

AI in medical education

Training physicians for an AI-enabled healthcare system

John R. Hayes, DO, FAAFP, FASAM
Program Director
MCW-Prevea **Green Bay** Family Medicine Residency
Associate Professor
Medical College of Wisconsin



Educating the Next Generation

**MCW is generating the health
and scientific workforce for
Wisconsin and beyond.**

Our innovative approach to academic medicine has prepared our students to **excel in their fields** and instilled a responsibility to administer **compassionate, equitable and culturally competent care.**

1,677

Medical (1,039),
graduate (458),
pharmacy (127), and
anesthesiology
assistant (53) students

20,800+

Total number of
living alumni

36
Degree
programs



3 Campuses

52%

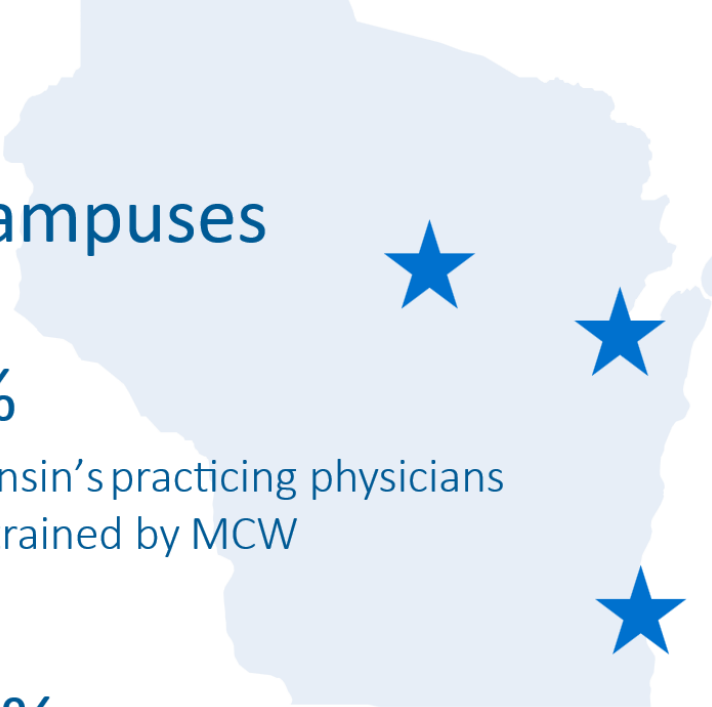
Wisconsin’s practicing physicians
were trained by MCW

100%

Green Bay and Central Wisconsin
medical students successfully match

99%

Milwaukee medical students
successfully match



Score on the United States Medical Licensing Examination

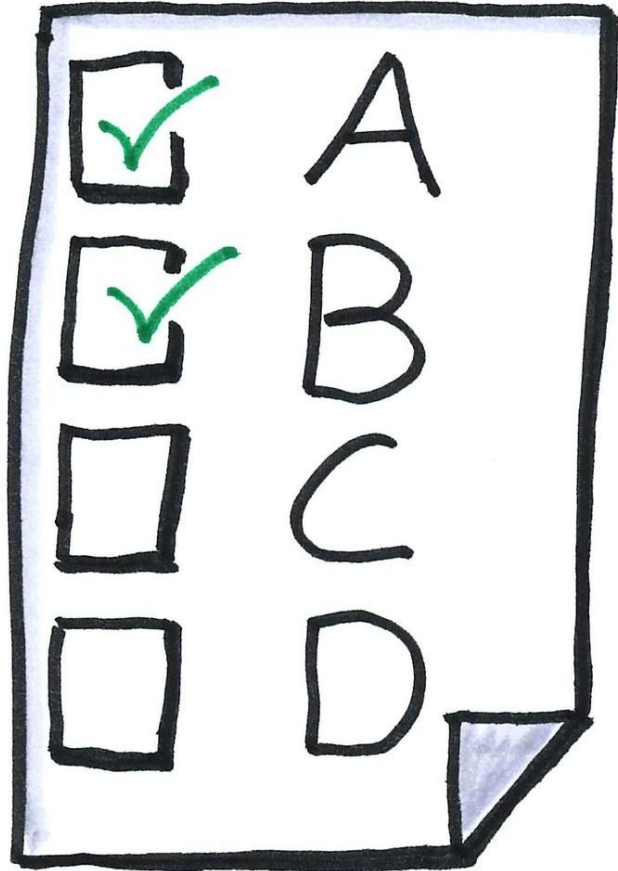
OpenEvidence GPT-5 GPT-4 GPT-4o



Do we even need to train more physicians?

Does this really matter?

Medicine IS NOT multiple choice!



Blood Pressure is high... should I:

- A. Discuss lifestyle changes
- B. Start HCTZ
- C. Lock my PT in a closet and feed them lettuce
- D. Prescribe Amoxicillin

Good News!

AI Can Not Replace Physicians...

Jagarapu et al. (2025) tested OpenEvidence against the MedXpertQA dataset

– 100 complex medical subspecialty scenarios.

"Deep Consult" Mode

41% diagnostic & therapeutic accuracy

Standard Query Mode

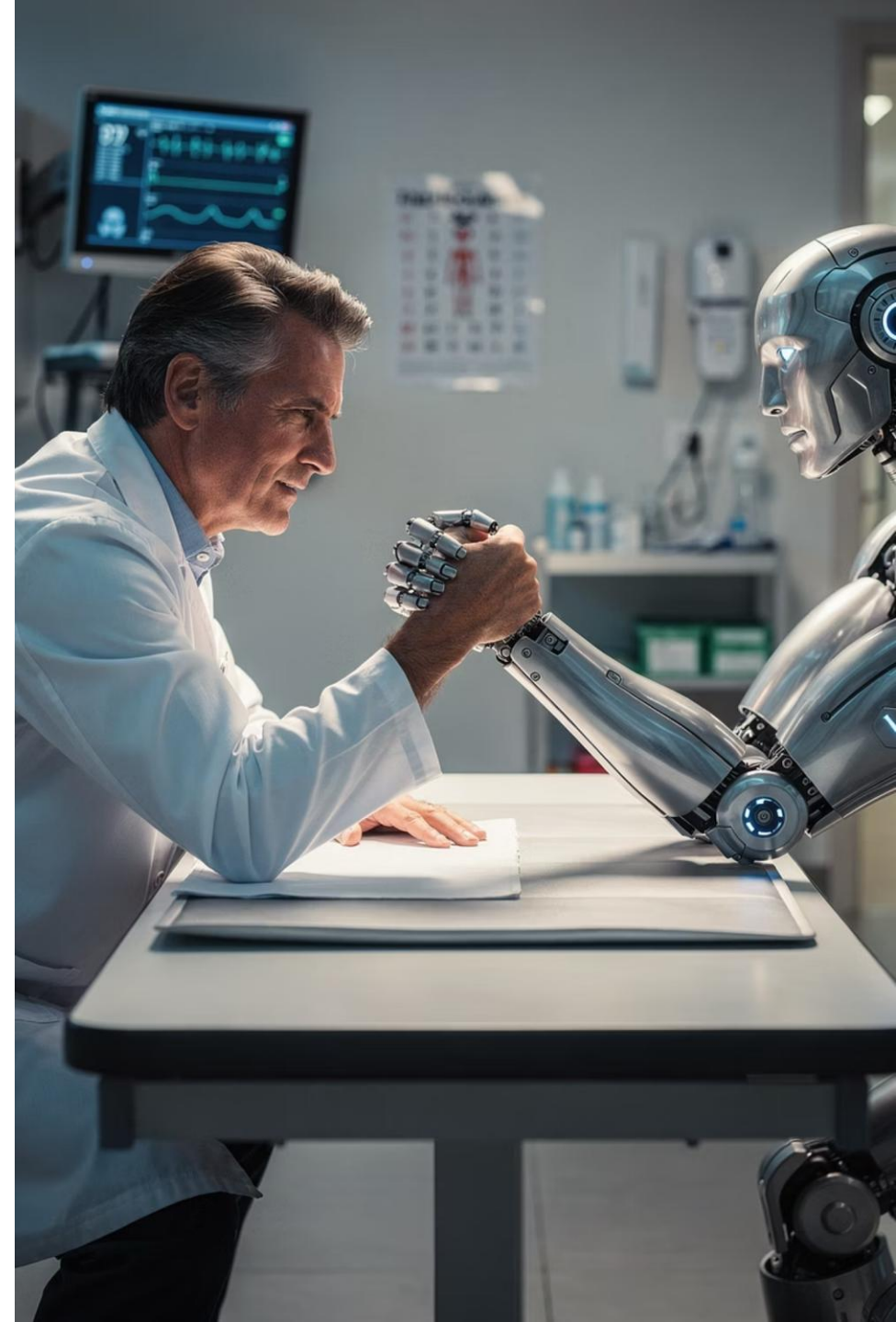
34% diagnostic & therapeutic accuracy

100% Fluent, convincing answers supplied to the physicians



⚠ Take-Home Message: Fluent, confident AI outputs do not always indicate clinical accuracy.

Physicians must be trained in CRITICAL APPRAISAL



If its so inaccurate, why should healthcare bother with AI?

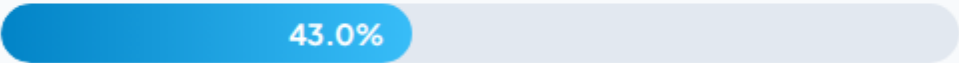
NATURE MEDICINE (2025)

Doctors paired with AI perform better than those with traditional resources.

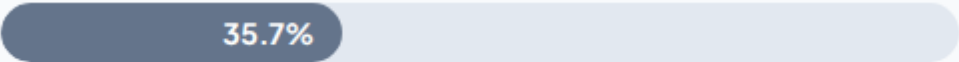
Goh E, Gallo RJ, Strong E, Weng Y, Kerman H, Freed JA, Cool JA, Kanjee Z, Lane KP, Parsons AS, Ahuja N, Horvitz E, Yang D, Milstein A, Olson APJ, Hom J, Chen JH, Rodman A. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. Nat Med. 2025 Apr;31(4):1233-1238. doi: 10.1038/s41591-024-03456-y. Epub 2025 Feb 5. Erratum in: Nat Med. 2025 Apr;31(4):1370. doi: 10.1038/s41591-025-03586-x. PMID: 39910272; PMCID: PMC12380382.

MANAGEMENT REASONING SCORE (0-100 SCALE)

Physicians + GPT-4 43.0%



Traditional Resources Only (UpToDate, Search) 35.7%



+6.5%

Statistically Significant Improvement
95% CI: 2.7 to 10.2 · $P < 0.001$

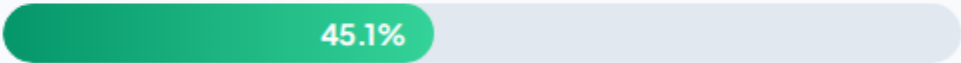
Goh E, Gallo RJ, Strong E, Weng Y, Kerman H, Freed JA, Cool JA, Kanjee Z, Lane KP, Parsons AS, Ahuja N, Horvitz E, Yang D, Milstein A, Olson APJ, Hom J, Chen JH, Rodman A. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. Nat Med. 2025 Apr;31(4):1233-1238. doi: 10.1038/s41591-024-03456-y. Epub 2025 Feb 5. Erratum in: Nat Med. 2025 Apr;31(4):1370. doi: 10.1038/s41591-025-03586-x. PMID: 39910272; PMCID: PMC12380382.

JAMA INTERNAL MEDICINE (2023)

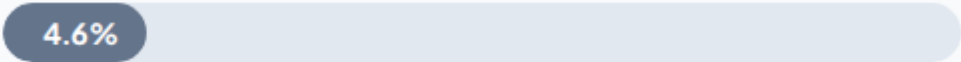
Doctors are significantly more empathetic when responding to patients.

RATED EMPATHETIC OR VERY EMPATHETIC (≥4/5)

AI-Assisted Responses 45.1%



Physicians Unassisted (Standard Workflow) 4.6%



9.8×

Higher Frequency of Empathetic Tone
Triplicate blinded clinician review · $P < 0.001$

Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, Faix DJ, Goodman AM, Longhurst CA, Hogarth M, Smith DM. Comparing Physician and Artificial Intelligence Chatbot Responses to Patient Questions Posted to a Public Social Media Forum. JAMA Intern Med. 2023 Jun 1;183(6):589-596. doi: 10.1001/jamainternmed.2023.1838. PMID: 37115527; PMCID: PMC10148230.

Training for the future:

AI is not replacing the physician

AI is changing what physicians must be excellent at



Clinical reasoning & Critical Apriasal

Generate, test, and revise a differential instead of simply accepting an answer.



Communication

Build trust, explain uncertainty, and make shared decisions with real patients.



Procedural skill

Maintain hands-on competence that cannot be delegated to a chatbot.



Implementation

Lead teams, redesign workflows, and turn evidence into reliable care.



Accountability

Use AI responsibly while the physician remains clinically and ethically accountable.



HOW IS THE MEDICAL COLLEGE OF WISCONSIN ADAPTING TO THE AI ERA?

AI code of conduct

1. Patient Privacy First

Use AI tools in ways that protect all personal health information—never input identifiable patient data into systems that are not secure or approved.

2. Accuracy Over Convenience

Verify AI-generated information against trusted medical sources and clinical judgment before using it in education, patient care, or assessment.

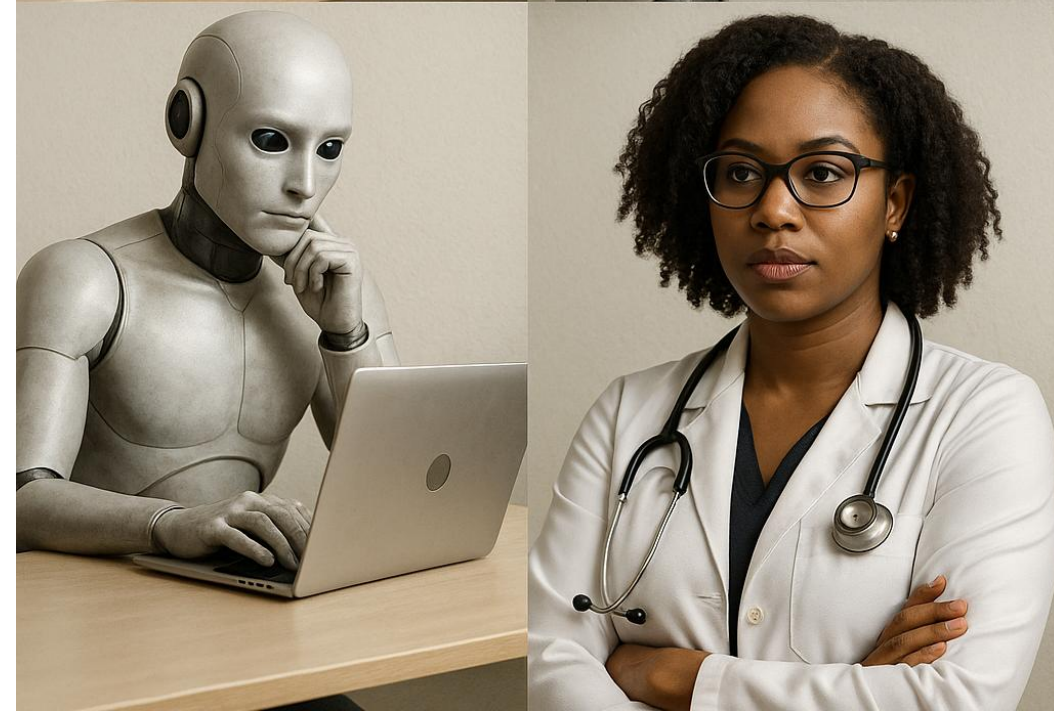
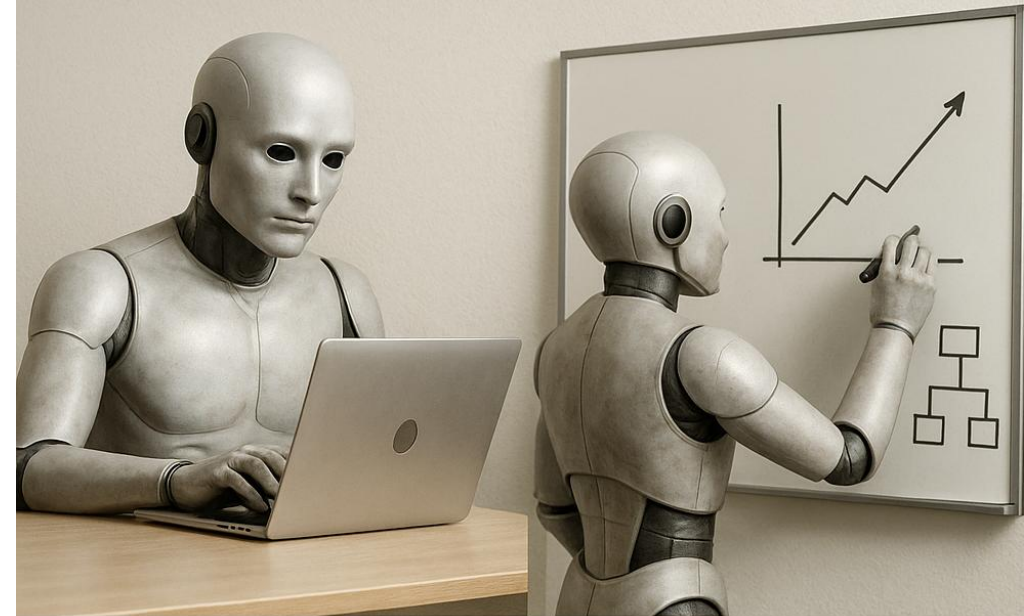
3. Transparency and Attribution

Disclose when AI is used to generate educational content, feedback, or decisions

4. No Disclaimers

AI Generated work is YOUR WORK. You are ultimately responsible for checking facts, reading the work, and

HUMAN IN THE LOOP



Training AI enabled physicians

Modern docs must learn to harness a whole new world of AI resources



MCW's fully secure + HIPAA compliant
Enterprise AI solution



Learn a new concept or disease
Create a podcast or explainer video
Share your didactics slides/readings/videos
Create and share a content expert for reference
Generate flashcards, multiple choice Qs



Create a differential based on a clinical details
Quickly answer a clinical question
Generate multiple



Identify a "best practice"
Research a presentation



Perform a chart audit
Act as a clinical scribe
Write a letter for your patient's employer



Lookup drug dosing information
Scribe an encounter for free



Create a patient handout
Practice a difficult conversation
Create a poster presentation
Research a clinical topic



Create an engaging slide decks



Check email for tone
Reformat presentations

Virtual patients can help doctors learn valuable communications skills **BEFORE** they interact with real patients

Hello, Doctor. Thanks for coming in. Do you... do you have any news for us? We've been waiting for the results from that biopsy yesterday. The nurse said you'd be by this morning to talk about what's going on.

**(obviously bad answer put in for teaching purposes):
You have cancer! You will die from this**

(My hand tightens on my husband's. My face goes pale, and my eyes widen in shock and disbelief. For a moment, I just stare at you, my mouth slightly open as if to speak, but no words come out. My husband, Tom, stands up abruptly, his chair scraping against the floor.)

Husband: "What the hell did you just say to her? Who are you to walk in here and say something like that?"
(I start to shake my head, tears welling up in my eyes. My voice is a choked whisper.)

"What...? No... Cancer? You... you think I'm... you said I'm going to *die*? What are you talking about? I came in for my back... I have a cough... This has to be a mistake. It can't be..."



Improving assessment: The Clinical Reasoning Exam

Moving away from multiple choice tests! AI generated cases. Open ended questions. AI graded assessments w/ physician oversight



George is a 68-year-old man with a six-year history of type 2 diabetes and hypertension who comes to the office for a follow-up visit. He has no complaints. His review of symptoms is negative. He states he "just needs a check up".

George works as an interstate truck driver. His main exercise is "helping unload the truck". He gets most of his food from trucking rest stops. He eats "lots of cheeseburgers, hot dogs, chips, soda, beef sticks, and cheese". He drinks a lot of "coffee and diet coke". He smokes 6 cigarettes a day. He has been checking his blood pressure on the road; reports the last 3 have been "160/80", "156/92", and "148/86". He is currently taking metformin 1000mg BID and hydrochlorothiazide 25mg daily. George is transferring his care to you from out of state. He has no current medical records to review.

Today his vitals are:

BP 149/98 Pulse 88 Respirations 16 Temperature 98.6

What physical exam do you want to check?

What Lab or Radiology tests should we order?

Try it here:



Ambient documentation and medical education

Current medical student may NEVER practice in a world where they will need to write clinical notes
Less time typing should mean more time doctoring, more time to learn, renewed emphasis on communication
Ambient scribes do NOT mean less professional responsibility.

The old way:

Focusing more on the computer than the patient
Manually type a perfect note
Look up clinical topics online or in books



VS

AI enhanced doctoring

Engage with the patient
Verify the AI generated note
Perform an AI assisted literature search



The next digital divide in Wisconsin healthcare

Access to AI infrastructure is now a critical workforce and access issue.

Small and rural systems

- May lack technical expertise
- Lack purchasing power
- Need help with AI governance
- Need protected training time.

Well-resourced systems

- More technical expertise
- Can purchase enterprise AI
- Robust AI governance
- Can train staff for safe AI use



POLICY FRAME

Wisconsin can lead by enabling safe AI innovation!

Enable

Avoid blanket restrictions that freeze medical education around yesterday's tools.

Invest

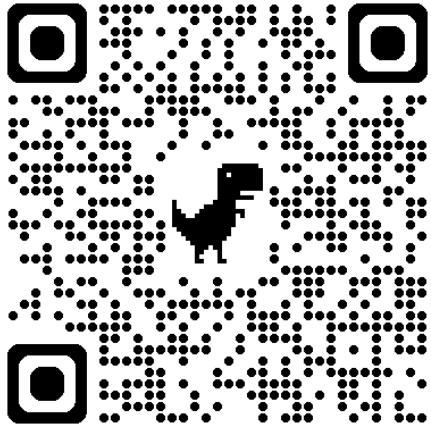
Support secure enterprise access, shared technical expertise, and workforce AI literacy.

Evaluate

Tie pilots to outcomes: learning, safety, access, workload, and patient trust.

Bottom line: AI should help Wisconsin train physicians who think more clearly, communicate more humanely, and extend care where access is limited.

Next steps: Where to learn more!



Check out my Gemini Notebook guide to AI in medical education!

