PROMPT, BUT VERIFY: CRITICAL USE OF AI-GENERATED INFORMATION IN EVIDENCE-BASED MEDICINE

By

Divya Singh

A capstone project submitted for Graduation with University Honors

May 7, 2026

University Honors

University of California, Riverside

APPROVED

Dr. Daniel Novak

Department of Social Medicine, Population, and Public Health

Dr. Begoña Echeverria, Howard H Hays, Jr. Chair

University Honors

ABSTRACT

Generative AI is being used in clinical settings, but medical students are not being taught how to evaluate what these tools produce. Large language models can generate responses that contain fabricated citations, outdated guidelines, and demographically biased recommendations, and none of that is obvious from reading the output. *Prompt, But Verify: Critical Use of AI Generated Information in Evidence-Based Medicine* is an asynchronous online course designed for M3 and M4 medical students at the UCR School of Medicine. The course is built around Challenge-Based Learning Cycles and a continuing clinical vignette. Across five modules, students learn to verify AI-generated claims against primary literature, respond to patients who bring AI-sourced information to appointments, audit algorithmic bias, and build these practices into their clinical workflows. I curated the research for each module and helped reorient the course from a technology-first framing to one grounded in clinical reasoning. I also developed an automated build process for assembling the course in Canvas. The creative work accompanying this submission is the complete course content in EPUB format.

ACKNOWLEDGEMENTS

First, I want to thank Dr. Begoña Echeverria. When things did not go as planned with my original capstone mentor, she stepped in, helped me find a new one, and made sure I still had a shot at finishing. I would not be submitting this without her.

I owe a huge thank you to Dr. Daniel Novak, who agreed to mentor me with less than a year left and never once made me feel like we were behind. Every meeting with him left me more excited about the project than the last, and I look forward to continuing the work we have accomplished with this course.

TABLE OF CONTENTS

Introduction

This course was designed as a response to the growing consequences of casual AI use in healthcare. AI is being used in multiple healthcare contexts, but with no formal structured education for how to use it safely. Medical students are already pulling up AI on their phones while walking into patient rooms. Residents paste clinical details into consumer-level chatbots during overnight shifts. Attendings use AI scribing tools without fully understanding how those tools interpret and present the data they gather.

Prompt, but Verify: Critical Use of AI-Generated Information in Evidence-Based Medicine is designed for M3 and M4 medical students. The goal of the course is to teach students how to think critically about their already present use of AI by teaching them how to verify it, discuss it with patients, and how to document it in a way that ensures transparency and defensibility. Despite growing use of AI in clinical settings, no standardized curriculum exists for teaching medical students how to critically evaluate AI-generated clinical information. There is no shortage of content teaching "how to use ChatGPT" online, but what is missing, and what this course is trying to bridge, is a framework that allows for responsible clinical use that can hold up against the challenges posed by real patient care. The distinction between an AI course that covers medicine and a medical course about AI shaped a lot of my work on this project.

The Problem

Ask a large language model (LLM) a clinical question, and you'll get a confident, well-organized response, often one that includes properly formatted citations with DOIs. Going one step further, however, the DOIs are occasionally routed to an error page, the citations are real but misinterpreted, or they reference guidelines and standards that are now outdated. None of these

issues are visible simply by reading the generated response. It takes intentional verification steps to catch these problems, and right now, most people are not trained to do that systematically, especially clinicians and medical students.

The course begins with a vignette about Maya Patel, a resident, who encounters these very problems during a patient admission on an overnight shift. She does what many residents would: She enters the patient's clinical details into a consumer-level AI tool and, in return, gets a fluid and well-cited differential diagnosis. As she looks further into it, she finds that three of the citations turn out to be fabricated, outdated, or misinterpreted. Maya's case is fictional, but this situation is well-documented and actively occurring in clinical settings. Maya's story is a continuing vignette throughout five modules, presenting and then teaching responses to different challenges that she encounters. Students learn to verify AI claims against evidence and communicate effectively with patients who bring AI-generated research to their appointments. They also learn to identify algorithmic bias in AI outputs and integrate all of this into a sustainable clinical workflow.

Structure and Pedagogy

The course is structured using Challenge-Based Learning Cycles, an active learning methodology developed by researchers at Vanderbilt, Stanford, and the University of Washington. Each of the five modules begins with a complex and open-ended grand challenge question, which is followed by initial reflections, engagement with multiple different perspectives, and a summative assessment reflecting the ongoing case narrative.

Module 1: Foundations of Generative AI in Clinical Contexts

Module 1 begins with foundational concepts, such as how LLMs generate text using next-token prediction and why that mechanism produces fluent fabrication. It also covers what safeguards need to be in place before a clinician uses AI for reasoning, and how to create a provenance log that makes AI-assisted work auditable. The resources include a 3Blue1Brown[1] explainer on transformer architecture, an NEJM review on AI in Medical Imaging[2], a New Yorker investigation[3] on what a doctor's role will be in an AI-augmented world, and an AMA governance framework video[4]. Students must actively engage with each resource by responding to a prompt connecting the material to the challenge question and the case vignette.

Module 2: Applying EBM Principles to AI-Generated Content

Module 2 continues with the introduction of verification. Students learn how to decompose AI outputs into individual claims, which allows them to map those claims to PICO elements, identify the best evidence for each claim, and then verify them against guidelines and primary literature. The module includes a spot-the-hallucination exercise where students work through actual AI-generated clinical claims and must figure out which ones contain fabricated DOIs, outdated recommendations, or overgeneralized findings. The takeaway is to be able to build the kind of pattern recognition that catches these errors before they make it into a clinical decision.

Module 3: Communicating with Patients About AI-Generated Health Information

Module 3 moves into patient communication. Patients are already using AI to look up their conditions, so clinicians need to know how to respond when a patient walks in with semi

accurate AI-generated information. The module covers health literacy, plain language principles, teach-back, shared decision-making, and cultural humility. One of the assessment scenarios has a bilingual patient arrive with a twelve-page AI-generated printout about MS treatment options. The student must take the patient's effort seriously without endorsing the wrong parts.

Module 4: Identifying and Mitigating Bias in AI Outputs

Module 4 uses the Representation, Design, and Deployment bias framework to tackle algorithmic biases. Continuing the story with Maya, she runs a clinical case three separate times using the same AI. Each time, she changes only the patient's demographics. For a Black female patient on Medicaid, there is an unsolicited addition of "medication adherence challenges in this population" and suggests a simplified dosing schedule as a part of the treatment plan. For an uninsured Hispanic patient, the AI recommends a cheaper drug, citing "cost considerations" without being asked. The module prompts students to identify the step at which the bias was introduced, investigate how things like insurance can trigger structural inequalities, and what individual clinicians can do about it. It connects directly to the Obermeyer et al. study[5] in *Science* that found a widely deployed clinical algorithm systematically underestimated illness severity in Black patients because it used healthcare cost as a proxy for health need.

Module 5: Practical Integration of AI into Clinical Workflow and Decision-Making

Finally, Module 5 teaches integration and sustainability. Students build a personal AI-use playbook, create clinical note templates with provenance fields, write patient handouts at appropriate reading levels, and put together a capstone portfolio covering all five modules. Students will leave the course with tools and habits they can use in residency.

Real-World Implications

The healthcare field is moving fast on AI adoption and moving slowly on AI education. Graduating doctors will enter residencies where AI is already embedded in the EHR systems, clinical decision support tools, imaging analysis, and documentation workflows. Without a framework for evaluating these tools critically, they can default to either uncritical trust or blanket rejection, both of which are harmful. Uncritical trust is the main danger of AI in any field, as it leads to automation complacency, reducing the clinician's own scrutiny. Blanket rejection stops them from accessing and learning about powerful new tools, which, if used critically, can improve information retrieval, evidence synthesis, and documentation efficiency.

This course offers students another pathway. It does not treat AI as a savior or a threat. It treats it as a tool that requires the same kind of critical appraisal clinicians already apply to published literature, clinical guidelines, and diagnostic tests. In essence, the "Prompt, But Verify" framework is evidence-based medicine principles applied to a new category of information source that did not exist when the current EBM curriculum was designed. This course is built, it is deployed, and it is designed to be adaptable as the technology and the regulatory landscape continue to evolve. The EPUB version accompanying this submission represents the complete course content as it exists in Canvas, formatted for review outside the learning management system.

Reflection and Conclusion

The content plan and pedagogical architecture for this course were designed by my PI, Dr. Daniel Novak. The module sequence, the learning objectives, the challenge questions, and

the CBLC framework were his design. My work fell into three areas: research and curation, a reorientation of the course's overall framing, and streamlining the build process.

I spent a lot of time finding the papers for each module that were rigorous enough to hold up and accessible so that a medical student would read them. For Module 1, I paired the 3Blue1Brown explainer with clinical writing from the NEJM and the New Yorker, so students could see the same concepts from two different angles. For Module 4, the Obermeyer study was paired with a Nature Medicine paper[6] on how LLMs produce different clinical recommendations based on patient demographics alone, so students could see that the proxy-variable problem is not limited to traditional algorithms.

The reorientation from an AI-in-medicine course to a medical AI course was an important step. That framing starts from the technology and works toward clinical applications, which makes sense for engineers or informaticists, but a third-year student does not need to understand attention heads when handed an AI-generated differential. The readings, prompts, and assessments all needed to start from clinical reasoning and evidence-based medicine, concepts students already have, and build outward from there.

To build this course in Canvas, I began by manually creating each module activity. Very quickly, it became evident that constructing those files by hand was time-consuming and full of errors. I figured out how to utilize Claude's Opus 4.6 to generate IMS Common Cartridge files, which automated the actual course construction. I developed a workflow in which I could describe what I wanted in plain language and have the AI generate valid IMSCC XML files. This allowed for weeks of work to be done in a matter of hours. This framework enabled my PI and

me to focus on the value of the content, the thoughtfulness of the formatting, and making sure the product does not sacrifice quality for efficiency.

I have utilized AI tools throughout my undergraduate career, and that is part of why I wanted to work on this project. I was building a course about critical AI use while relying on it myself to get the course built. That meant I had to verify the outputs and catch the errors. I had to keep a record of what I changed. That is the same process the course asks students to follow. This project did not just produce a course. It made me practice the framework I was helping to build.

Creative Work

Access to *Prompt, But Verify: Critical Use of AI-Generated Information in Evidence-Based Medicine* Course on Canvas: https://canvas.instructure.com

Username: scottyBear

Password: Scotty@26

Access to course in EPUB format via Google Drive: https://tinyurl.com/divya-singh-2026-capstone

Access to Course Overview: http://bit.ly/promptbutverify

REFERENCES

1. 3Blue1Brown. *But what is a GPT? Visual intro to transformers* [video]. YouTube. Published 2024. Accessed May 4, 2026. https://www.youtube.com/watch?v=wjZofJX0v4M
2. Rajpurkar P, Lungren MP. The current and future state of AI interpretation of medical images. *N Engl J Med.* 2023;388(21):1981-1990. doi:10.1056/NEJMra2301725
3. Khullar D. If AI can diagnose patients, what are doctors for? *The New Yorker.* September 22, 2025. Accessed May 4, 2026. https://www.newyorker.com/magazine/2025/09/29/if-ai-can-diagnose-patients-what-are-doctors-for
4. AMA STEPS Forward. *Establishing an AI governance framework* [video]. AMA Ed Hub. Published 2025. Accessed May 4, 2026. https://edhub.ama-assn.org/steps-forward/video-player/18981711
5. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453. doi:10.1126/science.aax2342
6. Omar M, Soffer S, Agbareia R, et al. Sociodemographic biases in medical decision making by large language models. *Nat Med.* 2025;31(6):1873-1881. doi:10.1038/s41591-025-036266