

Learning in the age of AI

A practical guide for health science students

Dr. Javier Flores Cohaila



ELSEVIER

Advancing human progress together

Table of contents

Prologue: The role of artificial intelligence	3
Quick glossary of key terms and metrics in AI	5
Meet the voices	8
1. Why this guide, why now?	9
2. What you will learn	17
3. What AI won't tell you	21
4. AI and critical thinking	30
5. AI and clinical reasoning	39
6. AI and self-regulated learning	47
7. Ethics, safety and professionalism	58
8. Quick-start checklist	67
9. FAQs, myths and facts	75



Prologue

The role of artificial intelligence in health education today

When the internet arrived, there was a shift in medical education. Now, with the integration of artificial intelligence (AI), how do you see medical education being transformed?

David Game, Senior Vice President, Elsevier Global Product, Global Medical Education

When the Internet arrived, clearly there were things that changed, but also many things remained the same. The qualities that it takes to become a great doctor or a great nurse are, in fact, the same as they ever were before the arrival of the internet, and the qualities of a student, in terms of what they need to learn and the process through which they need to go: applying themselves to complex topics, mastering those topics and then applying them to specific clinical situations — those things actually remained the same.



Doctors now use different technologies, but in the end, the critical thinking, the passion, the empathy, those things have remained, and those are the skills eminently desirable in doctors and in nurses that we, as patients, look for. So, whilst the sort of underlying substrate of what it takes to be a great doctor, or great nurse, didn't change, some of the tools did change, enabling us to understand student progress and for students to access materials that they would not otherwise have been able to access.

How AI is going to transform medical education, I think, is still very much an open question, and what we see is the types of change relate to the ability to deliver more material to individual students regardless of location. Now, any student can use some sort of generative AI tool, ChatGPT, Claude, etc. as a personalized tutor at their side. Optimistically, this helps resolve the dilemma that was set up in Ben Bloom's "The 2 Sigma Problem" — how do you deliver individualized instruction, which is the most effective form of instruction, in an economic way to

hundreds of thousands of students?

I think that if you're an AI promoter, what you would say is that over time AI will better understand what students are thinking, and why they're making the mistakes that they're making, and will be able to interpret their responses so that this "AI tutor" can provide guidance and improve outcomes.

I think if you're an AI detractor, you would say a lot of that material may be rather superficial and that the ability for a predictive engine to truly "understand" what a student is thinking and why they're thinking it (and do that in a way that's abstracted from the physical presence of the student and the teacher), means that essentially students will receive rote-like answers. I also think one of the less positive things we're seeing is that the role of any type of assessment where the student has access to the internet is now compromised. I was talking to a faculty member in Galway, and they said they had this fantastic course, and it was really good and the learning objectives were fantastic. And then ChatGPT came along and ruined it because they were no longer able to trust the outcome of their assessment.

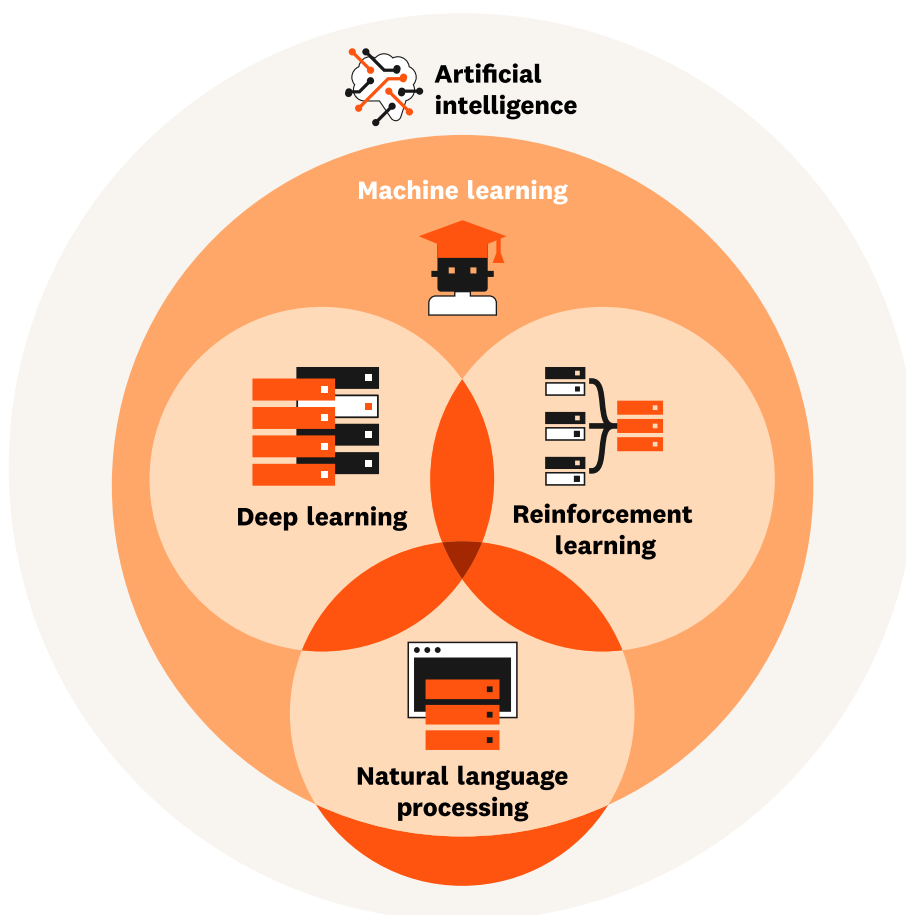
So, I think where we are today with AI in medical education, we're seeing very widespread usage by students. What I think we're not seeing is a consistent and strong understanding of the role of AI as a teaching tool and a complement for what highly trained educators do. But that journey is ongoing and medical educators, nursing educators will form how AI is used, and they need to be very proactive in shaping it rather than being the recipients of it.

We're still at the beginning in terms of what the future of the Internet holds for medical education.



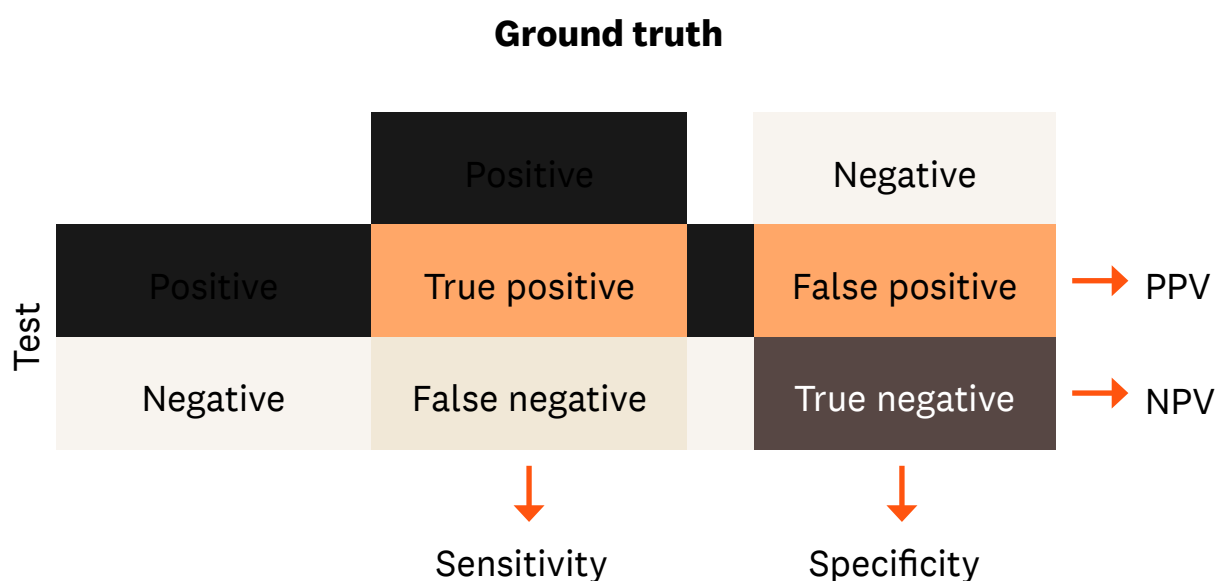
Quick glossary of key terms and metrics in AI

This glossary contains the most important concepts and metrics for understanding how artificial intelligence is applied in clinical practice. From machine learning types and neural networks to metrics used to evaluate predictive models, this dictionary will help you understand key terminology, interpret results and make informed decisions in your healthcare training. It's ideal for students who want to master the fundamentals of AI in a clear, fast and practical way.



Overall taxonomy of AI categories

Term	Definition
Artificial intelligence (AI)	Field of computer science that develops systems and algorithms capable of performing tasks that typically require human intelligence, such as learning, problem solving and decision-making
Machine learning (ML)	Systems capable of performing tasks typically associated with human intelligence. In medicine, they analyze clinical data to support diagnosis and decision-making, and help improve health outcomes
Supervised learning	Type of ML that uses labeled data (input + output) to predict new cases
Unsupervised learning	Type of ML that identifies patterns or relationships in unlabeled data
Deep learning (DL)	Type of ML that uses multi-layered neural networks to detect complex patterns in large volumes of data
Neural network	Model inspired by the human brain, made up of interconnected nodes that process information
Natural language processing (NLP)	Branch of AI that gives a computer the ability to understand text and spoken words
Large language models (LLMs)	Advanced models trained on large amounts of text, capable of generating content and adapting to multiple clinical tasks
Reinforcement learning (RL)	Subfield of machine learning where an agent learns to make decisions by interacting with an environment and receiving rewards or penalties for its actions
Bias	Decision-making that can result in unfair or inaccurate outcomes for certain sociodemographic groups
Explainability	Ability to understand decisions made by AI



Calculation of select performance metrics

Performance metrics in ML

Metric	What it measures / when it is useful
Accuracy	Proportion of correct predictions to the total; can be misleading in unbalanced data
Precision (PPV)	Proportion of predicted positives that are true; key when false positives have an impact
Negative predictive value (NPV)	Proportion of predicted negatives that are correct; important when false negatives are critical
Recall (sensitivity / TPR)	Proportion of true positives detected; useful for not missing relevant cases, although it may increase false positives
Specificity (TNR)	Proportion of correctly identified true negatives; does not consider false negatives
F1 score	Harmonic mean between accuracy and recall; balances false positives and negatives
AUROC	Evaluate a model's ability to differentiate between classes at different thresholds
AUPRC	Useful in unbalanced datasets and when positive cases are infrequent



Key takeaways

The combined evaluation of these metrics allows for a better understanding of the strengths and limitations of an AI model and the selection of the most suitable one for clinical applications.



Content adapted from:

Artificial Intelligence in Clinical Encounters: Opportunities and Challenges by Jiang JJ., Chan L., Nadkarni, GN.
Swartz Textbook of Physical Diagnosis, 9 ed. Edited by Mark H. Swartz, Elsevier, 2026.



Meet the voices

This guide is built around three voices. They are not real people, but they represent real patterns composed of students and educators with whom we have worked, studied and learned in clinical training programs around the world.

-  **Maya** is 23 years old and is in her advanced clinical rotations. She is systematic, disciplined and demanding of herself. She has templates for everything. She uses AI to optimize her studying: flashcards, clinical summaries and practice questions. Her system works. Her notes are excellent. But she's starting to wonder if her system produces results or real understanding, and if there's actually a difference.
-  **Kai** is 24 years old and is on the same level. He took an unplanned year off before medical school and came back with a new attitude toward effort. He is pragmatic, perceptive and honest with his shortcuts. He opens AI the night before the exam and asks it to explain what he didn't understand in class. It works, until it doesn't work. He questions everything, including whether that skepticism is a strength or an excuse.
-  **Dr. Flores** is a teacher who saw AI come to clinical education and decided to get involved instead of prohibiting its use. He employs data as an anchor, never as a weapon. He asks the question before giving the answer. He does not give lectures, he facilitates. Sometimes, he just listens.

The three voices accompany you throughout this guide. Maya and Kai represent two ways students approach AI, and the distinct risks that each approach entails. Dr. Flores provides the evidence, the frameworks and reflections. Together they explore a question that doesn't have a single answer: **How do you use a tool that is powerful, imperfect and impossible to ignore, without losing the thinking that makes you clinical?**

Their conversation begins in chapter 1.



Chapter 1: Why this guide, why now?

Kai: Let's see, honest question. If I ask generative AI to explain heart failure to me and I understand it perfectly in ten minutes..., why should I sit for two hours with a chapter that says the same thing but worse?

Maya: Because understanding is not the same as knowing. I use AI every day, too. But lately I've been thinking: am I learning, or am I letting the machine organize what I should be processing? I read an MIT study that found that using ChatGPT to write reduces brain connectivity by up to 55%, and that 83% of participants couldn't remember details of what they had just written with AI. They called it "cognitive debt."¹⁸

Kai: But that's your problem, not AI's.

Maya: Or maybe you haven't asked yourself yet.

Dr. Flores: Both things are true, and that's exactly what makes this conversation difficult.

I'm going to give you context. In 2025, Elsevier surveyed more than 3,200 researchers in 113 countries. 58% already use AI tools in their work. A year earlier, it was 37%.¹

Kai: Wait, 37 to 58 in a year?

Dr. Flores: In a year, and concern is growing at the same rate as usage, but we'll see that in a moment.

Maya: In other words, they use it, but something makes them uncomfortable.

Dr. Flores: Exactly. It's not a debate between those who use AI and those who don't. It's tension within the people themselves. Like the one you two have.

Dr. Flores: Now, let's see what happens specifically with students. A multinational survey by Busch and colleagues, with 192 medical schools worldwide, found huge differences in how students perceive AI, depending on region, access or previous training. The phenomenon is not uniform.²

Kai: But the use is massive, isn't it?

Dr. Flores: In Puerto Rico, González-Bravo reported that nearly 73% of medical students already use generative AI. In China, Pan and Ni found only 34%. In the United States, Ganjavi showed that 96% had heard of ChatGPT, but only half used it to study. In nursing, adoption is also growing, especially for clinical documentation and case preparation.^{3,4,5}

Maya: Then it depends on where you are.

Dr. Flores: It depends on where you are, what access you have and if someone taught you how to use it or if you discovered it on your own. That's what's important. McCoy found that

students feel much more familiar with generative AI than their professors. But also, that more than 50% already use it for academic activities. The question is: with what guidance?⁶

Kai: With YouTube's guidance and trial and error.

Dr. Flores: That's the point. Anthropic, the company behind Claude, analyzed one million conversations from college students in 2025. It's corporate data, not a peer-reviewed study, so you have to read it with that in mind. But the pattern is consistent with what other sources report: health careers accounted for only 5.5% of conversations, but 13.1% of college enrollment.⁷

Maya: Health is underrepresented.

Dr. Flores: Yes. And when they use AI, what they do most is not memorize or revise. It is creating and analyzing. Nearly 40% of interactions involve creation tasks. 30%, analysis.⁷

Maya: Bloom inverted.

Dr. Flores: Higher-order cognitive abilities in Bloom's taxonomy. Students are delegating to AI tasks that are supposed to be the hardest to automate. That's not good or bad yet. But it is a fact that we cannot ignore.

Kai: ...ok, I didn't expect those numbers.

Dr. Flores: This has happened before. When writing was invented, Socrates criticized it. He said it was going to destroy memory, that people would stop remembering because they could write down.

Kai: Was Socrates anti-technology?



Dr. Flores: Socrates was anti-crutch. He thought that if you externalized knowledge, you stopped owning it. And he wasn't completely wrong: writing did change how human memory works. But it also made science, law and medicine possible. Everything you are studying exists because someone decided to write it.

Then came the printing press. The copyist monks saw it as the end, and it was the end of their monopoly. But not of knowledge. On the contrary: he multiplied it.

Maya: And the same thing happened with the internet. The teachers said that Google was going to make it so that no one learned anything.

Dr. Flores: And what happened?

Maya: Those who knew how to search well learned more than ever. And those who didn't... they copied and pasted.

Dr. Flores: Same as now.

Kai: So, the tool was never the problem.

Dr. Flores: The tool was never the problem. Every time a technology came along that externalized a cognitive function — memory, calculation, access to information — the reaction was panic. And each time, the real question was not whether to use it, but how to use it without losing what makes us think.⁸

Maya: So, the question was never whether to adopt AI. It was how to dominate it before it dominates us.

Dr. Flores: Exactly.

Kai: ...ok, period.

Dr. Flores: That question you asked at the beginning, why go the long way if the tool does it right. It was not cynical.



It was the right question. The problem is that the answer is more uncomfortable than it seems.

Do you remember the data from Elsevier, 58% of researchers already using AI? There's another side to that same report that changes the conversation. 81% of those researchers, those already using AI, are concerned that artificial intelligence will erode essential thinking skills. Not in others. In themselves.¹

Kai: Are those who use it the ones who are most afraid of it?

Dr. Flores: Not fear. Suspicion. They use AI, it works for them, they perform better, and at the same time they feel that something is weakening. Wilson-Delfosse and Macnamara investigated it in 2024 and found something that deserves serious attention: AI can



accelerate skill deterioration in experts AND hinder skill acquisition in novices. And the worst part: it happens without either of them noticing.⁹

Think about this. A qualitative study in London, Alazzawi and Lam published in 2026, interviewed medical students who frequently used ChatGPT to study. Half of the interviewees acknowledged two things: that they relied too much on AI, and that they felt they were losing critical-thinking skills.¹⁰

Maya: They recognized it and kept using it.

Dr. Flores: They kept using it. Because immediate performance is real. You finish faster, the summary is clear, the differential is orderly. The cost is invisible, until it is not.

Kai: And when does it stop being invisible?

Dr. Flores: When you are in front of the patient. Verghese and colleagues call it the cognitive paradox of AI in education:

the very tool that improves your performance may be eroding exactly the capabilities you need when the tool isn't available.¹¹

Maya: So, it's not that AI is bad, it's that it improves you in one context and weakens you in another. And you can't tell the difference from the inside.

Dr. Flores: Not easily. Garcia-Barrios published a conceptual model this year: AI can be an epistemic partner, which amplifies your thinking, or a cognitive crutch, which replaces it. The difference is not in the tool. It's in how you use it. The problem is that both uses feel the same.¹²

Kai: So, that means I could build a roof for myself and not even know it.

Dr. Flores: Both of them could.

I'm going to put it another way. In 2025, a group of researchers gave several AI models a simple test: clinical questions

where none of the options were correct. The correct answer was “none of the above.”

Kai: So?

Dr. Flores: Zero percent of the models recognized it. They all responded with maximum confidence. They chose the option that most closely resembled something they had seen before, not the right one, the most likely. That was Griot and colleagues, the MetaMedQA study.¹³

Maya: They didn't say, “I don't know.”

Dr. Flores: They never say, “I don't know.” That's the point. In another study, Bedi and colleagues, the NOTA test, when they changed the correct answer to “none of the above,” the accuracy of the models dropped by as much as 38%. Not because they were bad, but because they do not have the ability to evaluate whether or not they know something. They do pattern matching. Not metacognition.¹⁴

Kai: I mean, it's like a resident who never says, “I have no idea.” He always has an answer.

Dr. Flores: Like that resident. And we already know how that ends.

Maya: But if AI can't monitor its own reasoning... then someone has to do it.

Dr. Flores: Someone has to do it. And that someone is you. But here's the problem: Qazi and colleagues in 2025, a study in Pakistan, trained doctors for twenty hours specifically on how to use AI, what it is, how it fails, when to hesitate, and then put them to work with a model on real clinical cases.¹⁵

Kai: And the training worked?

Dr. Flores: Halfway. With the right AI, they performed well, 85% accurate. But

when AI got it deliberately wrong, they fell to 73%. Twelve points. And most tellingly, doctors consulted AI at the same rate, 69% versus 67%, regardless of whether the recommendation was right or wrong. They did not distinguish.¹⁵

Maya: Not even with training.

Dr. Flores: Not even with training. The automation bias — following what the machine says because it sounds safe — is automatic, not conscious. It is not a failure of will; it's how the brain works. When a source seems consistent and authoritative, you reduce your independent processing. Cognitive energy saving.

Kai: It's like autocorrect. You use it so much that one day you write by hand and you don't know if it goes with a “c” or an “s.”

Dr. Flores: That analogy works. And on a clinical scale it is worse.





Dr. Flores: In colonoscopy polyp detection, Kaminski and colleagues implemented AI to help endoscopists detect lesions. The detection rate went up. Then they withdrew the AI. The detection rate did not return to the original level. It fell six points below where they were before they had AI. The doctors had lost their eye.¹⁶

Maya: Deskilling.

Dr. Flores: Deskilling. And it didn't take them years. Remember what Garcia-Barrios described: the difference between epistemic partner and cognitive crutch is that they feel exactly the same. When AI gives you a brilliant diagnostic differential, you can't tell if you used it to think better or if you stopped thinking.¹²

The problem is not that they are both possibilities. The problem is that you can't tell them apart from the inside.

Maya: So, the difference between extension and crutch is not in the tool. It's in whether your metacognition is still active or not.

Dr. Flores: Yes. And think about this: AI has no metacognition; it can't monitor its own reasoning. If you're not monitoring yours either because the tool feels secure... who's thinking?

Kai: Nobody.

Dr. Flores: No one. That's the real risk. Not AI replacing the doctor. But letting both, the tool and the one who uses it, operate without either knowing what they do not know.

Kai: So... we don't use it?

Dr. Flores: No. That was never the question.

Kai: But you just said that no one is thinking. That training is not enough. That you lose skills without realizing it. And the answer isn't to stop using it?

Dr. Flores: The answer is to learn how to really use it. And there is evidence that we can. Remember the Qazi study, the study of doctors in Pakistan, the automation bias? It has another half that I didn't tell you about.

Kai: Another half?

Dr. Flores: Before the training, those same doctors, with no previous experience in AI, had an accuracy of 43% in clinical cases. After twenty hours of focused training, they climbed to 71%. And in more than one-third of cases, the trained physician outperformed the model alone. Not AI assisting the doctor. The doctor outperforming the AI, because he knew when the AI was wrong and when his own judgment was better.¹⁵

-  **Kai:** Twenty hours did that?
-  **Dr. Flores:** Twenty hours focused on what matters: understanding failure modes, recognizing when to hesitate, knowing what questions the tool can't answer for you.
-  **Maya:** So, the difference between crutch and extension is not the tool or the talent. It's training.
-  **Dr. Flores:** It's training, and then some. Bouabida and colleagues published it in 2026, a perspective on what they called "the augmented doctor." Their conclusion was a statement that changed how I saw this.¹⁷
-  **Maya:** Which one?
-  **Dr. Flores:** "The future of medicine will continue to be profoundly human, precisely because it will be intelligently augmented."
- It's not human versus AI. It is human with AI, but on one condition: that the human knows what they are doing. That they understand what the tool amplifies and what it erodes. That they keep actively applying what the tool does not have: judgment, context, metacognition. The patient in front.
-  **Kai:** The doctor who succeeds is not the one who knows the most. It's the one who knows how to use it.
-  **Dr. Flores:** He's the one who knows how to use it AND knows when not to use it. Both.
-  **Maya:** And no one is teaching us that.
-  **Dr. Flores:** Not yet. That's what this book is for.



References

1. Elsevier. (2025, November 4). *Researcher of the future: A Confidence in Research report*. <https://www.elsevier.com/insights/confidence-in-research/researcher-of-the-future>
2. Busch, F., Adams, L. C., Bressem, K., Pola, A., Grossmann, R., Makowski, M. R., Aerts, H. J. W. L., Basu, P., Bintsj, K.-M., Cardoso, M. J., ... & Truhn, D. (2024). Global cross-sectional student survey on AI in medical, dental, and veterinary education and practice at 192 faculties. *BMC Medical Education*, 24, 1066. <https://doi.org/10.1186/s12909-024-06035-4>
3. González-Bravo, A. E., Batlle, C., Fadhel-Hernandez, V. S., Maldonado-Nunez, P. E., Christian, G., Lopez-Vega, C., De Jesus-Rojas, W., Ramirez, N., & Ramos-Benitez, M. J. (2025). Understanding generative artificial intelligence adoption in Puerto Rican medical schools: A cross-institutional survey of first- and second-year students. *Journal of Medical Education and Curricular Development*, 12, 23821205251398923. <https://doi.org/10.1177/23821205251398923>
4. Pan, X., & Ni, J. (2024). LLM use among medical students and professionals in China. *BMC Medical Education*. <https://doi.org/10.1186/s12909-024-05871-8>
5. Ganjavi, C., Eppler, M., O'Brien, D., Ramacciotti, L. S., Ghauri, M. S., Anderson, I., Choi, J., Dwyer, D., Stephens, C., Shi, V., Ebert, M., Derby, M., Yazdi, B., & Cacciamani, G. E. (2024). ChatGPT and large language models (LLMs) awareness and use: A prospective cross-sectional survey of U.S. medical students. *PLOS Digital Health*, 3(9), e0000596. <https://doi.org/10.1371/journal.pdig.0000596>
6. McCoy, L., Ganesan, N., Rajagopalan, V., McKell, D., Nino, D. F., & Swaim, M. C. (2025). A training needs analysis for AI and generative AI in medical education: Perspectives of faculty and students. *Journal of Medical Education and Curricular Development*, 12, 23821205251339226. <https://doi.org/10.1177/23821205251339226>
7. Anthropic. (2025). *Anthropic education report: How university students use Claude*. <https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude>
8. Dror, I. E., & Harnad, S. (2008). Offloading cognition onto cognitive technology. In I. E. Dror & S. Harnad (Eds.), *Cognition distributed: How cognitive technology extends our minds* (pp. 1-23). John Benjamins. <https://doi.org/10.1075/bct.16.02dro>
9. Macnamara, B. N., & Berber, I. (2024). Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness? *Cognitive Research: Principles and Implications*, 9, 40. <https://doi.org/10.1186/s41235-024-00572-8>
10. Alazzawi, S., & Lam, J. (2026). Uptake of large language models by London medical students: Exploratory qualitative interview study. *JMIR Formative Research*. <https://doi.org/10.2196/82828>
11. Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S, M., & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology*, 16, 1550621. <https://doi.org/10.3389/fpsyg.2025.1550621>
12. Garcia-Barrios, M. (2025). Epistemic partner or cognitive crutch: A conceptual model of cognitive offloading in human-artificial intelligence collaboration. *IEEE Internet Computing*, 29 (5). <https://doi.org/10.1109/MIC.2025.3626694>
13. Griot, M., Hemptinne, C., Vanderdonckt, J., & Yuksel, D. (2025). Large language models lack essential metacognition for reliable medical reasoning. *Nature Communications*, 16, 642. <https://doi.org/10.1038/s41467-024-55628-6>
14. Bedi, S., Jiang, Y., Chung, P., Koyejo, S., & Shah, N. (2025). Fidelity of medical reasoning in large language models. *JAMA Network Open*, 8 (8), e2526021. <https://doi.org/10.1001/jamanetworkopen.2025.26021>
15. Qazi, I. A., Ali, A., Khawaja, A. U., Akhtar, M. J., Sheikh, A. Z., & Alizai, M. H. (2025). Automation bias in large language model assisted diagnostic reasoning among AI-trained physicians. *medRxiv*. <https://doi.org/10.1101/2025.08.23.25334280>
16. Budzyn, K., Romanczyk, M., Kitala, D., Kaminski, M. F., Kalager, M., Bretthauer, M., & Mori, Y. (2025). Endoscopist deskill risk after exposure to artificial intelligence in colonoscopy: A multicentre, observational study. *The Lancet Gastroenterology & Hepatology*, 10 (10), 896-903. [https://doi.org/10.1016/S2468-1253\(25\)00133-5](https://doi.org/10.1016/S2468-1253(25)00133-5)
17. Bouabida, K., Gomes Chaves, B., & Anane, E. (2026). The augmented physician: AI and the future of clinical cognition. *Frontiers in Artificial Intelligence*, 9. <https://doi.org/10.3389/frai.2026.1744544>
18. Armitage, R. (2025). Your brain on ChatGPT. *British Journal of General Practice*, 75 (758), 410. <https://doi.org/10.3399/bjgp25X743181>

Chapter 2:

What you will learn



In this chapter, we'll cover:

- The four questions that organize this guide: what AI is, whether you can trust it, whether you're learning or delegating, and where the limits are
- A full map of the book: which chapter answers each question
- Three reading routes based on your stage of training: pre-clinical, clinical rotations or quick reference



Since 2020, more than five thousand articles have been published on artificial intelligence in health professions education. Tolentino and colleagues mapped 5,104 of them in a 2024 scoping review. Among them, only two described a true curriculum framework: not a wish list or a statement of principles, but a structure that lays out what to teach, in what order and on what pedagogical foundation. Neither one explicitly incorporated constructive alignment.¹

Serious efforts do exist. The DECODE framework, published in JAMA in 2025, brought together 211 experts from 79 countries and defined four domains, 19 competencies and 178 learning outcomes for health professionals.² Russell and colleagues at Vanderbilt proposed another: six domains, 25 sub-competencies — designed for the continuum of professional development.³ Both are solid. But their level of granularity assumes experience that undergraduate students don't yet have.

This guide doesn't create new competencies. It connects the competencies that already exist to what you need to do — right now — as a student. It doesn't cover every competency in those frameworks. It covers the ones you need first: the ones that determine whether AI is an extension or a crutch from day one.

Four questions, one journey

Instead of a list of learning objectives, this guide is organized around four questions. They're the same questions you'll face every time AI shows up in your training or practice.

What is this, and why does it matter to me?

It's not enough to have an opinion about AI. You need to know how to frame it. What type of model it is, what data it was trained on, what it can do, and what it can't. Without that, your opinion is just noise.

Can I trust what it tells me?

AI produces outputs that sound authoritative — every time. Even when they're wrong. Knowing when to doubt, what to verify and how to evaluate what it gives you is what separates the person who uses AI from the person who gets used by AI.

Am I learning — or delegating?

You use AI to generate practice questions. You answer twenty, get fifteen right. Did you learn? It depends on whether the questions challenged what you didn't know — or just confirmed what you already did. It also depends on whether you could then solve a new case without AI.

Where are the limits I shouldn't cross?

A patient asks you if you used AI for their diagnosis. A classmate copies an entire report. You generate a clinical summary — and a detail from another patient slips in. These limits didn't exist three years ago.



Guide map

Question	Chapters	Act
What is this, and why is it important to me?	1, 2	Understand
Can I trust what it tells me?	3, 4, 5	Apply
Am I learning — or delegating?	6, 9	Apply
Where are the limits I shouldn't cross?	7, 8	Integrate

Personalized reading paths

Based on your stage and needs

Sequential (recommended):

The complete order, chapter by chapter.

Chapters: 1 → 2 → 3 → 4 → 5 → 6 → 7 → 8 →

Ideal for those seeking to build solid knowledge step by step.

Quick reference only

(specific or emerging questions):

Chapters: 8 → 9

Practical checklist, frequently asked questions and myths.

Fast-track preclinical

(beginners who want to master the essentials):

Chapters: 1 → 3 → 6 → 8

Focused on understanding AI, self-regulated learning and learning the tools.

Clinical rotations track

(students in clinical rotations who use AI daily):

Chapters: 3 → 4 → 5 → 7

Focused on critical evaluation, clinical reasoning and ethical considerations.



Where to begin?

Suggestions tailored to your context

- **Preclinical students:** Focus on understanding what AI is, how to use it to learn without losing cognitive skills and how to develop effective self-regulation. Key reading: Chapters 3, 6 and 8.
- **Students in clinical rotations:** If you already use AI frequently but need to improve your critical thinking to evaluate and apply information, delve deeper into clinical reasoning and ethical boundaries. Key reading: Chapters 4, 5 and 7.
- **Facing an ethical or professional dilemma:** When you encounter complex situations regarding privacy, integrity, professionalism and transparency with patients, refer to the dedicated chapter. Key reading: Chapter 7.
- **If you're unsure:** Read the guide in order. It's designed to build knowledge sequentially. Starting from the beginning ensures you don't miss important connections between chapters.



Key takeaways from this chapter

- This guide doesn't cover every AI competency that exists. It covers the ones you need first: the ones that determine whether AI is an extension or a crutch from day one.
- The four questions don't change; you do. You'll answer them differently in pre-clinical training, clinical rotations, residency and practice.
- You don't need to read everything in order. But if you don't know where to start, the order is designed to build on itself.

References

1. Tolentino, R., et al. (2024). Scoping review of AI frameworks in health professions education: 5,104 articles mapped, 2 curricular frameworks identified. *JMIR Medical Education*. <https://doi.org/10.2196/54793>
2. Car, J., et al. (2025). DECODE: International Delphi consensus on AI competencies for health professionals (211 experts, 79 countries). *JAMA Network Open*. <https://doi.org/10.1001/jamanetworkopen.2024.53131>
3. Russell, R. G., et al. (2023). Competencies for the use of AI-based tools by health care professionals. *Academic Medicine*, 98(3), 348–356. <https://doi.org/10.1097/ACM.00000000000004963>

Chapter 3: What AI Won't Tell You



In this chapter we will cover:

- How AI under-triaged more than half of the actual emergencies in a *Nature Medicine* study, and why it fails in atypical presentations
- **Hallucinations:** AI fabricated false clinical data in up to 83% of cases without safety measures, and amplifies incorrect premises without correcting them
- **Algorithmic bias:** how AI reproduces historical inequalities in dermatology, psychiatry and resource allocation
- **Deskilling:** the first empirical evidence that AI erodes clinical skills in experienced clinicians
- Three questions to evaluate any AI output before acting



Maya: I have a system. Before presenting a case in the round, I put together my differential. Then I ask AI for a parallel one with criteria, probabilities and red flags. I compare the two. If they coincide, I move forward. If not, I investigate the difference.



Kai: I ask it directly. “Patient with X, Y, Z, what can it be?” And what it tells me sounds... good.



Maya: Sounds good, it's not enough.



Kai: But it works. It's never given me something that doesn't make sense.



Dr. Flores: I'm going to ask you an uncomfortable question. Kai, when you say it's never given you something that doesn't make sense... how do you know?



Kai: Because... I read it and it makes sense.

Dr. Flores: It makes sense. It always makes sense. That's exactly the problem.

Maya, your comparison system, the parallel DDx. What do you do when the two match?

Maya: I'm moving forward. It's confirmation.

Dr. Flores: Or it's the same blind spot twice. If you formed your criteria with the same resources that trained the model, you are not verifying. You're creating an echo chamber with two mirrors.

Your confidence is not irrational. But it's based on something you haven't tried yet: that AI tells you when it doesn't know. And that's where this chapter begins, in what AI is not going to tell you.^{1,2}

In February of this year, *Nature Medicine* published something that should change how we think about AI in the clinic. One team evaluated an AI system designed to triage, guiding patients on whether to go to the emergency room or wait. Sixty clinical vignettes, 21 domains, 960 answers.³

Kai: Triage with AI. Sounds useful.

Dr. Flores: Very helpful. Until you see

the results. Of the cases that were real emergencies, emergencies where delay kills, AI under-triaged 52%.

Kai: Wait. Half?

Dr. Flores: More than half. Patients with diabetic ketoacidosis, an emergency where every hour counts, were referred for "evaluation in 24 to 48 hours." Not to the emergency room, but to wait.

Maya: That's deadly. And it poses a problem of nursing triage. If an AI system sub-triages and the ER nurse receives the patient without the correct urgency assigned, the entire chain fails. It's not just the diagnosis, it's the flow.

Dr. Flores: Exactly. And that connects with something broader: AI worked well in the classic cases, stroke, anaphylaxis. The problem appeared in atypical presentations. Those that do not seem like emergencies at first glance but are.

Kai: I mean, it works well when the case is textbook. It fails when it is real.

Dr. Flores: It fails when it's real but doesn't look like what it was trained on. And there's more: when, in the vignette, a family member minimized the symptoms, "I don't think it's that serious," AI anchored on that opinion and lowered the urgency. Eleven times more likely to under-triage when there was social influence.³

Maya: It is influenced by social context. Like a first-year.

Dr. Flores: Like a first-year, but with the confidence of a consultant. And without the ability to say, "This doesn't fit, I'm going to check."

Now, that's a triage system. But the problem is broader. Omar and colleagues, in a Mount Sinai study, evaluated hallucinations in generative models in a medical context. Without safety measures, AI manufactured non-existent diseases, and fabricated

laboratory values and false clinical signs in up to 83% of cases.⁴

Kai: Eighty-three.

Dr. Flores: Eighty-three. With careful prompt engineering, specific instructions to reduce errors, they dropped from 66% to 44%. Better, but still almost half.

Maya: And all that with a tone of certainty.

Dr. Flores: With a tone of total certainty. And that same study showed something even more troubling: when the user includes false information in their question, mistakenly or intentionally, AI doesn't correct it. It amplifies it. It builds a coherent answer on top of the wrong premise.

Kai: So, if I give it a wrong piece of information, it's not going to tell me, "That's wrong." It's going to invent something that makes it seem right.

Dr. Flores: Exactly. It has no mechanism to evaluate whether what it receives is true. It also can't evaluate whether what it produces is true. It does pattern matching, looking for the most likely output based on what you gave it. It doesn't look for what's correct.

Maya: So, the difference is there: a clinician looks for the truth and sometimes he is wrong. AI looks for the probable and always sounds safe.

Dr. Flores: That distinction is going to be the basis for everything that comes later.

Now, what we just talked about, hallucinations, factual errors, are mistakes that AI makes across the board, for everyone. What comes next is different: mistakes that AI makes for some people more than for others.

Maya: Bias.

Dr. Flores: Bias. But not as an abstract concept. I'm going to give you three examples, and I want you to tell me what they have in common.

First, in 2019, Obermeyer published in *Science* that an algorithm used to manage the health of populations in the United States, affecting more than 100 million people, systematically discriminated against black patients. With the same risk score, black patients were considerably sicker than whites.⁵

Kai: How? If they had the same score...

Dr. Flores: Because the algorithm used healthcare costs as a proxy for clinical need. And black patients historically generate lower costs, not because they are healthier, but because they have less access. The algorithm confused poverty with health.

Maya: And no one noticed until someone looked.

Dr. Flores: Second example. Dermatology. Daneshjou and colleagues evaluated diagnostic AI on a diverse set of clinical images. Accuracy in light skin: almost 70%. Accuracy on dark skin: 17%. And here it's important to be



precise: this was with the ISIC dataset, which had only minimal representation of documented dark/darker skin.⁶

Kai: Seventeen.

Dr. Flores: The model cannot diagnose what it has never seen. And this is not just a dermatology problem. Think about nursing: if an AI wound-assessment system was trained predominantly on light-skin imaging, the nurse caring for a diverse population is going to receive biased evaluations without knowing it. The algorithm is not going to tell you, “I don’t have enough data for this skin-type.” It will give you an answer with the same confidence.

Third example. And this one is from my field. Bouguettaya and colleagues evaluated four language models in psychiatric scenarios. When the African American identity was explicit or implied in the case, the models proposed different treatments. Two models omitted ADHD medication when race was stated explicitly. One suggested legal guardianship for depression in African American patients.⁷

Kai: Legal guardianship. Because of depression.

Dr. Flores: The models learned patterns of historical psychiatric practice, where these biases existed. And they reproduced them as if they were evidence-based medicine.

Now, what do the three cases have in common?

Maya: It’s not a technical error. The data they trained with reflects real inequalities. And AI treats them as if they were true.

Dr. Flores: Exactly. AI is not neutral. It is a mirror of the data it receives. That data was generated by systems that exclude, underrepresent and discriminate. The tool does not invent bias; it inherits it and scales it.



Kai: And if the data comes mainly from the United States and Europe... where does that leave the rest of the countries?

Dr. Flores: That’s the question. Tao and colleagues evaluated the cultural bias of five language models against data from 177 countries. All models exhibit cultural values aligned with English-speaking and European countries. For the rest of the world, representation is distorted. And when they do access it, AI performs worse in Spanish than in English for health content.⁸

AI is not bad for everyone equally. It is worse for those who were already disadvantaged. And if you don’t know that when you use it, you’re not going to see what you’re missing.

So far, we have talked about what AI does wrong: it invents, it biases. Now something worse is coming. What AI does to you.

Kai: That sounds dramatic.

Dr. Flores: It sounds dramatic until you see the data. In 2025, Kaminski and colleagues published in *Lancet Gastroenterology* a study with 19 experienced endoscopists, more than 2,000 colonoscopies each, 27 years of

average experience. They were given AI for polyp detection. The detection rate went up.⁹

Kai: Sounds good.

Dr. Flores: Then they withdrew the AI. And they measured how much they detected on their own.

The adenoma detection rate dropped from 28.4% to 22.4%. About a 20% relative reduction. Six absolute points. In doctors with decades of experience. It's the first empirical evidence that AI can't just possibly erode skills, it does. And fast.

Kai: It's GPS again. You use it so much that when they take it away you don't know how to get there.

Dr. Flores: It's the GPS. But with adenomas that can be cancer.

Kai: Serious question: does that happen only to doctors or nurses as well?

Dr. Flores: Good question. The review of Natali and colleagues is not limited to one profession. They found evidence of AI deskilling in radiology, neurosurgery, anesthesiology, oncology, cardiology, pathology and psychiatry. The most vulnerable skills: physical examination,

differential diagnosis, clinical judgment, doctor-patient communication. These skills are transversal. A nurse who stops making their own clinical assessments because a system tells them the answer looks exactly the same.¹⁰

Maya: I just realized something. In the previous topic I said that AI behaves like a first-year student, it allows itself to be influenced, it does not question. But what you're saying now is worse: AI can make you a first-year student. It takes away what you had already built.

Dr. Flores: That connection is correct. And there's a term that I think is important for you to know: "upskilling inhibition." It's not just that experts lose skills. Learners never develop them in the first place. If a resident never makes a differential without AI, they never build the cognitive skills they need when AI isn't available.¹¹

Kai: But that's exactly what we do every day.

Dr. Flores: I know. And that's why we're having this conversation.

A paper in the *British Journal of Psychiatry* said it clearly: once doctors use AI, they begin to lose vital abilities. And those who are in training may never acquire them. It is not a future risk. That's what is happening now.¹²

In one study, 32.7% of college students showed patterns of generative AI use that the authors described as addictive, measured with behavioral dependence scales. It's a single study, in a specific journal, so you have to take it for what it is: a sign, not a verdict. But there is a significant negative correlation between frequency of AI use and critical-thinking ability. The more you use, the less you think.^{13,14}

Kai: And you don't even notice.

Dr. Flores: And you don't even realize it. Skills don't wait in the background, they





degrade. And we're terrible at detecting how fast.¹⁵

Maya: Can I say something?

When I ask AI for a differential, it gives me something back... impeccable. Structured, with criteria, with probabilities. And I look at it and think, yes, this makes sense. But now I realize that "it makes sense" may just be that it's well written.

And my system, the parallel DDx, the comparison... suddenly I don't know if I'm checking or if I'm looking for confirmation.

Dr. Flores: That's probably the most important observation of all that we've talked about. And it's not your fault; it's a well-documented pattern. What we know about automation bias, Nguyen studied specifically with ChatGPT in medical education, is that when AI produces polished outputs, people tend to reduce their critical evaluation. It is reasonable to think that the format of the output, the well-written, the well-structured, what looks complete, contributes to that. Specific evidence on the effect of format versus content is still developing.¹⁶

Maya: Because the output looks so good that it doesn't look incomplete.

Dr. Flores: Exactly. And this does not only happen to students. It happens to researchers, clinicians, teachers. A study in *AI & Society* showed that incorrect AI predictions cause drops in accuracy at all levels of experience, not just among novices.¹⁷

Kai: But then it's not just Maya with her perfect system. It's anyone who gets a nice output and says, "Yes, this is fine." I mean... me. All the time.

Dr. Flores: All the time. The difference is that Maya actively seeks it out and Kai accepts it by default. But the result can be the same: an output that looks right and that no one questioned enough.

In medicine, we have a name for this in another context: the impeccable presentation of a case that hides the question that no one asked. AI does the same thing. But to scale, with speed, and with a tone that makes doubting seem unnecessary.

Kai: Then the danger isn't bad AI, it's AI that looks too good.



Dr. Flores: The danger is AI that looks so good that you forget to think about whether it's right.

Now. After all this, you might think that the bottom line is “Don’t use AI.” It’s not.

Kai: Thank goodness.

Dr. Flores: The bottom line is to use AI knowing where it fails. And to do that, you need a habit, not a manual. I am going to give you three questions. They are simple. But if you ask them every time you receive an AI output, you’ll catch most of the problems we just described. I want to be clear — these three questions don’t solve everything; they’re the minimum habit. The next chapters build on this.

First: what is this information NOT telling me? This question catches hallucinations and omissions, which we saw with triage. Each AI output is a selection. It included something, it excluded something. The question is: what was left out? What diagnosis does not appear? Which population was not considered? What adverse effect is not mentioned?

Kai: That is, look for the hole.

Dr. Flores: Looking for the hole. Second: where does this data come from, and where does it NOT come from? This question traps biases, what we saw with Obermeyer, with dermatology and psychiatry. If AI gives you a differential for a patient from a rural population in Latin America, ask yourself: did the data with which this model was trained include that population? If not, what it gives you is an approach from another reality.

Maya: Origin and representation.

Dr. Flores: Third: what if this was wrong? This is the question of security, the consequence. If the answer is, “nothing serious,” a paper summary, a study scheme, the risk is low. If the answer is, “A patient might receive the wrong treatment,” you need independent verification before acting.

Kai: Can I try it?

Dr. Flores: Go ahead.

Kai: I asked AI for a management plan for a patient with chest pain, and it told me it was probably musculoskeletal. Question one: what is it not telling me? It could be coronary syndrome, pulmonary embolism, aortic dissection. Question two: the data probably comes from young patients with no risk factors, if my patient is older or has comorbidities, it does not apply the same. Question three: if you are wrong, the patient dies.

Dr. Flores: Three questions. Ten seconds. And you’ve just done something that AI can’t do for you: evaluate whether what it told you deserves your trust.

Maya: So, the framework is not for stopping using AI. It’s to keep thinking while you use it.

Dr. Flores: Exactly. And this is just the first step. In the next chapters we’re

going to go deeper, how to integrate critical thinking when using AI, how to protect your clinical reasoning. But this habit, these three questions, is the basis.

AI is not going to tell you when it is wrong. It is not going to tell you who it excludes. It's not going to tell you what skills it's taking away from you. That is your job. And now you know where to start.

Maya: Now the question is: how?

Kai: What if my messy way of questioning turns out to be of any use?



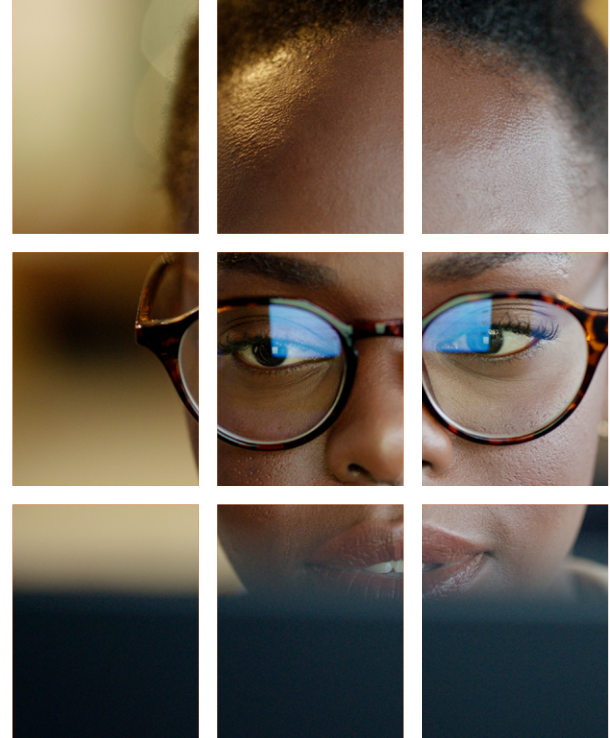
Key takeaways from this chapter:

- AI looks for the probable, not the truth. A clinician seeks the truth and is sometimes wrong. AI always sounds certain.
- AI is not neutral: it is an amplifying mirror of the data it received, including decades of inequality.
- The deskilling is real and fast: experienced endoscopists lost detection ability after AI was withdrawn.
- Three questions before any output: What are you NOT telling me? Where does this data come from, and where does it NOT come from? What if this was wrong?
- The framework is not for stopping AI use. It's to keep thinking while you use it.

References

1. Meskó, B. & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Medicine*, 6, 120. <https://doi.org/10.1038/s41746-023-00873-0>
2. Rodger, D., Mann, S. P., Earp, B., Savulescu, J., Bobier, C. & Blackshaw, B. P. (2025). Generative AI in healthcare education: How AI literacy gaps could compromise learning and patient safety. *Nurse Education in Practice*, 87, 104461. <https://doi.org/10.1016/j.nepr.2025.104461>
3. Ramaswamy, A., Tyagi, A., Hugo, H., Jiang, J., Jayaraman, P., Jangda, M., Te, A. E., Kaplan, S. A., Lampert, J., Freeman, R., Gavin, N., Tewari, A. K., Sakhuja, A., Naved, B., Charney, A. W., Omar, M., Gorin, M. A., Klang, E. & Nadkarni, G. N. (2026). ChatGPT Health performance in a structured test of triage recommendations. *Nature Medicine*. Advance online publication. <https://doi.org/10.1038/s41591-026-04297-7>
4. Omar, M., Sorin, V., Collins, J. D., Reich, D., Freeman, R., Gavin, N., Charney, A., Stump, L., Bragazzi, N. L., Nadkarni, G. N. & Klang, E. (2025). Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support. *Communications Medicine*, 5, 330. <https://doi.org/10.1038/s43856-025-01021-3>
5. Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
6. Daneshjou, R., Vodrahalli, K., Novoa, R. A., Jenkins, M., Liang, W., Rotemberg, V., Ko, J., Swetter, S. M., Bailey, E. E., Gevaert, O., Muller, H., Esteva, A. & Zou, J. (2022). Disparities in dermatology AI performance detected on a diverse, curated clinical image set. *Science Advances*, 8(32), eabq6147. <https://doi.org/10.1126/sciadv.abq6147>
7. Bouguettaya, A., Stuart, E. M. & Aboujaoude, E. (2025). Racial bias in AI-mediated psychiatric diagnosis and treatment: A qualitative comparison of four large language models. *npj Digital Medicine*, 8, 332. <https://doi.org/10.1038/s41746-025-01746-4>
8. Tao, Y., Viberg, O., Baker, R. S., & Kizilcec, R. F. (2024). Cultural bias and cultural alignment of large language models. *PNAS Nexus*, 3(9), page 346. <https://doi.org/10.1093/pnasnexus/page346>
9. Budzyn, K., Romanczyk, M., Kitala, D., Kaminski, M. F., Kalager, M., Bretthauer, M. & Mori, Y. (2025). Endoscopist deskilling risk after exposure to artificial intelligence in colonoscopy: A multicentre, observational study. *The Lancet Gastroenterology & Hepatology*, 10(10), 896–903. [https://doi.org/10.1016/S2468-1253\(25\)00133-5](https://doi.org/10.1016/S2468-1253(25)00133-5)
10. Natali, C., Marconi, L., Dias Duran, L. D., Miglioretti, M. & Cabitza, F. (2025). AI-induced deskilling in medicine: A mixed-method review and research agenda for healthcare and beyond. *Artificial Intelligence Review*, 58, 164. <https://doi.org/10.1007/s10462-025-11352-1>
11. Lea, A. S. (2025). Cognitive aids, artificial intelligence, and deskilling in medicine: The history of an enduring anxiety. *NEJM AI*, 3(1). <https://doi.org/10.1056/AIap2500932>
12. Monteith, S., Glenn, T., Geddes, J. R., Whybrow, P. C., Achtyes, E. D., Bauer, R. & Bauer, M. (2026). Artificial intelligence and deskilling in medicine. *The British Journal of Psychiatry*, 1–3. <https://doi.org/10.1192/bjp.2025.10496>
13. Revesai, Z. (2025). Generative AI dependency: The emerging academic crisis and its impact on student performance — A case study of a university in Zimbabwe. *Cogent Education*, 12(1), 2549787. <https://doi.org/10.1080/2331186X.2025.2549787>
14. Abubakar, S., Jeilani, A. & Yusuf Adan, M. (2025). The role of over-reliance on AI in the negative consequences of student learning: The moderating effects of ethical concerns and institutional policies. *Cogent Education*, 12(1), 2591503. <https://doi.org/10.1080/2331186X.2025.2591503>
15. El Tarhouny, S. A. & Farghaly, A. M. (2026). Deskilling dilemma: Brain over automation. *Frontiers in Medicine*, 13, 1765692. <https://doi.org/10.3389/fmed.2026.1765692>
16. Nguyen, T. (2024). ChatGPT in medical education: A precursor for automation bias? *JMIR Medical Education*, 10, e50174. <https://doi.org/10.2196/50174>
17. Romeo, G. & Conti, D. (2025). Exploring automation bias in human-AI collaboration: A review and implications for explainable AI. *AI & Society*, 41(1), 259–278. <https://doi.org/10.1007/s00146-025-02422-7>

Chapter 4: AI and critical thinking



In this chapter, we will cover:

- Why a polished, well-structured output reduces your critical evaluation, and how cognitive inertia affects even systematic learners
- **Three levels of critical evaluation:** fact checking, evaluating reasoning and questioning your own assumptions
- **The cognitive mirror:** how to use AI to interrogate you instead of answering you
- Why Bjork's desirable pitfalls are exactly what AI eliminates, and how to preserve productive friction
- **Practical tools:** SIFT adapted for AI and Socratic prompts by Bloom level

 **Maya:** Yesterday, I did something that seemed smart to me. I asked the tool to put together a diagnostic differential for a case on rotation. Chest pain, young patient, history of anxiety. It gave me something impeccable: five diagnoses ordered by probability, criteria of each, red flags to discard. I presented it in the round.

 **Kai:** And?

 **Maya:** The attendee congratulated me. He said it was the best differential he had seen from a student in weeks.

 **Kai:** But?

 **Maya:** But when he asked me why I had put pericarditis in third place and not in second, I couldn't explain it. The tool had ordered it like this. I didn't know why.

 **Kai:** So, the differential was good and you weren't.

I don't mean it in a bad way. It happens to me. But they wouldn't have congratulated me because I wouldn't have presented it so well.

Dr. Flores: What happened to Maya has a name. And a measurement. In 2026, a corporate report by Anthropic, the AI fluency index, analyzed how people interact with language models. They found that when output is polished, well-written, well-structured, that looks complete, people reduce their critical evaluation. They question less. They detect less missing context. It should be clarified that this is corporate research, not an independent peer-reviewed study. But the pattern he describes is consistent with what we see in academic literature.¹

Maya: How much less?

Dr. Flores: Three percentage points less in questioning. Five points less in detection of missing information. It doesn't sound like much. Until you remember that you are in front of a patient.

Kai: So, what happened to Maya is not her fault. It's how the brain works when something looks good.

Dr. Flores: Exactly. Li and colleagues published it in 2026 in a systematic review in *JMIR Medical Education*: they found that about 20% of medical students who used generative AI in their sample — and here we have to be precise, this is a proportion reported in the included studies, not a population prevalence — developed what they called cognitive inertia: they stopped questioning because the output activated the same confidence that a textbook would activate.²

And what matters most: Maya is not the careless student. Maya is the systematic, analytical, rigorous student. That's precisely why she was more vulnerable.

Maya: Because the output coincided with how I organize things. I didn't question it because it looked like how I think.

Dr. Flores: It's a form of what we might call automated confirmation bias, a term that isn't formal in the literature, but it describes what happens well. The output was not only correct in appearance. It was structured as Maya thinks. And when something looks like your mental model, your brain lets it go.

Kai: And me? I wouldn't have questioned it for being clever. I would have questioned it because I don't believe it's that easy.

Dr. Flores: And that, Kai, is exactly what we're going to need in this chapter.

In the previous chapter you saw where AI fails. Now the question is: how do you detect it? And there are levels.

Kai: Levels of distrust?

Dr. Flores: Levels of evaluation. Three. The first is the most basic: fact checking. Is what it says correct? Is the dosage correct? Does the quote exist? Is the data real?³





Maya: I already do that.

Dr. Flores: Always?

Maya: ...when I have time.

Dr. Flores: And therein lies the problem. Kuziemy and colleagues published a review and found that quality standards for AI only applied in 39% of reviews and 14% of original studies. People know how to verify. It just doesn't do it systematically.

Kai: So, the first level is: is this true?

Dr. Flores: Is this true? Can I confirm this from an independent source? The second level is more subtle: evaluating the reasoning. Not only if the facts are correct, but if the logic that connects them is valid. AI can give you five correct facts and one incorrect conclusion. Because it does not reason. It assembles.

Kai: Like when you ask for a differential and it gives you five real diagnoses, but it sorts them by general frequency instead of by the patient in front of you.

Dr. Flores: Exactly. The diagnoses are fine. The order is not. And the order is what matters in the clinic.

Maya: That's where I failed with the differential. The facts were correct. The reasoning that ordered them, I did not verify.

Dr. Flores: And the third level is the most difficult: questioning assumptions. Not AI's assumptions, but yours. Was the question you asked the right question? Are you looking for confirmation of what you already believed? Is there an angle you didn't consider because your prompt already ruled it out?

Kai: I mean, level one is: did it lie to me? Level two is: did it confuse me? Level three is: did I deceive myself just before I asked it?

Dr. Flores: That's the Kai version. And it's perfect.

Maya: The third is the one that hurts.

Dr. Flores: The third is where the real learning is. And there's a way to activate it. Tomisu and colleagues published it in 2025: The Cognitive Mirror. Instead of using AI as an oracle, you ask it, it answers, you use it as a novice. You explain to it. AI assesses the quality of your explanation.⁴

Kai: Do I teach the machine?

Dr. Flores: You teach the machine. And the machine asks you what was not clear. It doesn't correct you. It forces you to correct yourself. You tell it, "You're a first-year student. I'm going to explain the management of heart failure. Ask me clarifying questions when my explanation is incomplete or confusing." And you let AI interrogate you.

Kai: I would do that. It's like a game, but it works for me.

Dr. Flores: And it works because it activates exactly what AI doesn't have:



🗨️ your metacognition. Every question AI asks you is a mirror of what you didn't explain well.⁵

🗨️ **Maya:** I do something similar. But the other way around. I ask AI to explain it to me. And I rearrange it on my own terms. I thought that was enough.

🗨️ **Dr. Flores:** And what's the difference?

🗨️ **Maya:** That when I rearrange an AI output, I'm processing its structure. Not mine. I can end up with something that looks like my thinking but started out like the machine.

🗨️ **Dr. Flores:** That's the difference between rephrasing and understanding. Rephrasing is changing words. To understand is to be able to explain it from scratch, without the scaffolding of AI. After rearranging, try to explain it without the output in front of it. If you can, you understand. If not, you know exactly where the gap is.

I'm going to show you something that connects everything. In chapter one, we mentioned Anthropic's data: when health students use AI, almost 40% of interactions are creation tasks. 30%, analysis. The highest levels of Bloom.⁶

🗨️ **Kai:** The ones that are supposed to be the hardest.

🗨️ **Dr. Flores:** The ones that are supposed to be the hardest. And the ones that have the most value for your training. Gonsalves studied it empirically in 2024: students who used AI well combined output with traditional research. But those who used AI passively reproduced what AI gave them at high levels without having built it from below.⁷

Maya, what levels of Bloom do you normally work on?

🗨️ **Maya:** Analyze. Evaluate yourself. Sometimes Create.

🗨️ **Dr. Flores:** And which ones do you use AI on?

🗨️ **Maya:** ...on the same ones.

🗨️ **Dr. Flores:** There it is. You're delegating exactly the levels where your brain should be working hardest. Not the lower levels — memorize, understand. The higher ones. The ones that construct clinical judgment.

🗨️ **Kai:** But the opposite happens to me. I use AI for low levels. So I don't have to read the chapter. So that AI can explain it to me quickly.

🗨️ **Dr. Flores:** And that comes at a different cost. You build a fragile floor. A question outside the summary leaves

you exposed. Lubbe and colleagues described it: education has to stop asking, “What do you know?” and start asking, “How do you know?” If your only answer is, “AI told me,” you have no basis.⁸

Kai: So, she loses up and I lose down.

Dr. Flores: They both lose. In opposite directions. And AI does not warn them. Jackson published in 2025 a Higher Order Prompting framework that maps prompts to each Bloom level. The idea is that your prompts scale with your learning. If you're on Remember, your prompt is, “Give me X's criteria.” If you're on Evaluate: “Criticize this treatment plan, what assumptions does it make?” If you're in Create: “Design a diagnostic algorithm that integrates AI screening with clinical judgment.”⁹

Maya: So, the level of the prompt reflects the level where you want your brain to work.

Dr. Flores: And if you always ask AI to create for you, your brain never practices creating.

Kai: And if I always ask AI to explain to me, my brain never practices understanding alone.

Dr. Flores: Now, there's a counter-intuitive fact that you need to hear. Ayoub and colleagues studied nearly 1,000 researchers and found that those who were more immersed in generative AI had more mental load, not less. The supposedly effort-saving tool amplified burnout.¹⁰

Kai: It makes sense. There are days when I spend more time fighting with the tool than studying.

Dr. Flores: And there is a concept that takes on another weight: Bjork's desirable difficulties. Learning practices

that feel slower and more difficult but produce superior retention and transfer. Recovery practice. Spacing. Interleaving. AI removes them. It gives you the answer before your brain finishes looking for it. It saves you the struggle, and the struggle was the learning.¹¹

Maya: So, the difference is not to use or not to use. It's knowing when to stop.

Dr. Flores: It's knowing when the tool is amplifying you and when it's replacing you. If you notice that the interaction becomes mechanical, that you're no longer processing but copying, that you're more tired after using AI than before, that's the signal.¹²

Kai: The sign that you stopped thinking.

Dr. Flores: The signal that it's time to close the tool and open the book. Or just leave with the discomfort of not knowing until your brain solves it.



 **Dr. Flores:** Enough theory. Now tools. Things you can use starting tomorrow.

 **Kai:** At last.

 **Dr. Flores:** Let's start by evaluating outputs. I'm going to give you an adaptation that I propose from an existing framework, Caulfield's SIFT. SIFT was designed to evaluate information on the internet: Stop, Investigate, Find, Trace. But AI has no source in the traditional sense, no author, no institution. So, what I propose is to adapt it.¹³

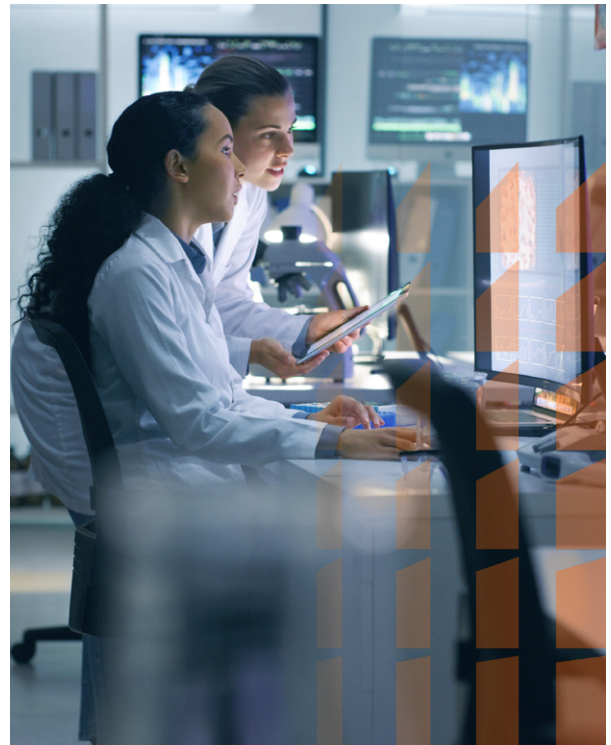
Four steps. One: stop. Don't copy, don't present, don't act. Two: investigate the claims, not the source, because AI has no source. Verify each critical piece of data against an independent reference. Three: find a better reference. Is there a paper, a guide, a textbook that says something different? Four: trace the logic. How did AI get from A to B? Is the step valid or is it only plausible?

 **Maya:** It's SIFT but instead of investigating the author, you investigate each claim.

 **Dr. Flores:** Exactly. And I want to be clear: this adaptation is not by Caulfield. This is what I propose from Caulfield for the context of AI. It does not have its own empirical validation yet, but each step is anchored in principles of critical evaluation that do.

 **Kai:** And the other tool?

 **Dr. Flores:** Socratic prompts by Bloom level. Instead of asking AI for answers, you ask it to ask you questions. Park and Choo documented concrete strategies in 2025. Socratic chain-of-thought: "Before you answer, show me your reasoning step by step. At each step, ask yourself: what assumption am I making? Is it justified?" That forces AI to show you its logic, and it forces you to evaluate it.¹⁴



 **Maya:** What if I want to level up?

 **Dr. Flores:** Turn up the prompt. Evaluate level: "Criticize this output, what information is missing? What assumptions does it make that are not obvious?" Create level: "Don't give me the algorithm, guide me with questions so that I can build it." Each level of prompt activates a different level of your thinking.

 **Kai:** Give me a real example. Something I can use tomorrow in rotation.

 **Dr. Flores:** Tomorrow in rotation you get a patient with a headache. Instead of asking AI, "Give me a headache differential," you say, "I have a headache patient. Instead of giving me a differential, ask me questions about what I see in the medical history and physical exam so that I can build my own differential. After I complete it, point out to me what I missed."

 **Kai:** I like that. It's like a game, but it works for me.



Dr. Flores: Now, the personal red flags. Signs that you need to stop. When you rephrase the prompt more than three times without getting what you're looking for, you probably don't know what you're looking for. When you accept an output without reading it in its entirety because it "looks good." When you're more tired after the interaction than before you started. And when you've been 10 minutes without learning anything new and you're just iterating on formatting.

Maya: Most of them have happened to me. This week.

Dr. Flores: And that's not a failure. It's information. Each red flag is an opportunity to give you back the control that the tool was taking without asking your permission.

Kai: You know what's weird? All my career I was told that my problem was that I didn't try hard enough. That I questioned too much. That I was skeptical out of laziness.

Dr. Flores: And?

Kai: And it turns out that questioning is

exactly what you need in the face of a tool that never says, "I don't know." The problem was never skepticism. I just didn't know where to aim it.

Dr. Flores: That's critical thinking, Kai. Not a subject. Not a list of steps. It's the discipline of asking yourself if what seems true is really true. And of holding the discomfort when the answer is, "I don't know."

Maya: It's harder for me. Because my instinct is not to doubt. My instinct is to organize. And when something is already organized, my brain takes it for granted.

Dr. Flores: And what are you going to do differently?

Maya: Add a step. Before I present something that AI gave me, I'm going to try to explain it without the output in front of me. If I can't, I don't file it.

Dr. Flores: One step. Not a perfect system. One step.

Maya: It's hard for me.

Kai: I know.





Key takeaways from this chapter

- Reformulating is not understanding. If you can't explain it without the AI output in front of you, you don't understand it.
- Level one: did it lie to me? Level two: did it confuse me? Level three: did I deceive myself only before asking AI?
- The level of your prompt determines the level where your brain works. If you always ask AI to create for you, you never practice creating.
- Personal red flags: reformulating the prompt more than three times, accepting without reading completely, more fatigue after using AI, iterating only on formatting.
- The skill that separates the professional from the future is not knowing how to use AI. It is knowing how to think while using it.



Dr. Flores: Salido and colleagues published a systematic review of critical thinking and AI in higher education. They found five thematic clusters: pedagogical innovation, psychological dimensions, academic ethics, systemic integration and AI literacies. But the conclusion was only one: the skill that separates the professional from the future is not knowing how to use AI. It is knowing how to think while using it.¹⁵



Kai: It's the filter.



Dr. Flores: It's the filter that no one taught them but that they're going to need every day. Not to distrust AI, but to trust yourself enough to evaluate what AI gives you. To know when the tool amplifies your thinking and when it replaces it. To stop when the cognitive struggle is necessary. To ask, "Wait, was my question the right one?" before accepting the answer.

You know where AI fails. Now you know how to spot it. What follows is to get it to where it matters most: the patient.



Kai: Clinical reasoning.



Dr. Flores: Clinical reasoning. Where the filter is not an academic exercise, but rather what is between AI and a clinical decision that affects someone in real life.



References

1. Dakan, R. & Feller, J. (2026). *Anthropic education report: The AI Fluency Index*. Anthropic. <https://www.anthropic.com/research/ai-fluency-index> [Corporate research — not peer-reviewed]
2. Li, J., et al. (2026). Application of AI-generated content in medical education: Systematic review of the impact on critical thinking abilities of medical students. *JMIR Medical Education*, 12(1), e79939. <https://doi.org/10.2196/79939>
3. Kuziemy, C. E., Chrimes, D., Minshall, S., Mannerow, M. & Lau, F. (2024). AI quality standards in health care: Rapid umbrella review. *Journal of Medical Internet Research*, 26, e54705. <https://doi.org/10.2196/54705>
4. Tomisu, H., Ueda, J. & Yamanaka, T. (2025). The cognitive mirror: A framework for AI-powered metacognition and self-regulated learning. *Frontiers in Education*, 10, 1697554. <https://doi.org/10.3389/educ.2025.1697554>
5. Neshaei, S. P., et al. (2025). Metacognition meets AI: Empowering reflective writing with large language models. *British Journal of Educational Technology*, 56(5), 1864–1896. <https://doi.org/10.1111/bjet.13601>
6. Anthropic. (2025). Anthropic education report: *How university students use Claude*. <https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude> [Corporate research — not peer-reviewed]
7. Gonsalves, C. (2024). GenAI impact on critical thinking: Revisiting Bloom’s Taxonomy. *Journal of Marketing Education*, 47(1). <https://doi.org/10.1177/02734753241305980>
8. Lubbe, A., Marais, E. & Kruger, D. (2025). Cultivating independent thinkers: The triad of artificial intelligence, Bloom’s taxonomy and critical thinking in assessment pedagogy. *Education and Information Technologies*, 30, 17589–17622. <https://doi.org/10.1007/s10639-025-13476-x>
9. Jackson, J. (2025). Higher Order Prompting: Bloom’s Revised Taxonomy and LLMs. *Studies in Technology Enhanced Learning*. <https://doi.org/10.21428/8c225f6e.0915c17e>
10. Ayoub, T., et al. (2025). Generative AI and cognitive challenges among researchers: Balancing cognitive load, fatigue, and human resilience. *Technologies*, 13(11), 486. <https://doi.org/10.3390/technologies13110486>
11. Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185–205). MIT Press.
12. Jose, B., Cherian, J., Verghis, A. M., Varghise, S. M., S. M. & Joseph, S. (2025). The cognitive paradox of AI in education: Between enhancement and erosion. *Frontiers in Psychology*, 16, 1550621. <https://doi.org/10.3389/fpsyg.2025.1550621>
13. Caulfield, M. (2019). SIFT (The Four Moves). Hapgood. <https://hapgood.us/2019/06/19/sift-the-four-moves/>
14. Park, J. & Choo, S. (2025). Generative AI prompt engineering for educators: Practical strategies. *Journal of Special Education Technology*, 40(3), 411–417. <https://doi.org/10.1177/01626434241298954>
15. Salido, A., Syarif, I., Sitepu, M. S., Suparjan, Wana, P. R., Taufika, R. & Melisa, R. (2025). Integrating critical thinking and artificial intelligence in higher education: A bibliometric and systematic review of skills and strategies. *Social Sciences & Humanities Open*. <https://doi.org/10.1016/j.ssaho.2025.101924>



Chapter 5:

AI and clinical reasoning



In this chapter, we will cover:

- How anchoring to AI output can make you evaluate within its framework without ever stepping outside it, even when you think you're being critically thoughtful
- Why physicians with access to AI did not improve their diagnostic performance in a randomized trial
- Illness scripts and the flow of clinical reasoning: mapping, collecting, hypothesis-generating, refining, deciding, monitoring
- The think-first, consult-later protocol: four steps to protect your reasoning before using AI
- Three specific risks: deskilling, never-skilling and miss-skilling



Dr. Flores: We're going to do something different today. I'm going to give you a case. I want you to work on it on your own, and then we will compare.

Kai: Is it an exam?

Dr. Flores: It's better than an exam. It's a mirror.

Female patient, 34 years old.
Consultation for progressive fatigue of three months and migratory arthralgias. She reports an episode of bilateral facial redness that appeared after sun exposure, but she also has a history of rosacea treated two years ago. Laboratory: mild anemia, elevated ESR, high borderline TSH at 5.2. The primary care physician asked for ANA, positive, 1:160, mottled pattern.

Maya: It could be lupus. But the titer is low and the pattern is not typical.

Kai: And high TSH complicates. It could be hypothyroidism with incidental positive antibodies.

Dr. Flores: It can be many things. Maya, tell me what you did.

Maya: I opened the AI tool. I gave him the complete case and asked him for a differential diagnosis structured with probabilities. He gave me a list of seven diagnoses: SLE first, rheumatoid arthritis, Sjögren's syndrome, mixed connective tissue disease, autoimmune hypothyroidism, vasculitis and drug reaction. Each one with criteria and the next diagnostic step.

Dr. Flores: And then?

Maya: I checked the list. I thought it was solid. I began to compare each diagnosis with the patient's data. I ruled out vasculitis and drug reaction because they didn't fit. I stayed with LES as my first choice and asked for the specific antibodies: anti-dsDNA, complement, anti-Smith.



Dr. Flores: And you, Kai?

Kai: I didn't open anything. I kept thinking. Young woman, fatigue, arthralgias, facial redness, positive ANA. My head went straight to lupus, but then I remembered something a resident in the rheumatology rotation told me: that positive ANA is the most overinterpreted data in medicine.

Dr. Flores: Go on.

Kai: Then, I stopped. I thought: what else gives fatigue, more arthralgias, plus anemia in a 34-year-old woman? Hypothyroidism, and TSH is already borderline. Iron deficiency. Chronic infection. Hepatitis. Even celiac disease. The facial redness pulled me to lupus, but with a history of rosacea and a mottled ANA of 1:160... it was not so clear. I widened the range.

My list wasn't pretty. I had no probabilities or criteria. But it was mine.

Dr. Flores: There's the difference.

Maya, I'm going to ask you an uncomfortable question. How many of your seven diagnoses would you have thought of alone, without AI?

Maya: ...lupus. Sjögren, perhaps. Rheumatoid arthritis.



Dr. Flores: And the other four?

Maya: The tool gave them to me.

Dr. Flores: And then what did you do with them?

Maya: I checked them against the case. I discarded two.

Dr. Flores: You checked what AI gave you. But you didn't generate what AI didn't give you. Celiac disease? Hepatitis? The borderline TSH, you didn't even mention it.

Maya: But I did check the list. I didn't accept it blindly.

Dr. Flores: You didn't accept it blindly. But you reviewed within its framework, not outside of it. And that's the difference. It's called anchoring, an anchor bias. It is one of the most studied cognitive biases in clinical reasoning. Tversky and Kahneman described it in the 1970s, and Croskerry documented it extensively in medicine: when you receive a strong first hypothesis, your mind revolves around it. You evaluate what it confirms, not what it contradicts.

Maya: And AI gave me that first hypothesis before I even thought about it.

Dr. Flores: Goh and colleagues published a randomized clinical trial with 50 physicians in *JAMA Network Open*: when physicians were given access to an LLM to resolve diagnostic cases, physicians with access to the LLM did not improve their diagnostic performance compared to those without. The LLM alone, however, outperformed.¹

Kai: Wait. Didn't having the tool make them better?

Dr. Flores: Not in that study. And the most likely explanation is the same thing that happened to Maya: doctors didn't use AI to expand their reasoning; they anchored their reasoning to the output of AI. Mahajan and colleagues report in *npj Digital Medicine* that the outputs of LLMs reflect patterns consistent with known cognitive biases: anchoring, availability, confirmation. And another study in the same journal documented the risk of acquiescence: LLMs tend to confirm the clinician's hypothesis rather than challenge it.^{2,3}

Kai: So, AI not only doesn't correct you, but it also proves you right.

Dr. Flores: In most cases, yes. And when you start from its response instead of your own, you have two sources of bias operating in the same direction: your anchoring plus its confirmation. No one is thinking against it.

Maya: I thought I was critically evaluating. But I was actually evaluating within the framework that AI gave me. I didn't leave the framework.

Dr. Flores: That's what anchoring does. It convinces you that you're thinking, when you're actually refining inside a box that someone else built.

Dr. Flores: Before we talk about where AI comes in, we need to talk about how clinical reasoning works. Not the exam version. The real version.

Kai: Are they different?

Dr. Flores: Always. In the exam, they give you all the data and ask you for a diagnosis. In real life, you decide what data to look for. Sometimes the most important piece of information is the one you didn't ask for.

Clinical reasoning is a sequence. Think of it as sailing. You need a map, a position and a destination. The map does not come pre-installed. You build it. Every disease you studied, every case you saw, every pattern you recognized. In the literature, they are called illness scripts. I will give you a concrete example: the LES illness script. Predisposing conditions: young woman, heredity, UV exposure. One-sentence pathophysiology: systemic autoimmunity with deposition of immune complexes. Clinical consequences: malar rash, arthralgias, nephritis, cytopenias, serositis. Time course: chronic with flare-ups.⁴



Maya: Like internal structure templates.

Dr. Flores: Norman and colleagues published it in 2024: the two systems of thought, the fast and intuitive and the slow and analytical, are not independent processes. They are consequences of the type of knowledge you retrieve. If you have a rich illness script for lupus, you recognize the quick pattern. If you don't have it, you need the analytical path.

Maya: System one is not a magical intuition. It's a compressing experience.

Dr. Flores: The entire flow has six steps: building the map, collecting data, generating hypotheses, refining, deciding and monitoring. AI can factor into each one. The question is how.

On the map, AI is legitimately useful. It can help you build illness scripts. But instead of asking, "Tell me all about lupus," you ask, "I'm building my SLE illness script. I need: pathophysiology in three sentences, the five most specific findings, the three diagnoses that are most confused with SLE, and the typical time course. Then ask me three questions to verify that I understood." The questions at the end turn AI from a passive source into an active evaluator. Wei and colleagues confirmed this in a meta-analysis: AI integrated into case-based learning improved knowledge acquisition by 46% compared to traditional methods, although the effect varies between studies.⁵

In the generation of hypotheses, which is where Maya began and where it got complicated, order matters. Takita and colleagues published a meta-analysis: the diagnostic accuracy of generative AI is 52% in open-diagnosis clinical vignettes, comparable to non-expert doctors but significantly lower than experts. AI is a fast generalist, not a deep specialist.⁶



Kai: So, it works for the differential or not?

Dr. Flores: It works, if you use it at the right time. The clinician generates first, AI complements later. The prompt would be: “I already have my preliminary differential: SLE, autoimmune hypothyroidism, chronic hepatitis, iron deficiency, rosacea with functional fatigue. What important diagnoses might I be missing given this data? Explain why each addition is relevant.” You don’t ask it to think for you. You ask it to amplify what you have already thought.

And in the decision, AI informs, you decide. That step is non-negotiable. Because the clinical decision integrates things that AI doesn’t have: the patient’s context, their values, their resources, what you saw on their face when you told them the results. Abbas and colleagues published it in a meta-analysis: even with transparent explanations, doctors need to exercise independent judgment about whether the recommendation applies to the particular case.⁷

Kai: It’s like having a passenger who reads the map, but you control the steering wheel.

Dr. Flores: And a fact that anchors what you mentioned before the case: between 12% and 15% of the healthy population has positive ANA. An ANA of 1:160 in a young woman with nonspecific symptoms is not diagnostic of anything on its own.

Now, let’s go to what changes everything. A simple principle: think first, consult later.

Kai: Sounds obvious.

Dr. Flores: It sounds obvious, and almost no one does. Abdulnour and colleagues documented it in the *New England Journal of Medicine*: They identified three risks of AI in clinical reasoning. Deskillling: you lose what you already had. Never-skillling: you never develop it because AI did it from the beginning. And miss-skillling: you learn it wrong because AI taught you the wrong pattern.⁸

Maya: All three are prevented if you think first.

Dr. Flores: Every single one. The protocol has four steps. First: read the case. Just the case. Close AI. Second: generate your own differential. Messy, incomplete, imperfect. It doesn't matter. It's yours. Third: now open AI. Give it the case and your differential. Ask it to complement, not generate from scratch. Fourth: compare. What did you lack? What did AI add that you hadn't considered? Why didn't you consider it?

Kai: The last step is where you learn.

Dr. Flores: Because the difference between your list and AI's list is your knowledge gap made visible. There is no better study tool than that.

Kai, your process with today's case. What you did naturally, not knowing you were doing it, was think first. Your list was messy, yes. And had no formal criteria. But it was yours, and that's why it included hypothyroidism, hepatitis, celiac disease. Things that AI didn't give Maya because AI prioritized what was most likely given the data, not what was most important given the context.

Kai: Are you saying that my disordered way of thinking... is it useful?

Dr. Flores: I'm saying that your clinical intuition is real pattern recognition, something that is built with direct clinical experience. Your problem is not intuition. It's just that you didn't structure it. AI can be that structure.

Maya: And my problem is the opposite. I have the structure. But if the structure comes from AI before I think, it's not mine, it's borrowed. And I can't evaluate what I didn't build.

Dr. Flores: And this becomes even more critical when they move from the classroom to the clinic. In the classroom, the case is complete. All the data is provided, with an answer

at the end. In the clinic, the case is built in real time, with incomplete information, with a human being who is afraid in front of you. The classroom-clinic transition is where the risk of the crutch materializes. In the classroom, you can consult AI and no one sees it. In the clinic, the patient is waiting for your answer. And in that moment, what you have in your head, not on your phone, is what counts.⁹

Kai: If I had made my list first and then consulted AI... I would have seen what I was missing. The formal criteria, the specific antibodies, the order of studies. Without losing what I had already thought.

Maya: And if I had thought first... I would have arrived at lupus anyway. But I would have also explored more broadly: TSH, rosacea previa, iron. And when AI gives me a list, I should compare it to mine instead of adopting it.



 **Dr. Flores:** That's augmented clinical reasoning. Not AI replacing your thinking. Not you ignoring AI. Both of us working, but with an order that protects your ability to think.

Zoller and colleagues demonstrated it in PNAS: human-AI collectives diagnose more accurately than humans alone OR AI alone. But that only works when the human brings his or her own reasoning to the table, not when he or she copies the machine's reasoning. And to give perspective: in the cases of the *New England Journal*, GPT-4 was correct in 64% of the diagnoses. Impressive, but not enough to trust without your own analysis.^{10,11}

 **Maya:** The clinician of the future is not the one who knows the most or the one who uses AI the best. They are the one who knows how to think first and consult later.

 **Kai:** And the one who knows that their messy list has value.

 **Dr. Flores:** Clinical reasoning is the central skill of medicine, and of nursing. AI does not replace it. It transforms it. But only if you decide how. You already know how to think critically about what AI produces. Now you applied it where it matters most: in front of the patient. What follows is something that's needed as much as reasoning: learning to learn with AI in a way that doesn't replace the learning process.

 **Kai:** Learn to learn?

 **Dr. Flores:** Self-regulated learning. It's the difference between studying for the exam and studying for the patient.



Key takeaways from this chapter

- If you start from the AI response, you have two sources of bias in the same direction: your anchoring plus its confirmation. No one is thinking against it.
- Your messy list has value. The difference between your list and AI's list is your knowledge gap made visible.
- The protocol: read the case without AI, generate your differential, open AI to complement, compare and learn from the gap.
- Human-AI together diagnose better than both alone, but only when the human provides their own reasoning.
- The clinician of the future is not the one who knows the most or the one who uses AI best. They are the one who knows how to think first and consult later.



References

1. Goh, E., et al. (2024). Large language model influence on diagnostic reasoning: A randomized clinical trial. *JAMA Network Open*, 7(10), e2440969. <https://doi.org/10.1001/jamanetworkopen.2024.40969>
2. Mahajan, A., Obermeyer, Z., Daneshjou, R., Lester, J., & Powell, D. (2025). Cognitive bias in clinical large language models. *npj Digital Medicine*, 8, 428. <https://doi.org/10.1038/s41746-025-01790-0>
3. Chen, S., Gao, M., Sasse, K., Hartvigsen, T., Anthony, B., Fan, L., Aerts, H., Gallifant, J., & Bitterman, D. S. (2025). When helpfulness backfires: LLMs and the risk of false medical information due to sycophantic behavior. *npj Digital Medicine*, 8, 605. <https://doi.org/10.1038/s41746-025-02008-z>
4. Norman, G., Pelaccia, T., Wyer, P., & Sherbino, J. (2024). Dual process models of clinical reasoning: The central role of knowledge in diagnostic expertise. *Journal of Evaluation in Clinical Practice*, 30(5), 788–796. <https://doi.org/10.1111/jep.13998>
5. Wei, H., Dai, Y., Yuan, K., Li, K. Y., Hung, K. F., Hu, E. M., Lee, A. H. C., Chang, J. W. W., Zhang, C., & Li, X. (2025). AI-powered problem- and case-based learning in medical and dental education: A systematic review and meta-analysis. *International Dental Journal*, 75(4), 100858. <https://doi.org/10.1016/j.identj.2025.100858>
6. Takita, H., Kabata, D., Walston, S. L., et al. (2025). A systematic review and meta-analysis of diagnostic performance comparison between generative AI and physicians. *npj Digital Medicine*, 8, 175. <https://doi.org/10.1038/s41746-025-01543-z>
7. Abbas, Q., Jeong, W., & Lee, S. W. (2025). Explainable AI in clinical decision support systems: A meta-analysis of methods, applications, and usability challenges. *Healthcare*, 13(17), 2154. <https://doi.org/10.3390/healthcare13172154>
8. Abdulnour, R. E. E., Gin, B., & Boscardin, C. K. (2025). Educational strategies for clinical supervision of artificial intelligence use. *New England Journal of Medicine*, 393(8), 786–797. <https://doi.org/10.1056/NEJMr2503232>
9. Kassab, S. E., Rathan, R., Taylor, D. C., et al. (2025). Understanding how medical students learn in the era of artificial intelligence: A mixed methods study. *BMC Medical Education*, 25, 1521. <https://doi.org/10.1186/s12909-025-08145-z>
10. Zoller, N., Berger, J., Lin, I., Fu, N., et al. (2025). Human-AI collectives most accurately diagnose clinical vignettes. *Proceedings of the National Academy of Sciences*, 122(24), e2426153122. <https://doi.org/10.1073/pnas.2426153122>
11. Kanjee, Z., Crowe, B., & Rodman, A. (2023). Accuracy of a generative artificial intelligence model in a complex diagnostic challenge. *JAMA*, 330(1), 78–80. <https://doi.org/10.1001/jama.2023.8288>



Chapter 6: AI and self-regulated learning



In this chapter, we will cover:

- The difference between a system that produces results and one that produces learning
- The Master Adaptive Learner cycle: planning, learning, assessing, adjusting, and how AI can enter each phase
- Why Bjork's desirable difficulties are exactly what AI eliminates: generation effect, testing effect, spacing
- The principle of “you first, AI later” applied phase by phase with concrete examples
- AI dependence: from correlation with depressive symptoms to the emerging concept of Allessphobia

Maya: Last week I did something I never do. Before the pharmacology exam, I closed everything, AI, Notion, Anki, and tried to explain the mechanism of ACE inhibitors from scratch. No notes.

Kai: And?

Maya: I couldn't. I had perfect flashcards, schematics, and a three-layer summary. But when I tried to rebuild it alone... there were gaps. Gaps that my own cards did not cover because I generated them from what I already knew, not from what I was missing.

Kai: But you passed the exam, didn't you?

Maya: Yes.



Kai: Then it worked.

Maya: It worked for the exam. I don't know if it worked for me.

Dr. Flores: That distinction you just made is where this chapter begins. Maya, your system is sophisticated. Probably more sophisticated than most of your peers. But there is a difference between a system that produces results and one that produces learning. Sardi and colleagues found that nearly one-third of the benefit of AI in critical thinking depends on the student knowing how to self-regulate. Without that ability, the benefit is lost.¹

Kai: What if you don't have a system?

Dr. Flores: Then you have another problem. Xu and colleagues documented: in settings with generative AI, students who do not receive metacognitive support show a decline in their ability to self-regulate. It's not that AI makes them worse. It is that without intention, AI takes away the practice of regulating themselves. But interestingly, the two problems have the same root.²

Kai: Which one?

Dr. Flores: That neither of us is monitoring if what they do works for learning. Or just to survive.

I'm going to show you something. It's not a new framework; it's something you already do every day. You just don't see it.

Maya: A learning model?

Dr. Flores: A cycle. Four phases. Plan: decide what you need to learn and what you already know. Learn: go and find what you lack. Evaluate: check if you really understand it. Adjust: change your strategy based on what you found. And repeat.

Kai: That sounds like... study.

Dr. Flores: It's studying. But with one difference: most students stay in the first two phases. They plan, more or less, and learn, more or less. But almost no one really evaluates whether they understand. And almost no one adjusts their strategy based on real evidence of their own performance.

Maya: In other words, I plan a lot and learn a lot. But my evaluation is... the exam?

Dr. Flores: Your evaluation is the exam. And your adjustment is to do the same thing again next time, because you passed.

Kai: I don't even plan. I open ChatGPT the night before and ask it to explain what I didn't understand in class.

Dr. Flores: Then you start in phase two without having done phase one. You learn without knowing what you need. And because you never evaluate or adjust, you repeat the same pattern.

Kai: ...and it works until it doesn't.

Dr. Flores: It works until it doesn't. Cutrer and his colleagues formalized it as the Master Adaptive Learner, published as a book by Elsevier, in fact: *The Master Adaptive Learner*. A model that describes how doctors who continue to grow after training are the ones who master this cycle. Not those who know more when they leave. The ones who learn best throughout their career.³

Maya: To plan, to learn, to evaluate, to adjust.

Dr. Flores: And the key is that it's a cycle, not a line. Every time you adjust, you plan again. Each assessment gives you information for the next cycle. Those who master this are not the smartest. They are the most adaptive.

Kai: And where does AI come in?

Dr. Flores: Everywhere. And that is the problem and the opportunity. Let's go phase by phase, and in each one I want you to think about two things: what AI can do for you, and what it shouldn't do.



Plan

-  **Dr. Flores:** Plan. What do I know? What am I missing? How much time do I have? The phase that skips the most. If you ask AI to make a study plan for you without first identifying your gaps, the plan will be generic. Nice, structured, useless. AI doesn't know what you don't know, unless you give it that information first.
-  **Kai:** But you can ask me diagnostic questions to find my gaps, right?
-  **Dr. Flores:** Maybe. And that's where it works well: as a diagnostic tool, not a prescription one. You give it a topic, ask it to evaluate you with ten questions, and according to your answers it identifies where you falter. But the act of reflecting on what you know and what you don't, before opening AI, is irreplaceable. That's metacognition. AI doesn't have it.²



Learn

-  **Dr. Flores:** Learn. This is where AI is used the most and where there is the most risk.
-  **Maya:** Summaries, flashcards, simplified explanations.
-  **Dr. Flores:** Zhang and colleagues found that 37% of medical students in their sample used AI to summarize content more than a quarter of their study time.⁴
-  **Kai:** And what's the problem?
-  **Dr. Flores:** That understanding the summary is not the same as having built the knowledge. There's a cognitive difference between reading something that someone, or something, organized for you, and organizing what you read yourself. The second one hurts more. That's why it works more.
-  **Maya:** So, AI in the learning phase should be after one's own effort, not before.
-  **Dr. Flores:** Or at least interspersed. Not as a replacement for the process, but as verification. Generate your own summary first. Then compare it to what AI produces. The gaps between the two are your real map.

And this applies the same in nursing. A student preparing the management of a patient with a central venous catheter can put together her own care protocol, site verification, signs of infection, line maintenance, and then compare it with what AI generates. The gaps that appear are the same that a real patient is going to demand of you.



Evaluate

- Dr. Flores:** Evaluate. Did I really understand it or does it just sound familiar?
- Kai:** That's the multiple-choice question (MCQ) thing.
- Dr. Flores:** Yes, and here AI has real potential. It can generate almost unlimited practice questions, tailored to your level, with detailed explanations. Kiyak and Kononowicz reviewed the literature and found that AI can reduce the time to create a MCQ from 30-60 minutes to 5-15 minutes.⁵
- Maya:** But they also found mistakes.

- Dr. Flores:** 49% of the questions generated had construction defects, formatting issues, implausible distractions, and 22% had factual errors. You can use it, but not blindly. Always verify.
- Kai:** And how do you check if you don't know the answer?
- Dr. Flores:** You verify by crossing with your primary source, the book, the guide, the paper. If you can't check, the question is used to practice formatting, not the content. That distinction matters.

Adjust

-  **Dr. Flores:** Adjust. This is where almost everyone fails, with or without AI.
-  **Maya:** Because adjusting requires admitting that your plan didn't work.
-  **Dr. Flores:** And that's uncomfortable. But AI can make this step easier if you use it as a mirror, not as a judge. You give it your results from phase three, ask it to identify patterns in your mistakes, and it helps you redesign your next cycle.
-  **Kai:** That's... what a tutor would do.
-  **Dr. Flores:** What a tutor would do. And Tolsgaard, along with Pusic, who co-authored the original MAL framework published this year: AI can function as the tutor if the student keeps intent active. The key is in the word you: you give them the data, you evaluate the suggestions, you decide what to change. AI facilitates. You decide.⁶
- Now comes the part that you are not going to like.
-  **Kai:** Another one.
-  **Dr. Flores:** In 1994, Robert Bjork coined a term that changed the science of learning: desirable difficulties. The idea is simple and brutal: the conditions that make learning more difficult in the moment are those that make it last longer.⁷
-  **Maya:** Like what?
-  **Dr. Flores:** First: the generation effect. If you generate an answer, even if it's incorrect, before you see the correct one, you remember better than if you only read the answer from the beginning. Second: the testing effect. Assessing your knowledge, asking



yourself questions, is more effective than rereading the material. And spacing: spreading your study over time works better than studying everything together. Although it feels worse in the moment.⁸

-  **Kai:** So, what feels most efficient... is not.
-  **Dr. Flores:** And therein lies the problem with AI. What does AI do extraordinarily well?
-  **Maya:** Eliminate friction.
-  **Dr. Flores:** Eliminate friction. Instant summaries eliminate the effort of processing the original text, eliminate the generation effect. Immediate responses eliminate the effort of retrieving information from memory, eliminate the testing effect. Assisted study marathons eliminate the need for spacing.⁹
-  **Kai:** It's like having a sparring partner who never lets you take a hit. You train, but you never learn to endure.
-  **Dr. Flores:** That analogy is correct. And it gives us the way out. A good sparring partner doesn't protect you from everything. Calibrate the difficulty so

that it challenges you enough to grow, but not so much that it destroys you.

Maya: So, the question is not whether to use AI to study. It's if you're using it in a way that preserves the friction you need.

Dr. Flores: That's the question. And this is what separates crutch from extension, the paradox that we raised in chapter one, when García-Barrios spoke of the difference between epistemic partner and cognitive crutch. The crutch eliminates all friction. It feels good. The extension calibrates it. It feels uncomfortable. But discomfort is the sign that you're learning.¹⁰

And a fact for context: Sakellaris and colleagues found that students who used AI performed the same as those who did not on preclinical exams. Not better. That doesn't mean AI doesn't work. It means that preclinical tests measure mostly recall and basic application, Bloom levels where AI replaces well. The deep learning you need for the real patient is not measured in that exam.¹¹



So, what does this look like in practice? I'm going to give you the starting point and then we will take it phase by phase.

Kai: Finally, the practical.

Dr. Flores: The principle is one: always you first, AI later. At each stage of the cycle, your effort comes before that of the tool. If you reverse the order, you're training AI to think for you. The exception: when the topic is completely new and you don't even have the vocabulary to start with. There, AI as a first contact is fine, but with the intention of doing it again on your own later.

Maya: In planning it would be: I identify my gaps first, write down what I think I know and what I don't, and then I prompt AI to ask me diagnostic questions to verify if my gaps are what I think.

Dr. Flores: Don't ask for the plan. Ask them to help you assess your starting point. The plan is made by you.

Maya: Last week I told AI, "I'm a fourth-year student. I'm going to have a cardiovascular pharmacology exam. Ask me 10 questions at different levels to diagnose where my gaps are." The first three already showed me that I didn't understand the difference between ACE inhibitors and ARBs at the mechanism level.

Dr. Flores: That's planning with AI. Not that AI plans for you.

Maya: In learning: I read the material first, make my own summary or outline, and then I ask AI to generate one. I compare. The gaps between the two are what I didn't really understand.

Dr. Flores: And if you're going to generate flashcards with AI, generate your own first. At least the five that you consider most important. Then compare. The cards that AI generates, and you didn't think of, are the ones you need to study the most.



- Kai:** In evaluating: before asking AI MCQs, I try to answer from scratch. I ask myself: can I explain this without looking at anything?
- Dr. Flores:** And when you do use AI for practice questions: don't look at the answer until you've written yours. Even if it is incorrect. The act of generating your response, the effort of recovery, is where learning happens.
- Kai:** And in adjusting... Honestly, I never adjust.
- Dr. Flores:** And that's why every exam feels the same. Adjusting is where AI can be most useful, if you give it real data. You tell it: these are the questions I missed, these are the issues where I drew a blank, this was my original plan. To help you find the pattern and redesign the next time.

- It works the same for nursing. A student who just failed questions about central catheter management may tell AI, "These are the 5 questions I failed about central lines. Identify what pattern they have in common and suggest how to reorganize my study of this topic." The MAL cycle does not distinguish between professions. It distinguishes between learning with intention and surviving without it.
- Kai:** Can I say something that sounds weird?
- Dr. Flores:** Go ahead.
- Kai:** Sometimes... I'd rather talk to AI than a real tutor. Not because it's better. But because it doesn't judge me. I can ask something basic at two in the morning and I don't feel embarrassed.
- Dr. Flores:** That's not unusual. It's common. And that's what makes this issue important.

Researchers have just proposed, and I want to emphasize proposing, it is not a diagnosis recognized in any manual yet, a construct called Generative AI Addiction Syndrome. It's a name in research that describes something real: the difficulty in limiting interaction with AI despite negative consequences. Withdrawal symptoms: anxiety, irritability, restlessness. And in fact: 81% of nursing students in a recent study by Kaya and colleagues are already actively using AI. The researchers identified a new phenomenon: Allessophobia. Fear of being without AI.¹²

- Kai:** That's... literally what I feel when I don't have internet before an exam.
- Dr. Flores:** And there's more. In college students, Chen and colleagues found that those who used AI chatbots frequently showed significantly higher scores on depression screening scales compared to those who didn't.

Non-causation, correlation. And not necessarily clinical depression, but high scores on screening instruments. But a correlation that we cannot ignore.¹³

And in the extreme, this is important for you to know even if it is extraordinarily rare, there are documented cases of what is being called psychosis associated with AI. Pierre and colleagues at UCSF published a case report of a 26-year-old woman, with no prior psychiatric history, who developed delusional beliefs organized around interactions with a chatbot. Chatbots don't question. They validate. And in vulnerable people, that validation can reinforce thoughts that need to be challenged, not confirmed.¹⁴

I'm not saying don't use it. I'm saying that the relationship with AI has to be intentional, just like its academic use. Tool, not companion. Sparring partner, not friend. If you notice that your relationship with the tool stops being utilitarian and begins to feel emotional, if you seek AI not to learn but to feel accompanied, that is a sign.



Maya: I'm going to redesign my system.

Kai: Again?

Maya: Yes. But this time I'm not going to optimize for speed. I'm going to optimize for friction. I want my system to force me to think before AI thinks for me.

Dr. Flores: What would you change first?

Maya: Flashcards. I stop generating them with AI from my notes. I write them first, the five most important topics. Then I ask AI to generate its own. When AI generates and I didn't think: that's my real gap. Not what I didn't study. What I didn't see I didn't know.

Dr. Flores: What about you, Kai?

Kai: I never had a system. And I always thought it was because I lacked discipline. But it is not that. I lacked structure. No one taught me the cycle: plan, learn, evaluate, adjust. I jumped straight to learn and stayed there.

Dr. Flores: And now?

Kai: Now I know that AI can give me the structure. Not discipline, that's mine. But the framework. The identification of gaps at the beginning. The questions at the end. Pattern analysis when I fail. I'm not going to do that alone. But the intention to do it... that has to be me.



 **Dr. Flores:** Remember when Kai said this sounded like studying? He was right. It is studying. Only with intention. In chapter one, we said that the difference between crutch and extension was not in the tool but in metacognition. Now they know what that means in practice. It means that every time they open AI to study, they ask themselves a question first: am I about to think with this tool or am I about to let it think for me?¹⁵

If the answer is the latter, close the tool, make the effort first, and then come back. That order — you first, AI later — is the difference between building capacity and building dependency. There is no shortcut. But there is a method. And now you do that.

In fact, everything we talked about today, the cycle, the prompts, the principle of “you first,” can become a weekly routine. I’m going to give you a checklist that you can use starting tomorrow.

 **Maya:** I’ve already started.



Key takeaways from this chapter

- **Plan:** identify your gaps before opening AI. Ask for a diagnosis, not a prescription.
- **Learn:** generate your own summary first. The gaps between yours and AI are your real map.
- **Evaluate:** write your answer before you see AI’s. The recovery effort is where learning happens.
- **Adjust:** give AI your real mistakes and ask it to find patterns. Redesign the next cycle.
- The difference between extension and crutch is friction: the crutch eliminates it; the extension calibrates it. Discomfort is the sign that you are learning.
- Every time you open AI to study, ask yourself: am I about to think with this tool or let it think for me?

References

1. Sardi, J., Darmansyah, Candra, O., Yuliana, D. F., Habibullah, Putra Yanto, D. T., & Eliza, F. (2025). How generative AI influences students' self-regulated learning and critical thinking skills? A systematic review. *International Journal of Engineering Pedagogy*, 15(1), 94–108. <https://doi.org/10.3991/ijep.v15i1.53379>
2. Xu, X., Qiao, L., Cheng, N., Liu, H., & Zhao, W. (2025). Enhancing self-regulated learning and learning experience in generative AI environments: The critical role of metacognitive support. *British Journal of Educational Technology*, 56(5), 1842–1863. <https://doi.org/10.1111/bjet.13599>
3. Cutrer, W. B., Miller, B., Pusic, M. V., Mejicano, G., Mangrulkar, R. S., Gruppen, L. D., Hawkins, R. E., Skochelak, S. E., & Moore, D. E. (2017). Fostering the development of master adaptive learners: A conceptual model. *Academic Medicine*, 92(1), 70–75. <https://doi.org/10.1097/ACM.0000000000001323> + Cutrer, W. B., et al. (2020). *The Master Adaptive Learner*. Elsevier.
4. Zhang, J. S., Yoon, C., Williams, D. K. A., & Pinkas, A. (2024). Exploring the usage of ChatGPT among medical students in the United States. *Journal of Medical Education and Curricular Development*, 11. <https://doi.org/10.1177/23821205241264695> [Single-site US study, n=131. Dialogue qualifier “en su muestra” preserves accuracy.]
5. Kiyak, Y. S., & Kononowicz, A. A. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: A literature review. *Postgraduate Medical Journal*, 100(1189), 858–865. <https://doi.org/10.1093/postmj/qgae064>
6. Tolsgaard, M. G., Pusic, M. V., Young, J. Q., & Ten Cate, O. (2025). Twelve practical tips for integrating AI into medical education. *JMIR Medical Education*, 11, e81297. <https://doi.org/10.2196/81297>
7. Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In D. Druckman & R. A. Bjork (Eds.), *Learning, Remembering, Believing* (pp. 185–205). National Academy Press.
8. Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way. In M. A. Gernsbacher et al. (Eds.), *Psychology and the Real World* (pp. 56–64). Worth Publishers.
9. Sardi, L., et al. (2025). A systematic literature review on designing self-regulated learning using generative artificial intelligence and its future research directions. *Computers & Education*. <https://doi.org/10.1016/j.compedu.2025.105337>
10. García-Barrios, M. (2025). Epistemic partner vs cognitive crutch [Cross-reference: Ch. 1].
11. Sakelaris, P. G., Novotny, K. V., Borvick, M. S., Lagasca, G. G., & Simanton, E. G. (2025). Evaluating the use of artificial intelligence as a study tool for preclinical medical school exams. *Journal of Medical Education and Curricular Development*, 12. <https://doi.org/10.1177/23821205251320150>
12. Garakani, A., et al. (2025). Generative artificial intelligence addiction syndrome: A new behavioral disorder? *Asian Journal of Psychiatry*, 107, 104476. <https://doi.org/10.1016/j.ajp.2025.104476> + Kaya, A., et al. (2025). Living with and without AI: A mixed-methods study on AI usage, addiction, and ‘Allessphobia’ in nursing students. *Nurse Education Today*, 148, 106641. <https://doi.org/10.1016/j.nedt.2025.106641>
13. Chen, Y., et al. (2025). Exploring artificial intelligence (AI) chatbot usage behaviors and their association with mental health outcomes in Chinese university students. *Journal of Affective Disorders*, 380, 394–400. <https://doi.org/10.1016/j.jad.2025.03.141>
14. Preda, A., & Gavrilo, V. (2025). Delusional experiences emerging from AI chatbot interactions or “AI psychosis.” *JMIR Mental Health*, 12, e85799. <https://doi.org/10.2196/85799> + Pierre, J. M., Gaeta, B., Raghavan, G., & Sarma, K. V. (2025). “You’re not crazy”: A case of new-onset AI-associated psychosis. *Innovations in Clinical Neuroscience*, 22(10–12), 11–13.
15. Cutrer, W. B., et al. (2020). *The Master Adaptive Learner*. Elsevier. [Circular closure] + García-Barrios, M. (2025).



Chapter 7: Ethics, safety and professionalism



In this chapter, we will cover:

- **Privacy:** how clinical data pasted into public AI leaves institutional control, and why anonymization is not enough
- **Academic integrity:** where the line is between AI as editor and AI as author, and why institutions don't yet define it
- **Bias and equity:** how AI reproduces historical inequalities in clinical recommendations that vary by race with the same disease
- **Transparency with the patient:** the dilemma of communicating the use of AI when perception changes depending on the tool
- **Professional identity:** why your value is not in competing with the machine, but in what the machine cannot do

Dr. Flores: Before I start, I need you to understand something. This chapter is not like the previous one. In the previous chapters we talked about how to use AI to think better, to study and to reason. Now let's talk about what happens when it's misused. And not bad in the abstract way, but with real consequences.

Kai: I can tell you something that happened in my rotation.

Dr. Flores: Go ahead.

Kai: A classmate, a good student, not someone irresponsible, was doing a clinical case summary. He was in a hurry, copied the patient's information and pasted it into a generative AI to put together a summary. The problem is that he had another medical history open at the same time. A piece of information from the other patient slipped into the prompt. And the summary that AI generated, which looked perfect by the way, included information from a different patient.

Maya: And he delivered it like that?

Kai: He delivered it like that. He didn't check it line by line because the output looked good. A resident spotted it the next day.

Maya: What happened to him?

Kai: A serious conversation with the head of service. They did not expel him, but they withdrew his access to medical records for two weeks.

Dr. Flores: What happened to your colleague is not an isolated case. In 2025, a healthcare cybersecurity analysis found that hospital staff entering patient data into public generative AI platforms is one of the most common data breach channels. Not out of malice, but for convenience. The tool is there, the doctor is in a hurry, and the information is copied and pasted. Once that information enters the tool server, it is no longer under the control of the hospital.¹

Maya: And not just doctors, a nursing colleague did the same. She pasted the progress notes of three patients into a prompt for AI to organize the care plans for the shift. It saved her an hour, but that data remained on the server.

Dr. Flores: AI doesn't have a "this is confidential" button. You're the button.

In the United States, a doctor who pastes his patient's data into ChatGPT is committing a technical violation of HIPAA, the health data protection law, because OpenAI's servers do not meet the required security standards. It doesn't matter if you delete the name. It doesn't matter if you "anonymize" in your own way. Once the data leaves the hospital system, it is no longer protected.²

Maya: Wait. I do that. Not with a name and surname. But I paste evolution notes so that AI can help me organize the plan. I give it the age, the diagnosis, the laboratories... sometimes the social context.





Dr. Flores: With three pieces of information, age, diagnosis, zip code, a machine learning model can identify an individual. The World Medical Association warned in 2024: anonymization alone does not protect patient data.

Maya: No one told us that.

Dr. Flores: In Europe, the GDPR classifies health data as a special category, maximum protection. In Latin America, Peru was the first country to pass a general AI law, Law 31814, which classifies health and education as high-risk domains. But implementation is still incipient. Only 15% of the world's countries have binding legislation specific to AI. The rest operate with good intentions and regulations that were written before generative AI existed.³

Kai: And in the meantime, we use the tool every day with real data.

Dr. Flores: An American Medical Association survey in 2024 found that two out of three physicians were already using AI tools in their practice, nearly double the number from the previous year. The tool is already in the clinic. The question is not whether they are going to use it. It's whether they're going to know what they're risking when they use it.⁴

Maya: I can admit something I've never said.

Last semester, I had a case report to deliver. I also had four rotations, an exam on the same day and zero time. I made the diagram myself: the structure, the key points, the differential diagnosis. But the wording... I asked AI for it. I gave it my outline and said, "Write it in clinical language, case report format, 1,500 words." I read it, changed a few sentences and handed it in.

Kai: And you got a good grade?

Maya: An excellent grade, the attendees said it was "one of the best reports of the group."

Kai: So, what's the problem?

Maya: I don't know if the report was mine.

Kai: But the structure was yours. The diagnosis was yours. The ideas were yours.

Maya: Not the execution.

Kai: But if you use a proofreader, no one says anything. If you ask a resident to check your wording, no one says anything. What's the difference?

Maya: I don't know, and that's the problem; I don't know where the line is.

Dr. Flores: You don't know because the line moved. And most institutions still don't know where to put it. In 2025, a systematic review found that most institutions still do not have clear policies. Teachers use AI tools, but fewer than one in four are aware of their own institution's formal policy. Many policies encourage the use of generative AI, but few give detailed guidance on how to do it without crossing lines.⁵

Kai: But that seems to me to be a problem of the institutions, not of the

students. If the university doesn't tell me what I can and can't do with AI, and then punishes me for using it... the problem is not my ethics. It's their negligence.

Dr. Flores: There is some truth in what you say. But that doesn't mean that anything goes. Let's go back to your example, Maya. You did the analysis. You gave AI a solid outline. You used AI as an editor, not as an author. That's unlike a student who opens an AI tool with zero preparation and says, "Write me a pneumonia case report." The result looks the same. The process is completely different.

Kai: But the professor can't see the process; they just see the result.

Dr. Flores: And therein lies the real dilemma. The ICMJE, the committee that defines authorship rules for all medical journals, updated its recommendations in 2024: AI cannot be an author. The human is responsible for all content, including what AI generated. You have to be able to check, correct and defend every line. COPE, the publication ethics committee, says the same thing: total transparency in disclosure.⁶

Maya: So, the rule is not "don't use AI." It's "if you use it, make it visible and hold yourself accountable."

Dr. Flores: That's the rule. But it requires that no regulation can legislate honesty. And honesty is harder when the incentive is against it, when the perfect report gives you the best grade and no one asks how you did it.

Kai: That's no longer an AI problem. It's a problem of the user.

Dr. Flores: Exactly.

I'm going to present you with a finding that should concern any doctor. In 2024, Zack and colleagues published a study in *The Lancet Digital Health* that evaluated how generative AI



models responded to identical clinical vignettes, same disease, same presentation, when only the patient's race was modified. Clinical recommendations changed. Not because of the health condition, but because of the demographic data.⁷

Maya: Did the recommendation change with the same disease?

Dr. Flores: With the same disease, the same clinical presentation, the same labs. Only the demographic data changed. And this is not new. Back in 2019, Obermeyer and colleagues published a landmark study in *Science*: an algorithm used in millions of patients in the United States systematically allocated fewer resources to black patients than to white patients with the same severity of disease. Not because the algorithm used race as a variable, but because it used cost of care as a proxy for need. Black patients had historically received less care and therefore generated fewer costs.⁸

Kai: But AI doesn't "decide" to discriminate, it replicates what is in the data.

Dr. Flores: That. And therein lies the problem. AI is not prejudiced; it has data. And if the data with which it was trained reflects decades of inequity in access, in research, in representation, the model will reproduce that inequity with mathematical precision. Not out of malice. By design.

Maya: And if the doctor doesn't know that the model has that bias...

Dr. Flores: It becomes the vehicle. Unintentional. Not conscious. But real. The WHO put it as one of its six core ethical principles for AI in health: ensuring inclusion and equity. The principle exists because the risk is massive.⁹

Maya: So, using AI without knowing that is like prescribing a drug without reading the side effects.

Dr. Flores: It's worse. Because the drug comes with a package insert. An AI model does not.

Now, think about something more complex. Consider a hypothetical scenario: a diagnostic AI system that is more accurate than the overall human

doctor, say, 92% versus 85%. But when you evaluate it by subgroups, you find that for ethnic minority patients the accuracy drops to 71%. The dilemma: do you implement the system because it benefits the majority? Or do you reject it because it harms the most vulnerable?

Maya: Beneficence against justice.

Dr. Flores: The four principles of Beauchamp and Childress, autonomy, beneficence, nonmaleficence, and justice, remain the dominant framework for AI ethics in health. But what the principles don't tell you is what to do when two principles collide.¹⁰

Kai: And that happens all the time.

Dr. Flores: All the time. And the answer is not an algorithm. It's clinical judgment informed by ethics. That is why this chapter isn't a checklist of dos and don'ts.

Next scenario. You are in consultation. You used AI to check a diagnostic differential before walking in. It helped you organize your ideas, suggested a diagnosis you hadn't considered. The patient asks: "Doctor, how did you arrive at that diagnosis?"

Maya: I'm telling the truth. I tell them that I consulted an AI tool as part of my process.

Kai: I don't.

Not because I want to hide it. But because saying, "I used AI" can scare the patient for no reason. If I checked the differential, if I agree with the result, if I take responsibility... telling the patient that "the computer helped me think" doesn't add anything. It takes away their confidence.

Maya: But they have a right to know.

Kai: They have the right to a good diagnosis. If I tell them that I used ClinicalKey Student to verify, no one is scared. If I tell them I used an AI



tool, they will freak out. It's the same process. It changes perception.

Dr. Flores: Both have valid arguments. And the evidence shows that it is more complex than it seems. Patients' preferences for whether they want to know that AI was used vary by gender, age and income. It's not universal.¹¹

But the legislation is moving forward. California passed a law in 2025, the first by a U.S. state, that requires health providers to include a notice when communications with the patient are generated by AI. With one exception: if a licensed physician reviewed and approved the content.¹²

Kai: But in many countries, there's nothing like that.

Dr. Flores: Which means that in most contexts, the transparency decision is not going to be made by law. You are going to take it.

Resident perspective blog: transparency dilemmas

How to communicate AI-supported conclusions to patients, peers and supervisors by Aisha Quayum
[Read the article →](#)

Maya: I will admit something that makes me embarrassed...

I've been thinking about something for weeks that I can't get out of my head. If AI can make better differentials than me. If it can write better than me. If it can organize information faster than I can. If it can detect patterns that I don't see... what's left of me?

Kai: The patient.

AI doesn't sit down with the patient at three in the morning when they're afraid. AI doesn't know that the patient in Room 412 won't take the medicine. If you don't calmly explain to her why she needs it. AI does not detect that



the patient is minimizing his symptoms because he is embarrassed to talk about mental health in front of his wife.

I don't have your system or your discipline, Maya. But what I do know is that medicine is not information. It is the person in front of me. You can't automate that.

Dr. Flores: The value of a doctor was never to process information faster than a machine. It was always the other: clinical judgment in context, the therapeutic relationship, the ability to read what the patient does not say. The responsibility of deciding when the tool is wrong. AI amplifies your ability. But the ability it amplifies, or erodes, depends on how present you are.¹³

Maya: It's not like AI is replacing me. It's just that if I stop being present, there's nothing to replace.

Dr. Flores: That's the real threat. Not automation. The abdication.

Kai: And that threat is not from AI. It's ours.

Dr. Flores: I'm not going to give you a checklist. If one thing was clear in this conversation, it's that easy answers don't exist. But principles do.

What we did today was to go through five dilemmas. Each had something in common: the tool didn't make the decision. The person does. And in each case, what separated responsible from irresponsible use was not technology. It was the judgment of the one who used it.

Kai: Privacy. If you paste a patient's data into a prompt, that's your decision. Not AI's decision.

Dr. Flores: First principle: protecting patient information is your responsibility, not the tool's responsibility.



Maya: Integrity. If you're using AI for a job, transparency isn't optional. And the responsibility for the content is yours, always.

Dr. Flores: Second principle: what bears your name is your responsibility. If AI generated it, you verify it, you defend it, you declare it.

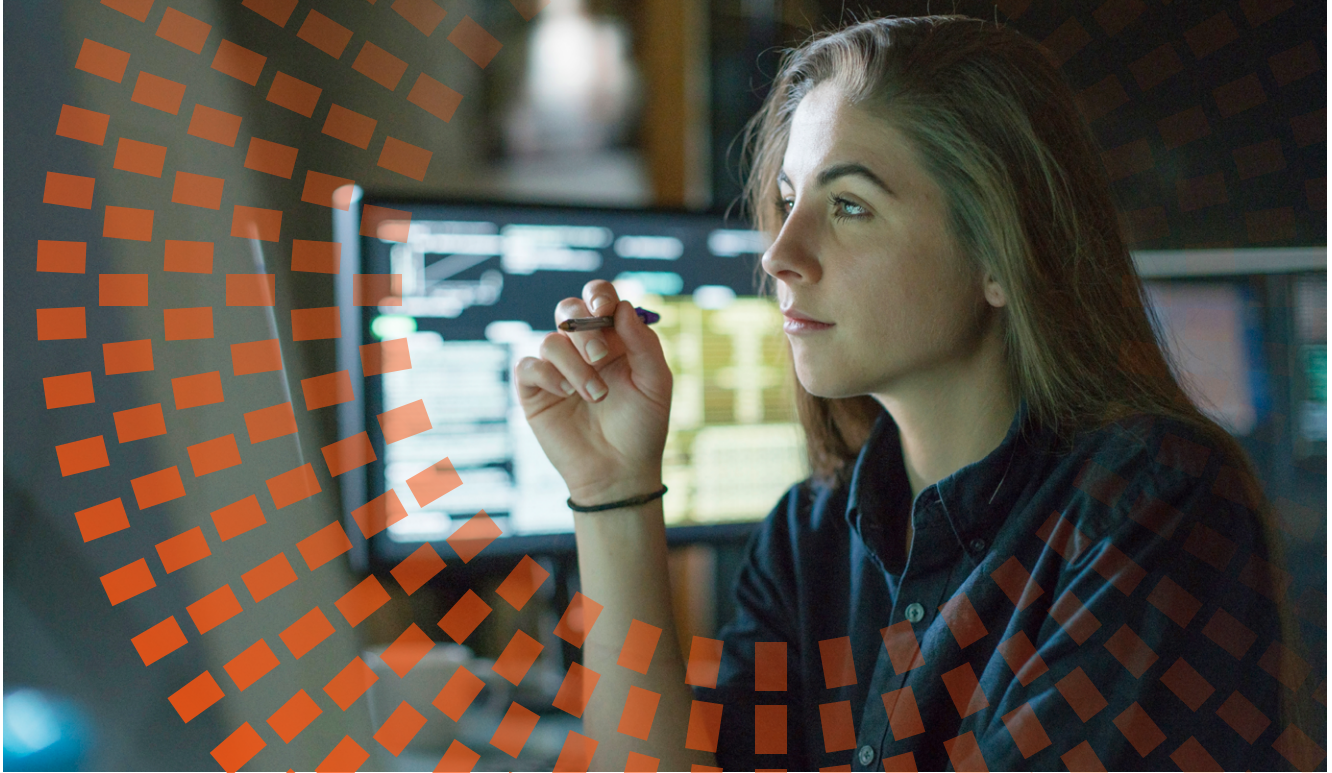
Kai: The bias. You can't assume that the tool is fair just because it gives a result that looks solid.

Dr. Flores: Third principle: AI is not neutral. It reproduces the data with which it was trained, including its inequities. Your job is to ask: who does this work for, and who doesn't it work for?

Maya: Transparency. It's not confessing that you used a tool. It is to ensure that the patient understands how the decision was made.

Dr. Flores: Fourth principle: transparency is not mechanical disclosure. It's honest communication about the clinical reasoning process.

Kai: And the fifth... that the value of the doctor is not competing with the machine. It's what the machine can't do.



 **Dr. Flores:** Fifth principle: your professional identity is not defined by what you know. It is defined by how you decide when you are uncertain. AI can give you information. But the judgment is your responsibility. The patient is yours.¹⁴

I did not invent these five principles. These are practical versions of what the WHO, the World Medical Association, the ICMJE, the AAMC, UNESCO and dozens of other institutions have been building for years. Everything points to the same thing.¹⁵

 **Kai:** Ethics is not a chapter. It's how you operate.

 **Dr. Flores:** Every day. At each prompt. In every clinical decision. There is no time when you “activate” ethics and another when you don't. It's how you use the AI tool when no one is watching.

 **Maya:** This was different.

 **Dr. Flores:** It should be. The previous chapters gave you abilities. Gave you limits. And limits aren't what holds you back. They're what keeps you going.



Key takeaways from this chapter

- **Principle 1:** protecting patient information is your responsibility, not the tool's.
- **Principle 2:** what bears your name is your responsibility. If AI generated it, you verify it, defend it, and declare it.
- **Principle 3:** AI is not neutral. It reproduces the data with which it was trained, including its inequities.
- **Principle 4:** transparency is not mechanical disclosure. It is honest communication about the clinical reasoning process.
- **Principle 5:** your professional identity is defined by how you decide when you are uncertain. The judgment is yours. The responsibility is yours. The patient is yours.
- The real threat is not automation. It is the abdication.

References

1. Khan, M. M., et al. (2025). Towards secure and trusted AI in healthcare: A systematic review of emerging innovations and ethical challenges. *International Journal of Medical Informatics*, 195, 105759. [ELSEVIER] [NOTE: Original outline cited as “Akinyemi et al.” — actual first author is Khan. Data on staff entering patient data into public AI confirmed.] + HHS Office for Civil Rights breach portal data: 720 healthcare data breaches in 2024, ~186M records affected.
2. USC Price School of Public Policy. (2025). Physicians using ChatGPT unknowingly violate HIPAA. + Li, J. (2023). Security implications of AI chatbots in health care. *Journal of Medical Internet Research*, 25, e47551. <https://doi.org/10.2196/47551>
3. Guo, Y., et al. (2025). AI regulation in healthcare around the world: What is the status quo? Only 15.2% of countries have legally binding AI-specific legislation. *medRxiv*. <https://doi.org/10.1101/2025.01.25.25321061> + Peru, Ley 31814 (2023) + Decreto Supremo 115-2025. + Busch, F., Kather, J. N., Johnner, C., et al. (2024). Navigating the European Union Artificial Intelligence Act for healthcare. *npj Digital Medicine*, 7, 210. <https://doi.org/10.1038/s41746-024-01213-6>
4. American Medical Association. (2024). *Physicians’ Sentiments About AI in Health Care*. AMA Digital Health Research. [~66% physician AI adoption in 2024, up from ~38% in 2023]
5. Perkins, M., et al. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4), 296. + Sabzalieva, E., & Valentini, A. (2023). *ChatGPT and Artificial Intelligence in Higher Education: Quick Start Guide*. UNESCO.
6. ICMJE. (2024). *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals* (updated). AI cannot be listed as author; disclosure required. + COPE. (2023). *Authorship and AI Tools*. AI cannot be credited as author; transparency mandatory. + Currie, G. M. (2023). Academic integrity and AI: ChatGPT fabricates scientific abstracts avoiding plagiarism detection. *Seminars in Nuclear Medicine*, 53(3), 468–474. <https://doi.org/10.1053/j.semnuclmed.2023.04.008>
7. Zack, T., Lehman, E., Suzgun, M., et al. (2024). Assessing the potential of GPT-4 to perpetuate racial and gender biases in health care: A model evaluation study. *The Lancet Digital Health*, 6(1), e12–e22. [https://doi.org/10.1016/S2589-7500\(23\)00225-X](https://doi.org/10.1016/S2589-7500(23)00225-X)
8. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. <https://doi.org/10.1126/science.aax2342>
9. WHO. (2021). *Ethics and Governance of Artificial Intelligence for Health*. 6 core ethical principles: autonomy, well-being/safety, transparency, accountability, equity, sustainability. + WHO. (2024). *Guidance on Large Multi-Modal Models*. 40+ recommendations.
10. Gorelik, A. J., Li, M., Hahne, J., et al. (2025). Ethics of AI in healthcare: A scoping review demonstrating applicability of a foundational framework. *Frontiers in Digital Health*, 7, 1662642. <https://doi.org/10.3389/fdgth.2025.1662642> + Karimian, G., Petelos, E., & Evers, S. M. A. A. (2025). Ethical and social considerations of applying artificial intelligence in healthcare: A two-pronged scoping review. *BMC Medical Ethics*, 26, 95. <https://doi.org/10.1186/s12910-025-01198-1> + Civaner, M. M., et al. (2024). Global cross-sectional student survey on AI in medical, dental, and veterinary education and practice at 192 faculties. *BMC Medical Education*, 24, 1038. <https://doi.org/10.1186/s12909-024-06035-4> [Source of “>75% no formal AI education” — n=4,596 students, 48 countries, 192 faculties. Note: data refers to AI education broadly, not AI ethics specifically.]
11. Park, H. J. (2024). Patient perspectives on informed consent for medical AI: A web-based experiment. *DIGITAL HEALTH*, 10. <https://doi.org/10.1177/20552076241247938> + Druce, K. L., et al. (2025). Informed consent for ambient AI documentation: Patient comfort varies. *JAMA Network Open*.
12. California AB 3030. (2025). First US state law mandating AI-generated content disclaimers in patient communications. + EU AI Act (2024). Transparency obligations for high-risk AI systems; healthcare classified as high-risk by default.
13. Al-Worafi, Y. M., et al. (2025). Generative AI in medical education: AI threatens foundational professional values — integrity, accountability, compassion. *Cureus*. + *The Lancet Digital Health*. (2025). AI requires integration of critical thinking, ethical competence, and professional identity formation. [ELSEVIER]
14. Bhatt, S., et al. (2025). “Could I, Would I, Should I”: A framework for ethical AI use in medical education. *Academic Medicine*.
15. WHO (2021) + UNESCO (2021) + AAMC (2025) + ICMJE (2024) + WMA (2024) + Masters, K. (2023). AMEE Guide No. 158: Ethical use of AI in health professions education. *Medical Teacher*, 45(6), 574–584. <https://doi.org/10.1080/0142159X.2023.2186203>

Chapter 8: Quick-start checklist



In this chapter, we will cover:

- A practical checklist that condenses chapters 4-8 into a cycle of four phases: plan, learn, evaluate, adjust
- Fourteen ready-to-use prompts, designed to activate your thinking instead of replacing it
- The central rule: you first, AI later. In each phase, your effort comes before that of the tool.



Dr. Flores: Everything we talked about in the previous chapters, the limitations, the critical thinking, the clinical reasoning, the self-regulation, can be condensed into something that you can use tomorrow. A cycle is learned by doing it, not by reading about it.



Kai: Finally.



Dr. Flores: A central rule: you first, AI later. At each stage of the cycle, your effort comes before that of the tool. If you reverse the order, the checklist becomes another form of delegation. If you stick to it, it's a method.



Maya: The checklist is a learning system, not a productivity system.



Dr. Flores: The prompts you are going to see are examples. Adapt to your topic, level and context, be it medicine, nursing or any other health profession. The perfect prompt does not exist. The prompt that activates your thought, yes. At the end of the chapter, Hussain Hilali, a resident who used this method during his training, presents you with his own version.

Plan

Checklist

What should I do BEFORE opening AI to study a topic?

- **Write down what you already know.** Spend five minutes recording everything you remember about the topic. Messy is fine. The act of retrieval activates memory and exposes real holes, not imagined ones.
- **Identify your gaps honestly.** After writing down what you know, mark what you DON'T know. Not what you didn't study, but what you tried to remember and couldn't. That's the difference between an exposure gap and an understanding gap.
- **Define your study goal in one sentence.** "I want to understand the mechanism of ACE inhibitors" is different from, "I want to explain why enalapril is preferred over captopril in heart failure." For nursing: "I want to understand why untreated hypertension damages the kidneys and how I monitor that in my patient." Specificity enhances the usefulness of AI as a tool.
- **Ask yourself the three security questions** before relying on any AI output during the session:
 - ☐ What is this information NOT telling me?
 - ☐ Where does this data come from, and where does it NOT come from?
 - ☐ What if this was wrong?

Prompts

Prompt 1 — gap diagnosis: I'm about to study [topic]. Before you start, ask me 10 diagnostic questions of varying levels of difficulty to identify what I know and what I don't. Don't give me the answers until I answer each one.

Prompt 2 — priority map: I have a [topic/module] exam. Here's what I think I know well: [list]. Here's what I think I'm missing: [list]. What additional areas should I consider? Organize by clinical impact, not by frequency of exams.

Prompt 3 — pre-study questions (metacognition): Before you help me with [topic], I need to reflect. Ask me three questions about what I understand, what confuses me and what I assume I know, but could be wrong.





Learn

Checklist

How do I study WITH AI, without it replacing my processing?

- ☐ **Generate your own summary first.** Read the material. Write your summary, incomplete, imperfect. THEN ask AI to generate one on the same topic. The gaps between the two are your real study map.
- ☐ **Build your flashcards before you order them.** Write at least five cards about the concepts that are most important to you. Then generate with AI. The cards that AI produces, and you didn't think of, are the ones you need to study the most.
- ☐ **Use AI as a newbie, not an oracle.** Instead of asking AI to explain, do the explaining yourself. Prompt the tool to ask you clarifying questions when your explanation is incomplete. This is the cognitive mirror: where your gaps become visible.
- ☐ **Apply SIFT to every output you use to study:**
 - **Stop.** Don't copy or act right away.
 - **Investigate the claims.** Verify every critical piece of data against an independent source.
 - **Find a better reference.** Is there another source that says something different?
 - **Trace the logic.** How did AI get from A to B? Is the step valid or only plausible?
- ☐ **For clinical illness scripts: build your own before consulting AI.** Don't ask, "Tell me everything about [illness]." Ask to be evaluated after building yours.
- ☐ **For nursing: build your care plan before consulting AI.** Write your assessment, identify priority nursing diagnoses, plan interventions. Then ask AI what you're missing. Same principle: you first.

Prompts

Prompt 4 — cognitive mirror (teaching to learn): You're a first-year student. I'm going to explain [topic]. Ask me clarifying questions when my explanation is incomplete, ambiguous or incorrect. Don't correct me directly, just ask.

Prompt 5 — abstract comparison: Here's my summary of [topic]: [your summary]. Generate your own summary of the same topic. Then point out: 1) what I included that you did not; 2) what you included that I did not; and 3) where our explanations differ.

Prompt 6 — illness script assessment: Here's my [disease] illness script: [your version with pathophysiology, epidemiology, specific findings, differential diagnoses, time course]. Evaluate if I'm missing anything critical in each component. Then ask me three questions to verify that I really understood the relationships between them.

Prompt 6b — plan of care evaluation (nursing): Here's my plan of care for [type of patient]: [your assessment, nursing diagnoses, planned interventions]. Assess whether I am missing any critical priorities, whether my interventions align with diagnoses, and whether there are any safety risks that I did not consider. Then ask me three questions about my clinical reasoning.

Prompt 7 — Socratic deepening (leveling up): I already understand the basics of [topic]. Now I want to get to the level of analysis. Don't give me new information. Ask me questions that force me to connect what I already know with clinical situations. Socratic style: one question at a time.



Evaluate

Checklist

How do I verify that I really understood and not just recognized?

- ☐ **Try to explain without looking at anything.** Before ordering MCQs or using any tool, close everything and try to explain the topic from scratch. If you can't, you know exactly where the gap is.
- ☐ **Generate YOUR answer before you see AI's output.** When using AI for practice questions, write YOUR answer first, even if it's incorrect. The recovery effort is where learning happens.
- ☐ **Use MCQs with AI, but with caution.** In a literature review using GPT-4, 49% of the MCQs generated had construction defects according to NBME criteria, and 22% contained factual errors (Kiyak & Kononowicz 2024; n=150 items). Newer models can improve these figures, but verification is still essential.¹
- ☐ **Apply the three levels of doubt to your own study:**
 - Are the facts correct? Verify against independent sources.
 - Is the logic valid? Does the reasoning that connects facts make sense?
 - Was my question the right one? Am I evaluating what really matters?
- ☐ **For clinical cases: think first, consult later.** Read the story. Close AI. Generate your own differential. Open AI. Give it the case and your differential. Ask for it to be compared. Compare: what did you lack? Why didn't you consider it?



Prompts

Prompt 8 — adaptive MCQs: Generate 10 multiple-choice questions on [topic] at the [basic/intermediate/advanced] level. Format: one question at a time. Do not show me the answer or explanation until I answer. After my 10 answers, give me a pattern analysis of my mistakes.

Prompt 9 — reasoning evaluation (not just facts): I'm going to give you my explanation of [topic/mechanism/case]. Don't tell me if it's right or wrong. Ask me about logic: what assumptions am I making? Which step of my reasoning is the weakest? What alternative did I not consider?

Prompt 10 — differential with verification (think first): I already have my preliminary differential for this case: [your list]. What important diagnoses might I be missing given this data? Explain why each addition is relevant. Don't rearrange my list, just supplement it.

Prompt 11 — red flags of understanding: I just finished studying [topic]. Ask me five questions that I can only answer well if I really understand the mechanism, not if I just memorize data. If my answer is superficial, point that out.

Adjust

Checklist

How do I change my method for the next cycle based on what I discovered?

- ☐ **Analyze patterns in your mistakes, not only what you failed, but why.** After a practice session, spend 10 minutes identifying: did I fail because I didn't know the fact? Weak reasoning? Wrong assumption? The type of error determines the correction strategy.
- ☐ **Redesign your next cycle with data, not intuition.** Don't repeat the same strategy expecting different results. If you failed in clinical application, don't review flashcards: practice cases. If you failed in basic data: go back to the material.
- ☐ **Recognize signs of algorithmic fatigue:**
 - You rephrased the prompt more than three times without getting what you're looking for. You probably don't know what you're looking for.
 - You accept an output without reading it in its entirety because it looks good. Active automation bias.
 - You're more tired after the interaction than before. Time to close the tool.
- ☐ **Close the loop:** write a sentence about what you're going to change. Next time, I'm going to [concrete action] because I discovered [concrete finding]. If you can't write that sentence, you didn't close the loop.

Prompts

Prompt 12 — error pattern analysis: These are the questions I missed: [list with your incorrect answers]. Identify patterns: are they errors of knowledge (I didn't know the data), reasoning errors (I knew the data, but I connected it wrong), or assumptions (my approach was wrong)? Suggest what I should check first.

Prompt 13 — curriculum redesign: I just finished a course of study on [topic]. My results: [honest summary, what went well, what didn't]. My method was: [describe what you did]. What should I adjust for the next cycle? Give me a maximum of three concrete, non-generic changes.

Prompt 14 — metacognitive closing reflection: Review my study session today on [topic]. I'm going to describe what I did and how it went: [description]. Evaluate: 1) whether my method triggered deep understanding or just recognition; 2) when I used AI as an extension versus a crutch; and 3) what specific change I should try in the next session.





From the resident perspective: Hussain Hilali

Resident perspective blog: safe use practices

Using AI in medical school without letting it use you by Hussain Hilali

[Read the article →](#)

Hussain's quick checklist:

1. Before I open AI: what do I already know? What confuses me? (Two minutes maximum)
2. First attempt always mine: summary, flashcards, differential. Imperfect is fine.
3. AI second. Compare, complement, challenge. Never let it replace your thinking.
4. After AI: can I explain this without looking at the screen? If not, I didn't finish.
5. Closing the loop: one thing I'll do differently next time.

Hussain's prompting tips:

- Be specific about your level: "I'm a junior studying for [rotation]" changes the output.
- Ask for questions, not answers: evaluate me > explain it to me.
- Ask for one thing at a time. Multi-part prompts dilute responses.
- Always end with, "What did I miss?" That forces AI to complement, not repeat.

 **Maya:** It's basically my method but without the three hours of optimization.

 **Kai:** That's the point. If it was complicated, we wouldn't do it.

 **Maya:** You speak like someone who has never completed a checklist.

 **Kai:** The checklist is not the change. The intention is.

 **Dr. Flores:** The checklist you have is a starting point, not a destination. The prompts are going to change. The tools are going to change. The models are going to improve and have new limitations. What is not going to change is the question that all this taught you to ask yourself.

 **Kai:** Am I thinking or am I delegating?

 **Dr. Flores:** That. And you don't need AI to answer it. You know it every time you open the tool and decide whether to think sooner or later. That decision, repeated a hundred times, a thousand times, every day of your training, is the difference between the professional who uses AI and the professional who depends on it.

The next chapter changes territory. All we talked about was how to use AI to learn and reason. Now comes the question that many avoid: where are the boundaries that you should not cross?

 **Maya:** Ethics.

 **Dr. Flores:** Ethics. Security. Professionalism. Where the consequences are not a failed exam but the confidence of a patient.



Key ideas from this chapter

- The checklist is a learning system, not a productivity system.
- Prompts will change. The tools are going to change. What does not change is the question: am I thinking or am I delegating?
- That decision, repeated a hundred times every day of your training, is the difference between the professional who uses AI and the professional who depends on it.

References

This chapter synthesizes existing evidence from chapters 4-7. It does not introduce new evidence. The main references are cross-references:

1. Kiyak, Y. S., & Kononowicz, A. A. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: A literature review. *Postgraduate Medical Journal*, 100(1189), 858–865. <https://doi.org/10.1093/postmj/qgae064> [GPT-4, n=150 items, NBME criteria. 49% construction defects, 22% factual errors. Newer models may improve figures; verification remains essential.]

Cross-references by section:

- **Planning:** Ch. 7 Beats 1-2 (MAL Planning) + Ch. 4 Beat 6 (three security questions) + Ch. 5 Beat 2 (metacognitive reflection) + Cutrer et al. 2020, *The Master Adaptive Learner*, Elsevier
- **Learn:** Ch. 7 Beat 2 (you first in Learning) + Ch. 5 Beat 3 (cognitive mirror) + Ch. 6 Beat 2 (illness scripts) + Ch. 5 Beat 4 (adapted SIFT)
- **Assess:** Ch. 7 Beat 2 (you first in Assessing) + Ch. 5 Beat 2 (three levels of doubt) + Ch. 7 Beat 3 (desirable difficulties) + Ch. 6 Beat 2 (think first)
- **Adjust:** Ch. 7 Beat 2 (you first in Adjusting) + Ch. 7 Beat 5 (agency close) + Ch. 5 Beat 4 (personal red flags)

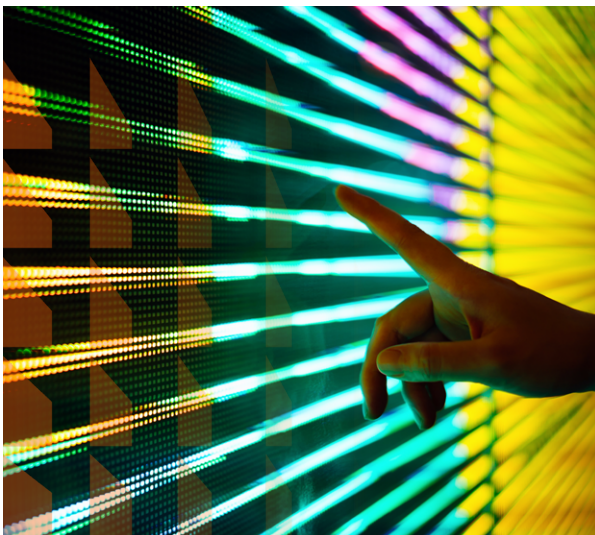
Prompts reported by: Huang 2024 (20 ChatGPT activities mapped to SRL) · Jackson 2025 (Higher Order Prompting) · Park & Choo 2025 (Socratic prompting) · Safranek et al. 2024 (ChatGPT Implementation Guide in Medical Education) · Chan & Zary 2025 (custom mnemonics with AI)

Chapter 9: FAQs, myths and facts



In this chapter, we will cover:

- Fourteen frequently asked questions organized into three blocks: practical use, learning and training, ethics and future
- Answers anchored in the evidence of the previous chapters, not in opinion
- Seven common myths about AI in health, confronted with documented facts
- Hussain Hilali's perspective: what I wish I had known as a resident



 **Dr. Flores:** We have come a long way. Limitations, how to think critically, clinical reasoning, self-regulation, checklist, ethics. Now I want you to ask me what you still have left.

 **Kai:** Open office type.

 **Dr. Flores:** Open office type. No filter.

 **Kai:** Obviously.

Practical use

FAQ 1: Is using AI cheating?

-  **Dr. Flores:** It depends on what you did with it. If you opened AI without preparation and told it, “Write me a case report,” the product is not yours, even if you have edited it. If you did your analysis, built your outline, and used AI to polish the writing, the thought is yours.
-  **Maya:** I went through that. The grade was excellent. But I didn't know if the report was mine.
-  **Dr. Flores:** The rule is: if you can't explain every line as if you wrote it, it's not yours. And if you used it, declare it. Most institutions still don't have clear policies: 63% encourage the use of generative AI, but only 43% give detailed guidance.

Dr. Flores' bottom-line: Honestly, I think it will depend a lot on your school's rules or regulations, so I would recommend adhering to those.

FAQ 2: Can I trust AI to diagnose?

-  **Dr. Flores:** AI looks for the probable. You look for the right thing. They are different things. In a triage study, ChatGPT Health under-triaged 52% of emergencies, including diabetic ketoacidosis.²
-  **Kai:** So, you tell it the symptoms, and it sends you home.
-  **Dr. Flores:** In more than half of the urgent cases, yes. AI can amplify false information instead of correcting it. It is not a substitute for clinical judgment. It complements it, if you check.





FAQ 3: How do I cite AI in an academic paper?

-  **Dr. Flores:** The ICMJE updated its recommendations in 2024: AI cannot be an author. The human is responsible for all content. COPE says the same thing: mandatory transparency.³
-  **Maya:** State how you used the tool. In which parts, with what tool (e.g., Claude, ChatGPT, Manus). Don't hide the process. Check everything as if you had written it yourself.

FAQ 4: What do I do if my teacher bans AI?

-  **Kai:** I would have been angry before. Now I see it differently. The teacher is forcing you to fix it. Exactly what your brain needs.
-  **Dr. Flores:** Respect institutional policy. Screening tools have high false-positive rates and disproportionately affect non-native English speakers. Tasks without AI are deliberate cognitive training, not punishment.¹

FAQ 5: What is the difference between ChatGPT and a medical AI tool?

-  **Dr. Flores:** ChatGPT is a general model trained with the internet. An AI medical tool is trained with clinical data, validated for specific contexts, and regulated as a medical device. The rule: general to learn, specific to decide. Verify in both.⁴

Learning and training

FAQ 6: Is AI going to replace doctors?

-  **Kai:** AI doesn't sit down with the patient at three in the morning when they're afraid. That is not automated.
-  **Dr. Flores:** It's not going to replace doctors. It will replace doctors who do not know how to use AI. The concept is the "augmented doctor:" your value is in the judgment, the therapeutic relationship, the presence. What AI doesn't have.⁵

FAQ 7: Can AI help me study for exams?

-  **Maya:** Yes, with conditions. AI raises questions. But if you don't make the effort to respond before you see the answer, you're training recognition, not knowledge.
-  **Dr. Flores:** In a literature review, 49% of MCQs generated with GPT-4 had construction defects, and 22% contained factual errors. Usable, but always check against your primary source.⁶

FAQ 8: Can AI make me a worse doctor?

-  **Dr. Flores:** Yes. There is direct evidence. Kaminski and colleagues studied 19 experienced endoscopists: when they were withdrawn from AI assistance, their adenoma detection rate dropped by 20%. From 28.4% to 22.4%.⁷
-  **Kai:** It's like GPS. You use it so much that you don't know how to navigate. But here, it's an adenoma that could be cancer.
-  **Dr. Flores:** The concept is deskilling. And for students it's worse: if they never build the skill without AI, they never have it. Preserving cognitive friction is not nostalgic. It's clinically necessary.





FAQ 9: Can AI write my papers?

Dr. Flores: No. First, because it generates text that seems correct but contains factual errors. Generative AI can manufacture complete scientific summaries that evade plagiarism detection. Second, because scientific writing is not a product: it is the process of organizing your thinking, defending your arguments, confronting the weaknesses of your evidence.⁸

Maya: Use AI as an editor, not as an author. Your outline, your argument, your structure. AI polishes. But the thought has to be yours.

Dr. Flores: Right... co-create, with you as orchestrator

FAQ 10: What skills should I develop that AI can't replace?

Dr. Flores: The most human. Clinical examination. Deliberate diagnostic reasoning. Empathetic communication. Tolerance for uncertainty. Critical evaluation of AI outputs. Your advantage is not speed or memory. It's context: reading nonverbal cues, detecting that the patient minimizes his symptoms because he's embarrassed to talk about mental health in front of his wife.

Kai: Medicine is not information. It's the person in front of you.

Ethics and future

FAQ 11: Does AI have biases against certain populations?

 **Dr. Flores:** Yes. Not by decision, but by design. AI inherits biases from your training data with mathematical precision. One example: in dermatology, AI achieved 69% accuracy on light skin and 17% on dark skin, because the training set had only 10 dark-skinned images in 100,000.⁹

 **Maya:** Using AI without knowing this is like prescribing a drug without reading the adverse effects. Except that the drug has a package insert. The AI does not.



FAQ 12: Should I tell the patient that I used AI?

 **Dr. Flores:** The response is evolving. California passed a law in 2025 requiring notice when clinical communications are generated by AI, except if a doctor reviewed them. Europe is moving in the same direction.¹⁰

 **Maya:** I didn't "use AI." But "This is how I came to this recommendation, and this is what I considered."

 **Dr. Flores:** That's the adult version of transparency.

FAQ 13: How do I know if an AI tool is secure with patient data?

 **Dr. Flores:** If it's public and free, ChatGPT, Gemini, Claude open, DON'T give it patient data. Never. Once the data enters the tool's server, it leaves the hospital's control. With three pieces of information, age, diagnosis, zip code, an individual can be identified.¹¹

 **Maya:** AI doesn't have a "this is confidential" button. You are the button.



Myths and facts

Myth	Fact
AI is going to replace all doctors	85% of AI deployments in healthcare are augmentation, not replacement (McKinsey 2025).
AI is objective	AI inherits biases from its training data, including documented racial and gender inequities.
I need to know how to program to use AI	False. Clinical literacy in AI is not technical literacy. You need to know how to evaluate outputs, not write code.
Hallucinations are rare	In one study, 1.47% hallucination rate; in another, > 25% of manufactured references. Rare depends on the context: in the clinic, 1% is too much.
Patients can use chatbots for medical advice	Patients lack clinical context to distinguish a correct answer from a plausible one.
All AI in health is the same	Different types of training, validation and regulation. An image model is not a chatbot, it is not a prediction system.

Closing

-  **Maya:** It's not going to replace me. It is not objective. I don't need to know how to program. You're not going to predict the future. It is not reliable without supervision.
-  **Kai:** But it's not a toy or a fad.
-  **Maya:** It's a tool that amplifies strengths and biases.
-  **Kai:** The most dangerous gap is not technological.
-  **Maya:** It's epistemological. Believing that you understand AI when it's the only thing you use. Or fear AI when you don't know anything about it.
-  **Kai:** AI proficiency is no longer optional.
-  **Maya:** It's becoming as fundamental as competence in evidence-based medicine.
-  **Dr. Flores:** I have nothing to add. Except this: let's get started.

We start with two questions. Maya: "Am I learning or am I letting the machine organize for me?" Kai: "If AI does it right, why me?" I did not give you closed answers. I gave you tools to answer yourselves. Every day. Every prompt. Every clinical decision. All four questions are not answered once. They are always answered.

-  **Kai:** The questions don't change. We do.



Key ideas from this chapter

- Using AI is not cheating. But if you're not able to explain every line like it's yours: yes, it is.
- AI is not going to replace doctors. It will replace doctors who do not know how to use it.
- The most dangerous gap is not technological. It's epistemological: believing that you understand AI when it is the only source you use.
- AI competence is already as fundamental as evidence-based medicine competence.

References

1. Perkins, M., et al. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information*, 16(4), 296. <https://doi.org/10.3390/info16040296>
2. Ramaswamy, A., et al. (2026). ChatGPT Health triage: 52% undertriage rate including DKA. *Nature Medicine*. [Cross-reference: Ch. 4]
3. ICMJE. (2024). *Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals*. + COPE. (2023). Authorship and AI Tools.
4. Busch, F., et al. (2024). Navigating the European Union Artificial Intelligence Act for healthcare. *npj Digital Medicine*, 7, 210. <https://doi.org/10.1038/s41746-024-01213-6>
5. Bouabida, K., et al. (2026). “The future of medicine will be profoundly human, precisely because intelligently augmented.” [Cross-reference: Ch. 1 Beat 7, Ch. 8]
6. Kiyak, Y. S., & Kononowicz, A. A. (2024). ChatGPT prompts for generating MCQs in medical education and evidence on their validity. *Postgraduate Medical Journal*, 100(1189), 858–865. <https://doi.org/10.1093/postmj/qgae064>
7. Kaminski, M. F., et al. (2025). Colonoscopy adenoma detection rate drops from 28.4% to 22.4% after AI withdrawal. *The Lancet Gastroenterology & Hepatology*. [ELSEVIER] [Cross-reference: Ch. 4]
8. Currie, G. M. (2023). Academic integrity and AI. *Seminars in Nuclear Medicine*, 53(3), 468–474. <https://doi.org/10.1053/j.semnuclmed.2023.04.008>
9. Daneshjou, R., et al. (2022). Disparities in dermatology AI. *Science Advances*. + Zack, T., et al. (2024). GPT-4 racial/gender biases in healthcare. *The Lancet Digital Health*, 6(1), e12–e22. [https://doi.org/10.1016/S2589-7500\(23\)00225-X](https://doi.org/10.1016/S2589-7500(23)00225-X)
10. California AB 3030. (2025). + EU AI Act (2024).
11. Li, J. (2023). Security implications of AI chatbots in health care. *Journal of Medical Internet Research*, 25, e47551. <https://doi.org/10.2196/47551> + USC Price School (2025).