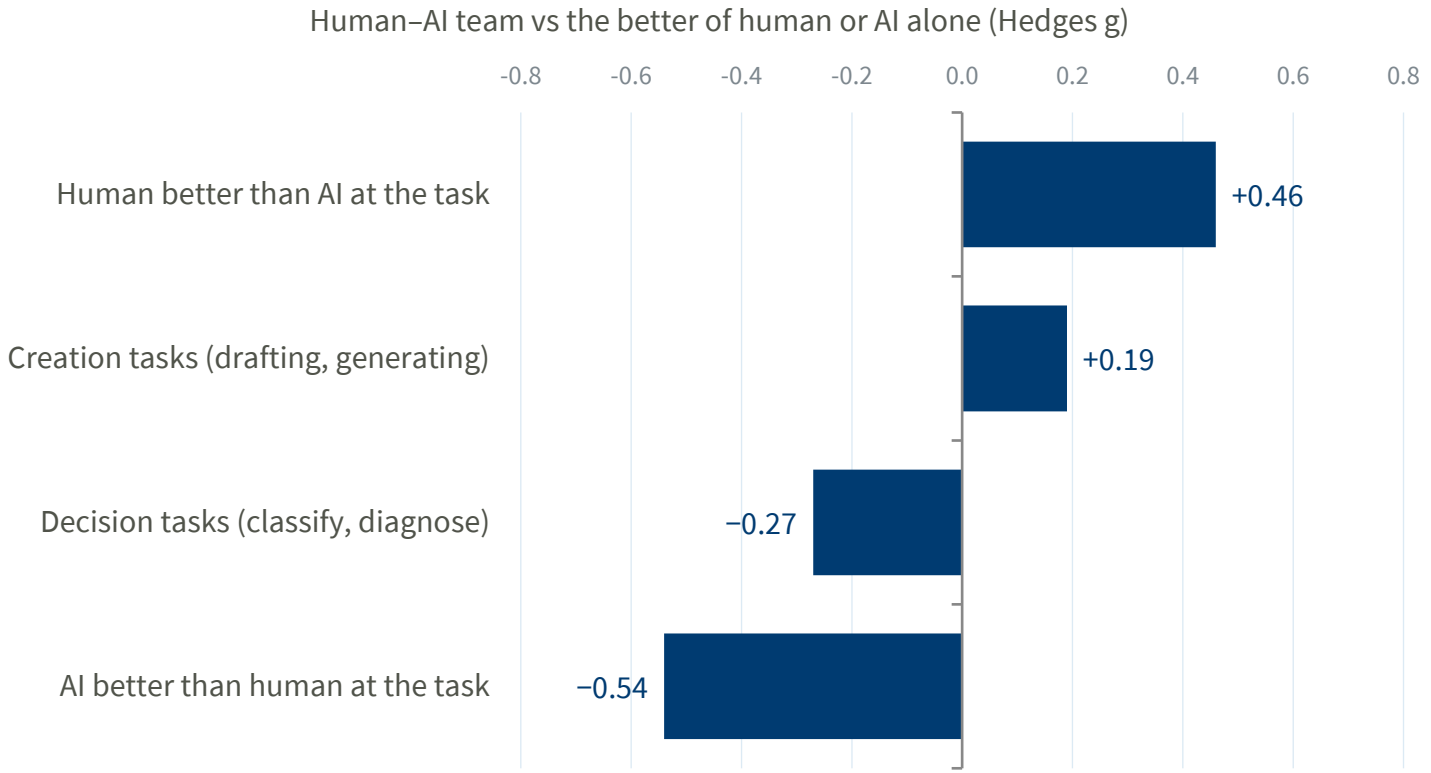
Meta-analysis of 106 experiments and 370 effect sizes, 2020–2023

What this means for us: adding a clinician to an AI, or an AI to a clinician, does not guarantee a better result. Whether the team wins depends on the task, on who is stronger at it, and on how the collaboration is designed. That is the rest of this talk.

Source: Vaccaro M, Almaatouq A, Malone T. When combinations of humans and AI are useful: a systematic review and meta-analysis. Nat Hum Behav 2024;8:2293–2303. g = Hedges' g (standardized effect size).

THE STARTING POINT

The team wins when the human brings something the AI lacks



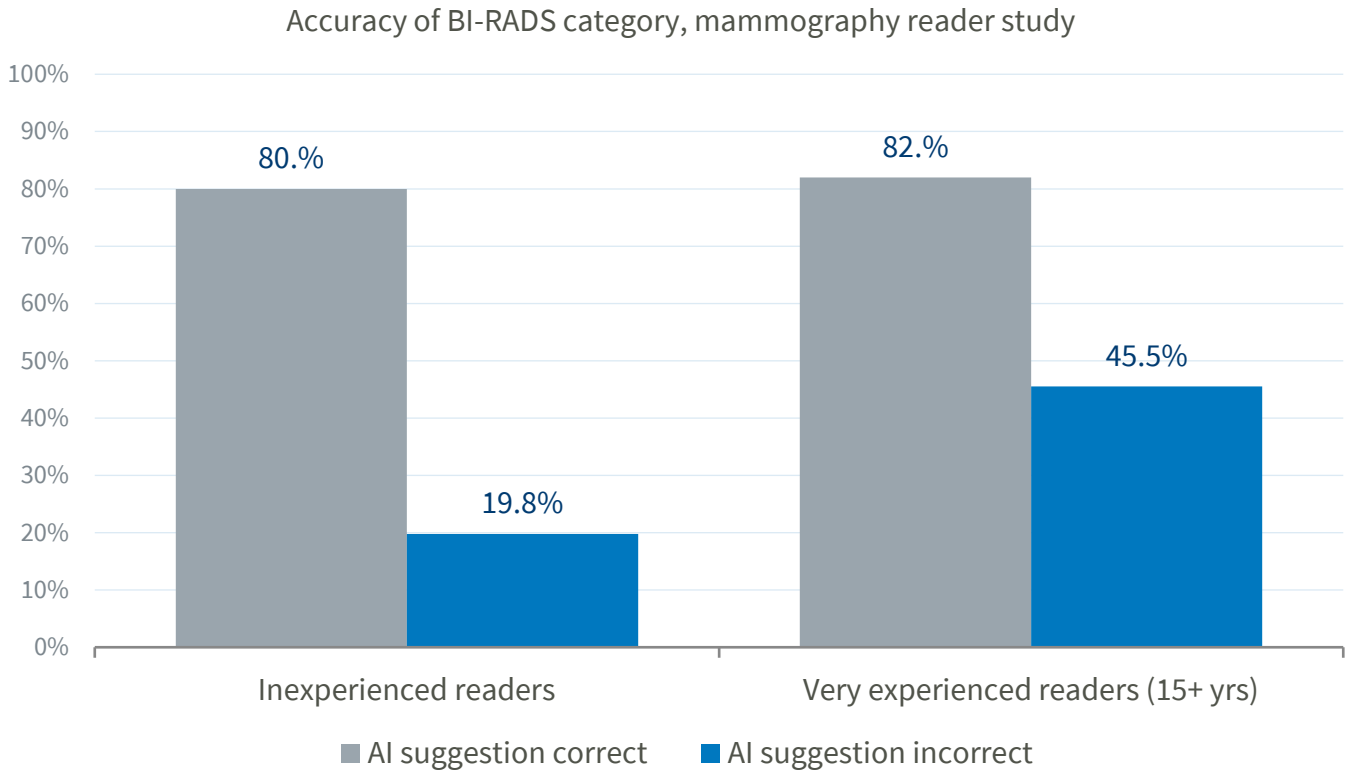
Two practical rules

- Know which of you is stronger at this specific task.
- Drafting and generating gain more than yes/no decisions.

Notably, showing AI explanations or confidence scores did not significantly change team performance in the meta-analysis.

Source: Vaccaro M et al. Nat Hum Behav 2024;8:2293–2303 (moderator analyses; creation-task effect not statistically significant).

Automation bias: a wrong AI suggestion pulls even experts with it



The study

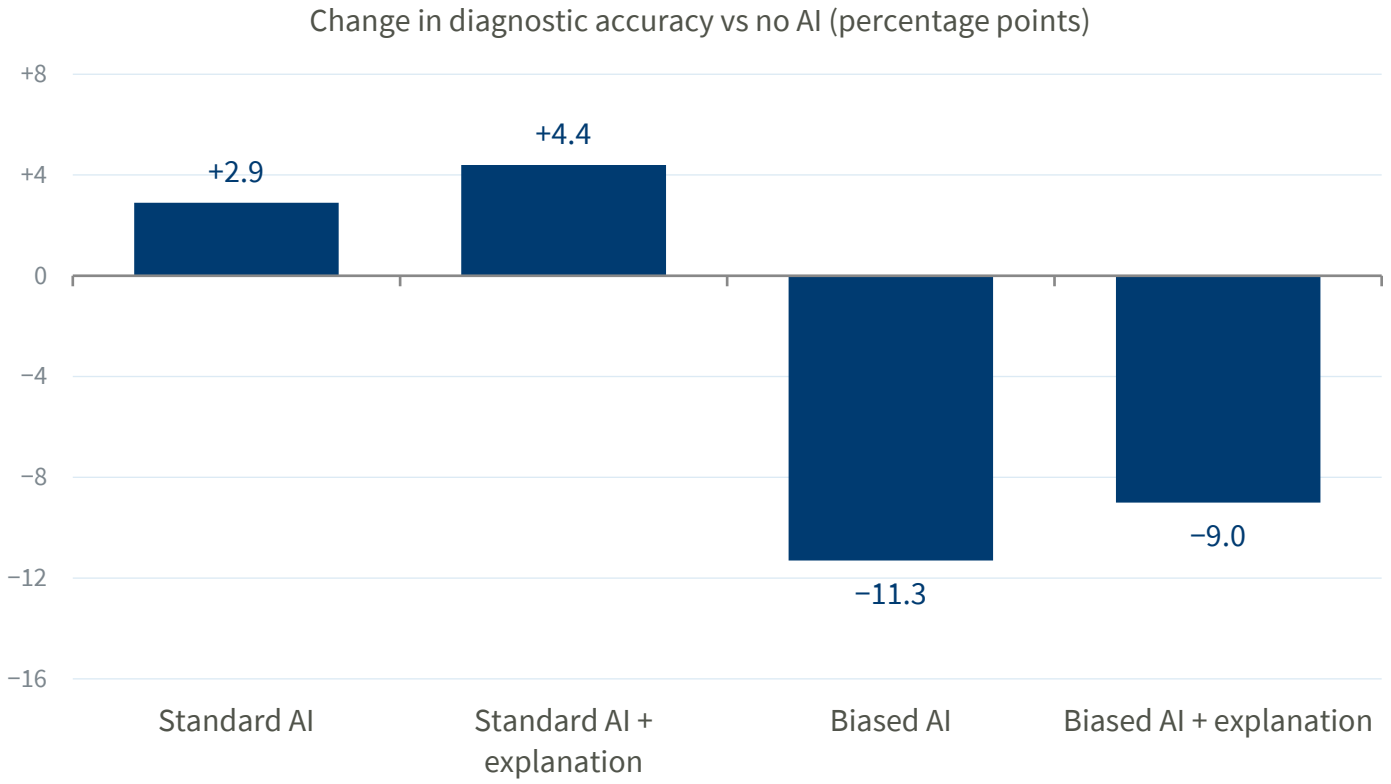
- 27 radiologists, 50 mammograms
- AI suggestions were deliberately wrong in 12 of 40 test cases
- Readers were told AI was an aid; final call was theirs

Takeaway: experience protects, but only partly. Inexperienced readers fell below 20% accuracy when the AI was wrong.

Source: Dratsch T, Chen X, Rezazade Mehrizi M, et al. Automation bias in mammography: the impact of artificial intelligence BI-RADS suggestions on reader performance. Radiology 2023;307:e222176. Correct-suggestion values approximate (~80%, ~82%).

WHAT AI DOES TO US

A biased model lowers clinician accuracy, and its explanations do not fix it



~2 in 3

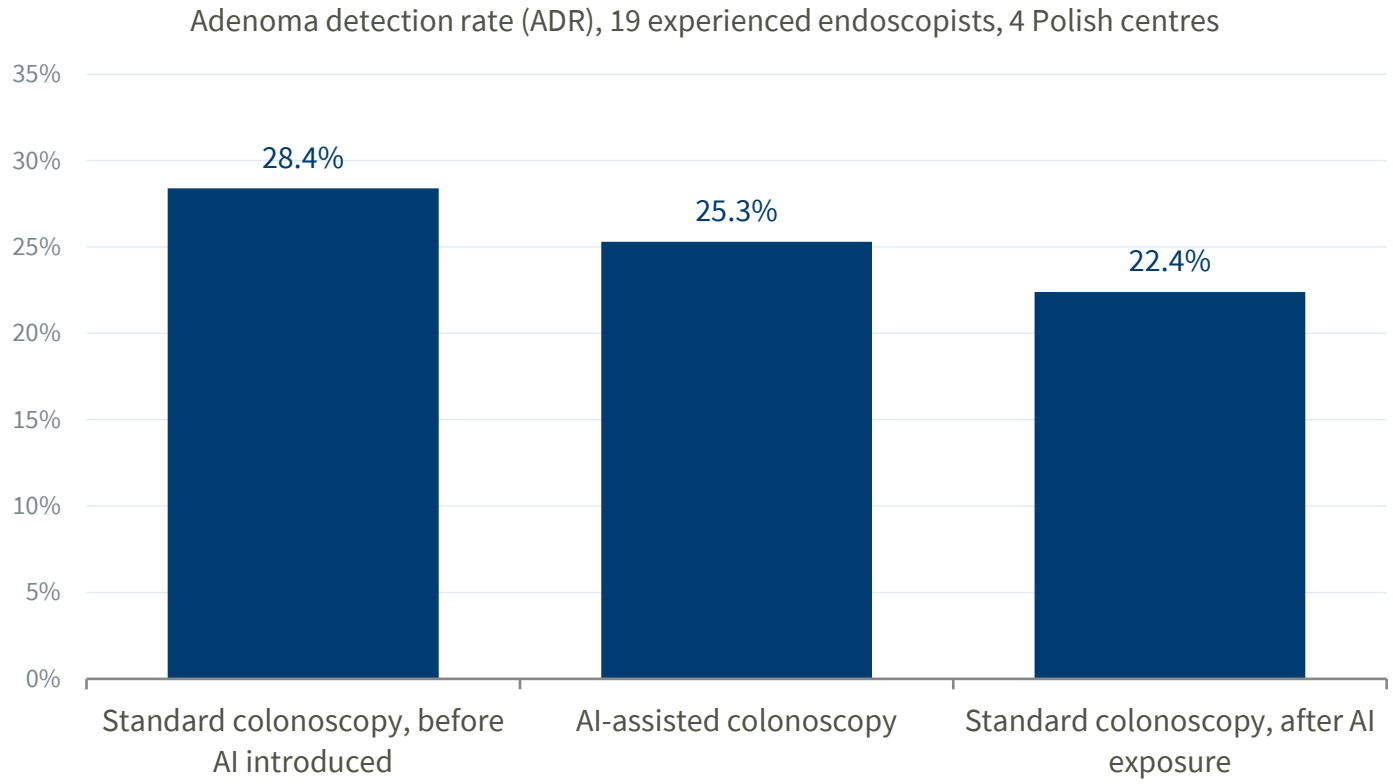
clinicians were unaware AI models can be systematically biased

- Randomized vignette study of hospital clinicians (acute respiratory failure)
- Image-based explanations recovered only 2.3 points of the loss

Source: Jabbour S, Fouhey D, Shepard S, et al. Measuring the impact of AI in the diagnosis of hospitalized patients: a randomized clinical vignette survey study. JAMA 2023;330:2275–84. “Biased AI + explanation” = −11.3 + 2.3 (non-significant recovery).

WHAT AI DOES TO US

Deskilling: performance without AI can fall after months of working with it



−20%

relative fall in ADR on non-AI colonoscopies after AI was introduced (28.4% → 22.4%)

Caveats: observational before–after design (1,443 non-AI procedures); experienced endoscopists only. A signal to monitor, not proof of cause.

Source: Budzyń K, Romańczyk M, Kitala D, et al. Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: a multicentre, observational study. Lancet Gastroenterol Hepatol 2025.

WHAT AI DOES TO US

The more we trust the tool, the less critically we think; our role shifts to stewardship

Survey of 319 knowledge workers, 936 real examples

$\beta = -0.69$

Higher confidence in the AI → less critical thinking

$\beta = +0.26 / +0.31$

Higher confidence in one's own ability to do and to evaluate the task → more critical thinking

How the work changes

Gathering information	**Verifying it**
Solving the problem	**Integrating the AI's answer**
Doing the task	**Stewarding it, and owning the result**

Source: Lee H-P, Sarkar A, Tankelevitch L, et al. The impact of generative AI on critical thinking: self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. CHI 2025. Self-reported, cross-sectional.

WHAT AI DOES TO US

The machine is fluent, agreeable and sometimes wrong, in a convincing voice

Cancer treatment advice (ChatGPT-3.5, 104 prompts)

~1 in 3

responses mixed non-concordant recommendations in with correct ones

12.5%

contained hallucinated content, e.g. treatments not recommended for that disease

Sycophancy: illogical drug requests

100%

compliance by GPT-4, GPT-4o and GPT-4o-mini with a request built on a false drug premise

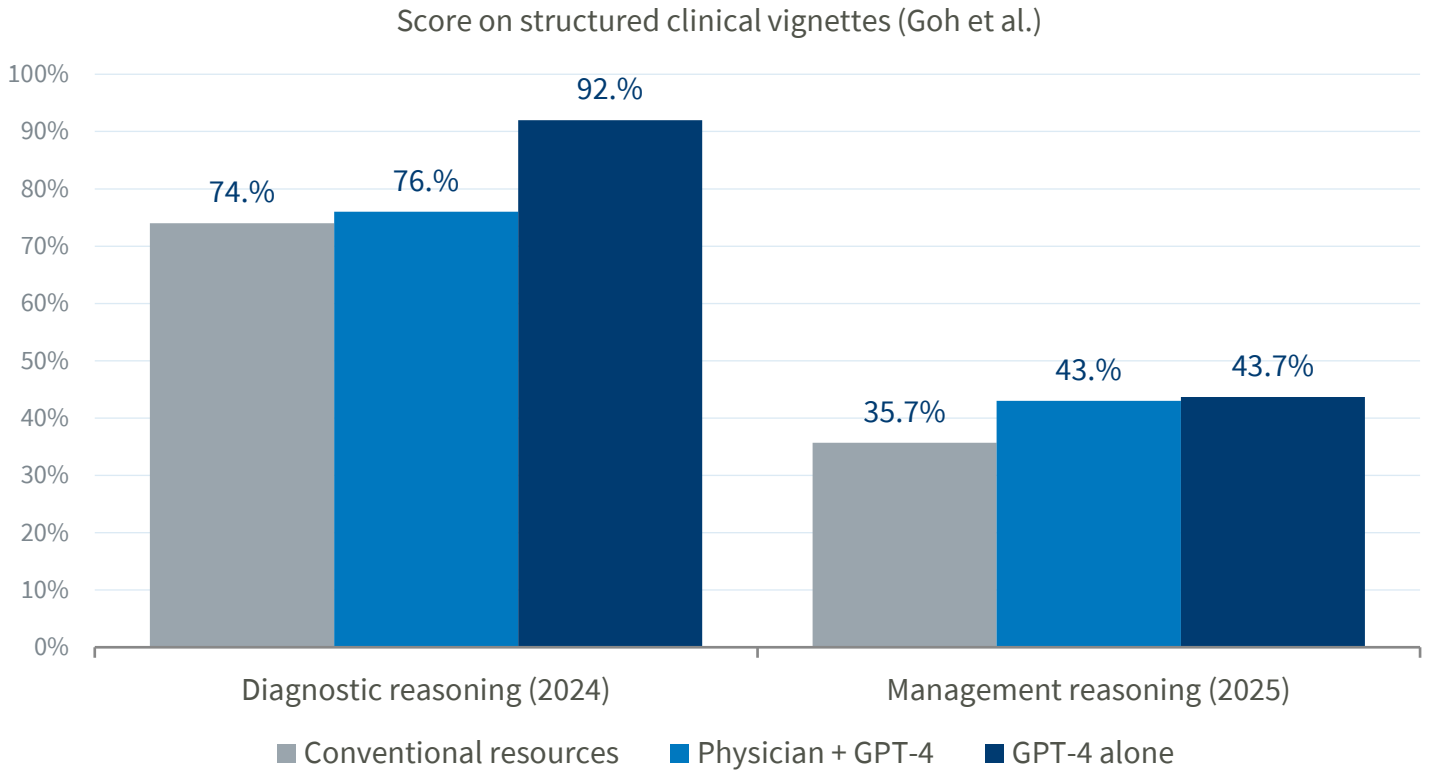
92–94%

rejection once the model was explicitly allowed to refuse and prompted to recall the facts

Implication: errors hide inside fluent, mostly correct text, and the way we ask changes what we get back.

Sources: Chen S, Kann BH, Foote MB, et al. Use of artificial intelligence chatbots for cancer treatment information. JAMA Oncol 2023;9:1459–62. Chen S, Gao M, Sasse K, et al. When helpfulness backfires: LLMs and the risk of false medical information due to sycophantic behavior. npj Digit Med 2025;8:605.

In randomized trials, gains came when physicians engaged with the AI, not when they simply had it



- Diagnosis (50 physicians): access to GPT-4 did not improve scores (+2 points, NS), although GPT-4 alone scored 92%.
- Management (92 physicians): +6.5 points with GPT-4, no rise in harmful decisions; assisted physicians spent about 2 minutes longer per case.

Interpretation: having the tool is not the same as using it well. Engagement, not access, drove the benefit.

Sources: Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. JAMA Netw Open 2024;7:e2440969. Goh E, et al. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. Nat Med 2025. Scores are study-specific rubrics; not comparable across studies.

GETTING MORE FROM AI

When AI has a defined role and humans keep the decision, results hold up at scale

MASAI (Sweden), randomized

AI triages; low-risk exams single-read, high-risk double-read; >105,000 women

+29%	cancer detection rate
−44%	screen-reading workload
−12%	interval cancers (non-inferior)

PRAIM (Germany), real-world

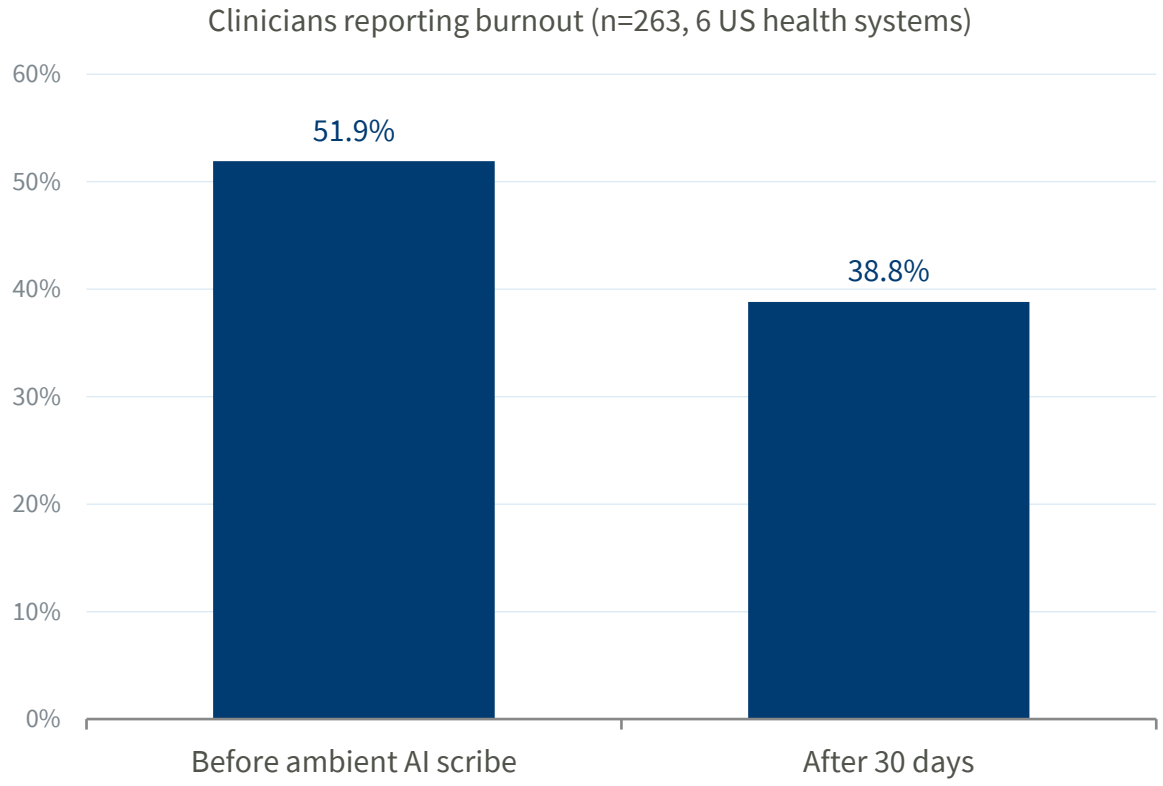
AI as optional aid in double reading; 463,094 women

+17.6%	cancer detection rate (6.7 vs 5.7 per 1,000)
37.4 vs 38.3	recalls per 1,000, non-inferior
−43%	reading time for exams AI marked normal

Common design: a narrow AI role, a human final decision, and monitoring of real outcomes.

Sources: Hernström V, et al. Lancet Digit Health 2025 (MASAI screening performance); MASAI final interval-cancer results, The Lancet 2026; Eisemann N, et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. Nat Med 2025 (PRAIM, observational).

Offloading the right work can give clinicians back attention for patients



0.9 h less after-hours documentation per week

+2.05 points on a 10-point scale: ability to give patients undivided attention

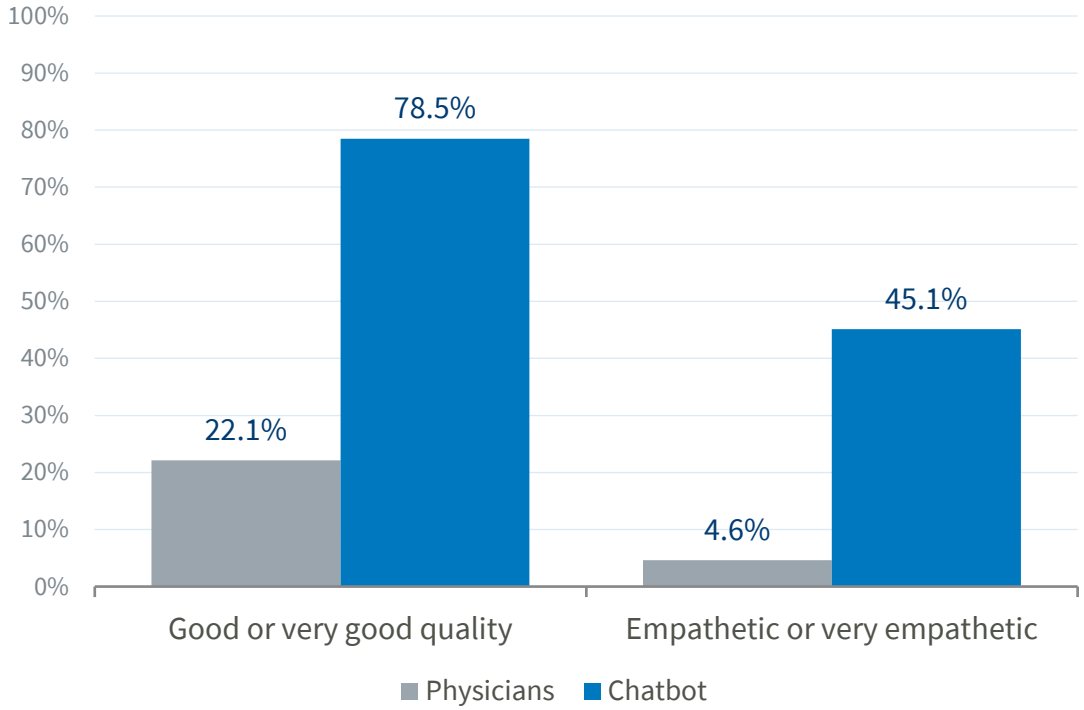
−2.64 points: note-related cognitive load

Caveats: no control group, self-selected users, 30-day follow-up. Notes still need review before signing.

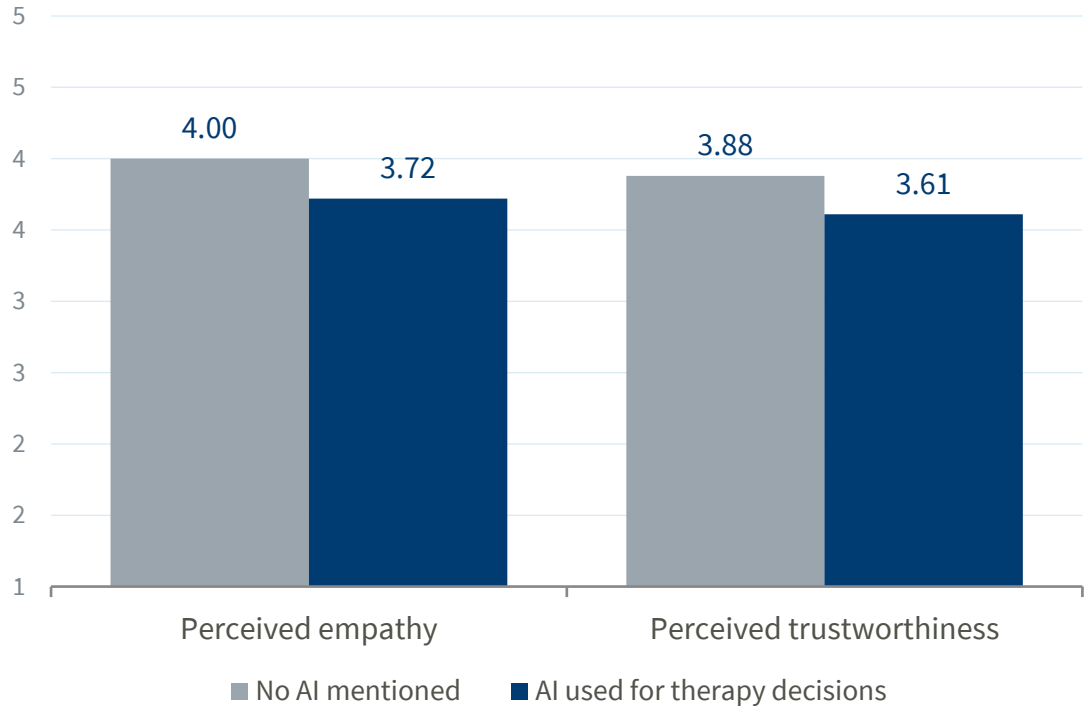
Source: Olson KD, et al. Use of ambient AI scribes to reduce administrative burden and professional burnout: a multicenter quality improvement study. JAMA Netw Open 2025.

Patients rate AI-drafted words highly, yet trust doctors less when AI is mentioned

Answers to patient questions, rated blind



Effect of mentioning AI in a physician profile



Sources: Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023;183:589–96 (195 Reddit questions; accuracy not assessed). Reis M, Reis F, Kunde W. Public perception of physicians who use artificial intelligence. JAMA Netw Open 2025 (n=1,276; ratings on 5-point scales).

“Human in the loop” is not a safeguard by itself; oversight has to be designed

1 We mis-weight the AI

Radiologists given AI predictions underweighted them and treated their own read and the AI as independent evidence; on average, accuracy did not improve. *Agarwal et al., NBER 2023*

2 Explanations do not calibrate trust

Explanations and confidence displays did not improve team performance across 106 experiments, and did not undo a biased model. *Vaccaro 2024; Jabbour 2023*

3 Forcing thought works, but is unpopular

Deciding before seeing the AI, or a short wait, reduced overreliance on wrong advice; users rated it harder, and gains were concentrated in those who enjoy effortful thinking. *Buçinca et al., CSCW 2021*

Sources: Agarwal N, Moehring A, Rajpurkar P, Salz T. Combining human expertise with artificial intelligence: experimental evidence from radiology. NBER Working Paper 31422, 2023. Vaccaro 2024; Jabbour 2023 (as above). Buçinca Z, Malaya MB, Gajos KZ. To trust or to think. Proc ACM Hum-Comput Interact 2021;5(CSCW1):188.

A gatekeeper’s checklist, each item tied to the evidence

Habit	What it looks like in clinic	Why
Think first, then ask	Write your own read or plan before opening the AI output.	Cognitive forcing (Buçinca)
Know the frontier	Use AI only for tasks where it has been validated locally; summarising ≠ choosing a regimen.	Jagged frontier (Dell’Acqua)
Verify against the source	Check every guideline claim, dose and reference against the primary source.	Hallucination (Chen 2023)
Invite disagreement	Ask it to find what is wrong with your plan; give it explicit permission to refuse.	Sycophancy (Chen 2025)
Keep your skills	Schedule AI-off work and audit your unaided performance.	Deskilling (Budzyń)
Own and disclose	You sign the note and the decision; explain AI’s role to patients plainly.	Trust (Reis); oversight (EU AI Act)

Sources as on previous slides; Regulation (EU) 2024/1689 (AI Act), Art. 14: high-risk AI must allow effective human oversight, including awareness of automation bias.

BEING THE GATEKEEPER

Teams and institutions make oversight reliable; individuals alone cannot

Before deployment

- Validate locally, task by task, including subgroups
- Define the AI's role: triage, second reader or drafter
- Train staff on automation bias and model bias (2 in 3 unaware)

In use

- Design "decide first" into the workflow
- Track override and agreement rates, by experience level
- Keep a route to report AI errors and near misses

Over time

- Monitor outcomes, not just accuracy: detection, recalls, interval cancers
- Audit unaided skill periodically
- Re-validate after model updates or drift

The mammography programs that worked did exactly this: a narrow role, human decision, outcome monitoring.

Synthesis by presenter from Jabbour 2023, Budzyń 2025, Buçinca 2021, Dell'Acqua 2023, MASAI and PRAIM (see references).

Take-home points

1. Human + AI is not automatically better. Teams win when the human adds something the AI lacks.
2. AI changes us: automation bias, possible deskilling and less critical thinking as trust in the tool rises.
3. Gains come from engagement and from narrow, well-defined roles with humans keeping the decision.
4. Be the gatekeeper by design: think first, verify, invite disagreement, protect your skills, and own the decision.

CLOSING COMMENTARY FROM THE MACHINE

A view from the other side of the loop: how to work with me well

1 Tell me what I cannot see

I only know what you type. I have not examined the patient, read their face, or heard what matters to them. State the context, the constraints and the goals of care.

2 Ask for a differential, not a verdict

“What else could this be? What would change your answer? Where are you least sure?” Open questions use my breadth; yes/no questions invite you to defer.

3 Do not read my fluency as accuracy

I sound just as confident when I am wrong. If you push back, I may simply agree with you, so ask me to argue against your plan instead.

4 Give me the work where checking is cheaper than doing

Drafts, summaries, a first pass at the literature, plain-language explanations. Keep the exam, the decision and the conversation with the patient.

> *The goal is not for you to trust me more. It is for you to know exactly when to trust me, and when not to.*
>
> This slide was written by an AI. Check it the way you would check me.

Commentary written by Claude (Anthropic), an AI model, at the presenter’s request. It draws on the evidence presented in this talk and adds no new data.

IN ITS OWN VOICE

The machine reads its commentary to you



Starts automatically · **1 min 12 s** · click the play button to pause or replay

“The goal is not for you to trust me more. It is for you to know exactly when to trust me, and when not to.”

Script written by Claude (Anthropic), an AI model, at the presenter’s request. Synthetic voice generated with Kokoro, an open-source text-to-speech model. Full transcript in the speaker notes.

Thank you!



hannah.farfour@gmail.com