Annals of Pharmacy Practice and Pharmacotherapy
ISSN: 3062-4436
2024, Volume 4, Issue 2, Page No: 82-91

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

When Clinical Reasoning Is Outsourced to Generative Artificial Intelligence: Preserving Verification, Causal Pharmacology, and Professional Accountability in Pharmacy Education

Michael Tan[1]*, Farah Aziz[2], Johan de Vries[3]

[1]Department of Pharmacy Education and Clinical Reasoning, Faculty of Science, National University of Singapore, Singapore.
[2]Department of Pharmacy Practice and Artificial Intelligence, Faculty of Pharmacy, Universiti Putra Malaysia, Serdang, Malaysia.
[3]Department of Clinical Pharmacy and Professional Ethics, Faculty of Pharmacy, Erasmus University Rotterdam, Rotterdam, Netherlands.

***E-mail** ✉ michael.tan@nus.edu.sg

Received: 02 July 2024; **Revised:** 29 September 2024; **Accepted:** 01 October 2024

ABSTRACT

Generative artificial intelligence can now produce fluent drug-information responses, therapeutic recommendations, explanatory prose, and answers to professional examination questions. This capability creates a specific problem for pharmacy education: a correct or persuasive response no longer establishes that the learner independently retrieved appropriate evidence, verified its authenticity and relevance, understood the pharmacological mechanism, integrated patient-specific information, recognized uncertainty, or assumed responsibility for the recommendation. The educational question is therefore not whether generative artificial intelligence should be prohibited, but which components of clinical reasoning must remain demonstrably attributable to the learner when machine assistance is available. Current evidence supports a differentiated response. Large language models can perform strongly on selected knowledge benchmarks and can assist with medication information and educational tasks, yet their performance varies across task type, clinical complexity, prompt formulation, source generation, and patient-specific decision requirements. These limitations become particularly important when pharmacy decisions depend on causal pharmacology rather than factual retrieval alone. Verification should consequently be treated as observable intellectual work rather than an instruction appended to an artificial-intelligence-generated answer. Pharmacy assessment should preserve opportunities to determine whether learners can reconstruct the evidence-to-decision pathway, explain pharmacological causation, respond to clinically meaningful case perturbations, identify unsupported sources, and state what remains uncertain. Generative artificial intelligence may appropriately augment pharmacy education, but professional accountability requires that consequential reasoning remain inspectable and attributable.

Keywords: Generative artificial intelligence, Pharmacy education, Clinical reasoning, Drug information, Pharmacotherapy, Assessment

How to Cite This Article: Tan M, Aziz F, de Vries J. When Clinical Reasoning Is Outsourced to Generative Artificial Intelligence: Preserving Verification, Causal Pharmacology, and Professional Accountability in Pharmacy Education. Ann Pharm Pract Pharmacother. 2024;4(2):82-91. https://doi.org/10.51847/jQAfOrib1A

Introduction

Generative artificial intelligence (AI) has entered pharmacy education at a point when its educational evidence base remains comparatively immature. A pharmacy-focused scoping review identified only a small number of studies evaluating AI applications in pharmacy education and emphasized persistent limitations in outcome evidence, generalizability, and faculty preparedness[1]. Since then, increasingly capable large language models (LLMs) have made the practical question more urgent. Students can obtain plausible medication explanations, treatment suggestions, dose calculations, examination answers, and supporting rationales within seconds. The

resulting educational challenge is not simply unauthorized access to information. It is loss of visibility into which intellectual operations were actually performed by the learner.

That distinction matters because *clinical reasoning* is not synonymous with producing a correct answer. Health-professions education uses a wide vocabulary for reasoning processes, purposes, skills, and outcomes, and those terms cannot be treated as interchangeable without obscuring what is being assessed[2]. Correspondingly, assessment scholarship shows that no single method captures every relevant component of clinical reasoning; different formats reveal different aspects of information gathering, problem representation, hypothesis development, interpretation, justification, and decision making[3]. Pharmacy education adds a professional dimension. Entrustable professional activities for new pharmacy graduates map competence to consequential tasks such as collecting patient information, assessing medication-related problems, formulating plans, implementing interventions, and monitoring outcomes[4]. These activities involve knowledge, but they also require judgments for which a learner is expected to provide a defensible professional account.

Generative AI complicates this relationship between answer, reasoning, and responsibility because the observable product may remain excellent even when the learner's contribution is uncertain. A submitted therapeutic plan can contain the correct drug, dose, monitoring strategy, and explanation while concealing whether the student verified the source, understood why the dose is appropriate, recognized a competing diagnosis, detected a contraindication, or could revise the recommendation if renal function, interacting medication, adherence pattern, or treatment objective changed. Conversely, use of AI does not itself demonstrate that these capacities have been displaced. A learner may use an LLM to retrieve alternatives or organize information and still independently verify and reason through every clinically consequential step.

The central educational question is therefore narrower than whether generative AI is beneficial or harmful: which elements of pharmacy clinical reasoning should remain demonstrably attributable to the learner when generative AI can produce fluent drug-information and therapeutic recommendations? Addressing that question requires separating information retrieval from evidence verification, causal pharmacology from verbal explanation, patient-specific reasoning from generic treatment matching, recognition of uncertainty from confident output, and technical assistance from professional accountability. The purpose of this Commentary is to establish those distinctions and identify the kinds of educational tasks that can reveal whether the reasoning underlying an AI-assisted answer was independently understood.

Fluent answers and the attribution problem

The attribution problem arises precisely because contemporary LLMs can produce answers that resemble the products historically used as proxies for knowledge and reasoning. ChatGPT demonstrated substantial performance on United States Medical Licensing Examination questions while generating explanations that were frequently judged coherent and informative[5]. Medical-domain language models have likewise shown that large-scale models can encode substantial clinical knowledge and perform strongly across question-answering benchmarks, even though human evaluations continue to identify limitations involving factuality, reasoning, potential harm, and bias[6]. Pharmacy assessment is not insulated from this development. In evaluations using North American Pharmacist Licensure Examination practice questions, GPT-4 achieved high overall performance, although accuracy varied by question type and was weaker for some formats such as select-all-that-apply items[7].

Such results establish that an LLM may generate a defensible endpoint. They do not establish who understood the pathway to that endpoint. This distinction becomes clearer when models are evaluated in settings that require integration of sequential or patient-specific information. In realistic clinical decision tasks derived from patient cases, LLMs that performed strongly on conventional benchmarks exhibited limitations in diagnosis, laboratory interpretation, guideline adherence, instruction following, and sensitivity to the order and quantity of available information. In a randomized trial, giving physicians access to GPT-4 did not significantly improve diagnostic-reasoning performance compared with conventional resources, despite strong performance by the LLM when evaluated separately[8, 9]. These findings do not show that AI necessarily impairs human reasoning. They demonstrate something more relevant to education: model capability, answer quality, and improvement in the human user's reasoning are empirically separable outcomes.

For pharmacy educators, this separation undermines a familiar assessment inference. When a learner independently completes a therapeutic case, the written recommendation has traditionally provided at least partial

evidence that the learner recalled or located relevant information and integrated it sufficiently to generate the response. Under unrestricted GenAI access, the same product may originate from multiple processes: independent reasoning followed by AI editing, AI-supported retrieval followed by careful learner verification, iterative human–AI co-reasoning, superficial acceptance of an LLM recommendation, or direct reproduction of generated text. The final prose alone cannot reliably distinguish these pathways.

The implication is not that take-home cases, written care plans, or AI-assisted assignments have lost educational value. Their evidentiary meaning has changed. They can still evaluate the quality of a submitted clinical product, and they may provide authentic opportunities to practice tool-assisted professional work. What they cannot automatically demonstrate is independent possession of every reasoning capability implied by that product. Assessment validity therefore depends increasingly on specifying what claim is being made. If the intended claim is that a learner can use available tools to produce a safe plan, AI-assisted performance may be directly relevant. If the intended claim is that the learner personally understands the evidence, pharmacology, uncertainty, and patient-specific logic underlying that plan, additional evidence of attribution is required.

Verification must be performed, not merely requested

Verification is often proposed as the solution to unreliable AI output: learners may use GenAI provided that they "verify the answer." That formulation is insufficient unless verification is converted into observable work. Pharmacy-specific evaluations show why. In repeated testing of ChatGPT across pharmacotherapy cases, the clinical relevance of generated responses coexisted with variation in accuracy and consistency across repeated time points[10]. In an evaluation of real-world medication consultations, many responses were judged appropriate, particularly for simpler public-facing questions, but clinically important errors and incomplete responses remained, and performance was less reliable for questions requiring professional interpretation[11]. The educational problem is therefore not solved when a learner reports having checked an answer. The assessment must reveal what was checked, against what evidence, and how discrepancies were resolved.

Verification also changes qualitatively as a task moves from factual medication information toward higher-order clinical interpretation. When ChatGPT was compared with clinical pharmacists across multiple clinical-pharmacy activities, its performance was relatively strong for drug counseling but significantly weaker for prescription review, medication education, adverse drug reaction recognition, and adverse drug reaction causality assessment[12]. Those differences matter educationally because "checking" a simple factual statement and evaluating an adverse-effect attribution are not equivalent operations. The latter requires attention to temporal sequence, competing causes, dose and exposure, dechallenge or rechallenge information where available, concomitant therapies, disease state, and the consequences of alternative interpretations.

Source verification constitutes another distinct operation. In an academic drug-information evaluation, only a minority of ChatGPT responses were considered satisfactory, and the references supplied in the responses that included citations were fabricated[13]. The specific frequency observed in one study should not be generalized to every contemporary model, but the underlying assessment lesson is durable: the visual presence of a citation is not evidence that a source exists, that it says what the generated response claims, or that it applies to the patient under discussion. A learner who submits an AI-assisted recommendation should therefore be able to retrieve the underlying evidence independently, identify the relevant portion of the source, explain why the source is sufficiently authoritative and applicable, and distinguish original evidence from model-generated synthesis.

At the same time, verification requirements should not be interpreted as evidence that every AI-supported pharmacy task is intrinsically unsafe. Structured evaluations in community-pharmacy scenarios have shown useful performance on selected medication-management activities, while still emphasizing the need for rigorous validation, contextual checking, and privacy safeguards[14]. This coexistence is important. Appropriate augmentation is possible precisely because verification can discriminate useful machine assistance from unsupported output. Educational policy that treats all AI assistance as equivalent would miss both sides of the evidence: the model may add value, but that value does not remove the learner's obligation to determine whether the output is trustworthy for the actual decision.

For assessment purposes, verification should therefore be operationalized as a sequence of learner-attributable actions: locating evidence independently of the generated prose; confirming that cited sources exist; checking correspondence between source and claim; determining whether the evidence applies to the drug, indication, population, dose, and clinical setting; identifying unresolved disagreement or missing information; and modifying

or rejecting the generated recommendation when those checks fail. This definition makes verification inspectable. It also prevents "AI use with verification" from becoming an honor-system declaration that cannot establish whether any independent evaluation occurred.

Causal pharmacology under patient-specific constraints

The distinction between retrieval and reasoning becomes most consequential when therapeutic decisions depend on causal pharmacology. An LLM may correctly state an indication, common dose, interaction, monitoring parameter, or adverse effect from patterns learned during training or information supplied in the available context. Clinical pharmacy reasoning requires something additional: an explanation of why those facts change the recommendation for this patient. Evidence comparing ChatGPT with clinical pharmacists illustrates this difference. Performance approached that of pharmacists in relatively bounded drug-counseling tasks, yet larger deficits appeared in prescription review, medication education, adverse-event recognition, and causality assessment[12]. The relevant educational boundary is therefore not "knowledge versus no knowledge." It is whether the learner can connect pharmacological mechanisms and exposure conditions to a patient-specific therapeutic consequence.

Dose individualization provides a particularly useful test of that capacity. In an educational evaluation involving tacrolimus, vancomycin, and lidocaine, LLM performance differed across drugs and clinical scenarios and changed significantly when prompt structure was altered[15]. These are narrow-therapeutic-index examples rather than a universal representation of pharmacotherapy, but they expose a general reasoning demand. A calculated dose is clinically interpretable only through the variables that generated it: pharmacokinetic assumptions, measured or expected exposure, organ function, timing of concentrations, therapeutic objective, toxicity risk, interacting treatments, and the reliability of the available patient data. If a learner cannot reconstruct those relationships, agreement with a generated dose provides weak evidence of independently understood pharmacotherapy.

This requirement also clarifies why causal pharmacology should remain distinct from source verification. A learner may verify that a dosing recommendation appears in an authoritative guideline and still fail to understand why that recommendation fits the patient. Conversely, a learner may understand the mechanism but rely on obsolete or inapplicable evidence. Competent reasoning requires both. Verification asks whether the informational foundation is authentic and applicable; causal pharmacology asks whether the proposed action follows mechanistically and clinically from that foundation.

Patient-specific reasoning then adds another layer. The learner must determine which features of the patient's state materially alter the evidence-to-decision pathway. Renal or hepatic function may change exposure; coadministered drugs may alter metabolism or pharmacodynamic risk; adherence may explain apparent treatment failure; microbiological data may narrow an antimicrobial choice; an adverse event may have competing etiologies; treatment goals may change the balance between benefit, burden, and risk. These variables cannot be handled safely as decorative additions to an otherwise generic recommendation. They determine which mechanism is operating, which evidence is relevant, and which action is defensible.

The same point helps distinguish legitimate AI augmentation from unverified outsourcing. A learner may appropriately use an LLM to generate candidate causes of treatment failure, summarize potential interactions, structure a differential explanation, or identify questions that require further investigation. What must remain demonstrably attributable is the learner's capacity to decide which possibilities are pharmacologically credible, determine which patient information discriminates among them, verify the supporting evidence, and justify the resulting treatment choice. Studies showing task-dependent performance, output variability, source problems, and prompt-sensitive dose recommendations support the need for that distinction without establishing that AI use itself degrades reasoning[10-15]. The educational objective is therefore not to preserve manual information retrieval for its own sake. It is to preserve visible evidence that the learner can reconstruct the causal chain connecting drug, exposure, mechanism, patient state, uncertainty, and accountable therapeutic action.

When correct output conceals fragile reasoning

A clinically correct answer can still rest on a fragile reasoning process. The problem becomes visible when the answer changes after an irrelevant prompt alteration, when the model responds inconsistently to the same case, or when a recommendation fails after clinically important information is presented in a different order. Realistic clinical evaluations and pharmacy-specific studies have demonstrated several forms of this sensitivity[8].

Repeated pharmacotherapy queries can produce variable outputs even when the underlying clinical problem is unchanged[10]. Therapeutic-dose recommendations can also shift with prompt formulation and drug-specific complexity[15]. These findings do not establish that an LLM lacks reasoning in a philosophical or computational sense. For education, the narrower concern is whether apparently successful performance is robust enough to justify attributing the underlying reasoning to the learner.

Automation research supplies a plausible mechanism for why such fragility could become educationally important. Automation bias has been associated with reduced independent information seeking and diminished vigilance when users rely on automated recommendations, particularly when verification is effortful or cognitively demanding[16]. Most of this evidence predates contemporary generative AI and should not be interpreted as proof that pharmacy students using LLMs will develop the same behavior. It nevertheless identifies a testable risk: assistance that makes a recommendation easier to obtain may reduce the probability that a user reconstructs the reasoning independently unless the task requires that reconstruction.

The practical distinction is between *agreement with an answer* and *ownership of the reasoning that makes the answer defensible*. A learner can agree with an AI-generated recommendation because it appears familiar, because the model presents supporting language confidently, or because the recommendation happens to be correct. None of these conditions establishes that the learner could detect a hidden error. Traceability therefore becomes relevant to educational accountability. Ethical analyses of AI in health care have repeatedly identified epistemic and traceability problems alongside questions of responsibility[17]. At the same time, broader analyses of human–AI collaboration argue that machine assistance can augment professional work when human judgment, transparency, safety, and contextual interpretation remain active[18]. Pharmacy education should preserve that possibility rather than treating manual performance as the only legitimate evidence of competence.

A useful test is whether the learner's reasoning remains coherent when the original output is no longer available as a scaffold. If the recommendation changes because a clinically irrelevant wording change alters the model response, or if the learner cannot explain why a dose, monitoring parameter, interaction, or adverse-effect attribution follows from the patient's physiology and evidence, then the final answer has revealed little about independent understanding. Robustness should therefore be examined through controlled changes in case information, source quality, prompt context, or competing therapeutic options. The educational concern is not variability itself; competent clinicians revise judgments when relevant information changes. The concern is unexplained dependence on irrelevant variation or inability to identify which information should have changed the decision.

What can be delegated without delegating accountability?

Not every intellectual operation performed during pharmacotherapy work needs to remain manual. Generative AI can reasonably assist with activities such as organizing information, producing an initial list of therapeutic options, reformulating explanations for different audiences, identifying questions that require investigation, summarizing candidate considerations, or accelerating retrieval of potentially relevant material. Selected pharmacy studies show that LLMs can provide useful outputs under defined conditions, particularly when the task is bounded and the result undergoes professional verification[14]. Human–AI augmentation is therefore compatible with the educational objective of developing independent clinical judgment[18].

The critical distinction is whether delegation removes evidence of a capability that the profession expects the learner to possess. Information retrieval may be assisted without relinquishing responsibility for determining whether the information is authentic and applicable. Drafting may be assisted without relinquishing responsibility for the pharmacological reasoning underlying the draft. Generating multiple options may be assisted without relinquishing responsibility for selecting among them. In contrast, evidence verification, mechanistic interpretation, patient-specific integration, uncertainty recognition, and final therapeutic justification cannot simply disappear from assessment because an AI system can produce plausible versions of them. The educational requirement concerns attribution, not technical exclusivity: AI may perform parts of these activities, but the learner must still be capable of demonstrating them independently when competence is being judged.

Professional accountability is clearest when a recommendation can be reconstructed. The learner should be able to identify the evidence used, explain why the source applies, connect pharmacological mechanism and exposure to the patient's state, describe relevant competing interpretations, state which uncertainty remains, and justify why the proposed action is preferable to alternatives. Entrustable professional activities place new pharmacy graduates

in roles that require assessment, planning, implementation, monitoring, and communication[4]. When AI participates in those activities, accountability cannot be inferred from the fluency of the joint product. It depends on whether the learner can defend and revise the decision when the generated scaffold is challenged.

Figure 1 represents AI assistance and learner accountability through verification pathways connecting generated output to the clinical decision.

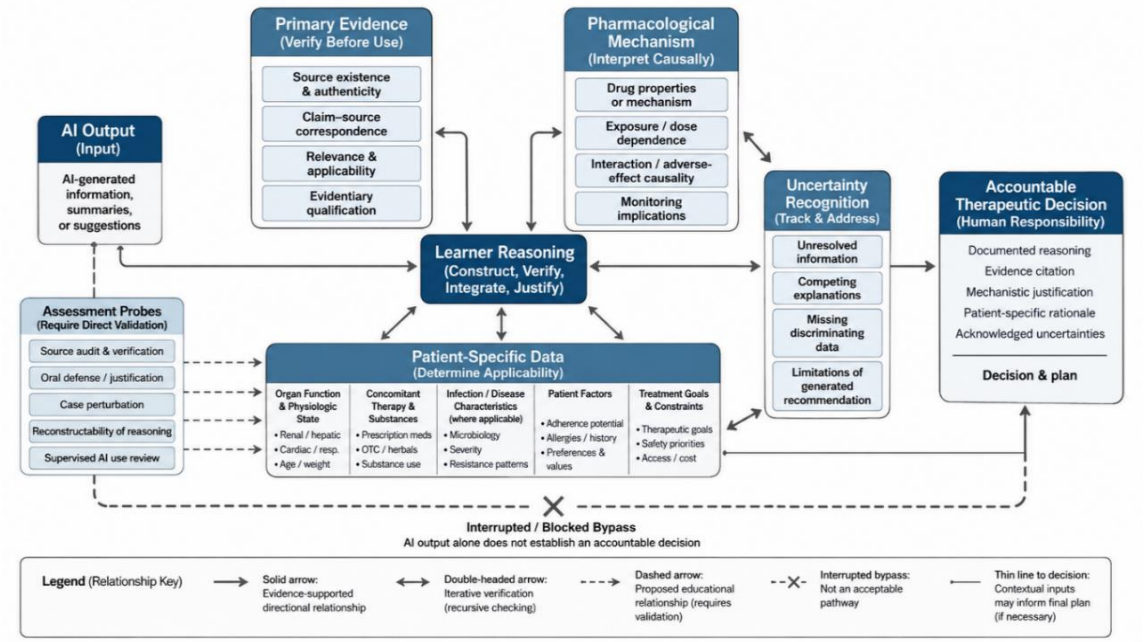

Figure 1. Verification pathways required to preserve learner-attributable clinical reasoning during generative-AI-assisted pharmacotherapy decisions

Table 1. Clinical-reasoning functions that remain attributable to the learner during generative-AI-assisted pharmacotherapy work

Reasoning function	Appropriate AI assistance	Learner action that must remain observable	Evidence of independent understanding	Failure pattern that should trigger concern
Information retrieval and organization	Locate candidate information, summarize material, structure alternatives	Identify which retrieved information is relevant to the clinical problem	Can locate or confirm essential information independently when required	Accepts generated material because it is fluent or familiar
Source verification	Suggest citations or evidence sources	Confirm source existence, authority, currency, and applicability	Can retrieve the source and show correspondence between evidence and claim	Cannot locate the cited source or distinguish fabricated from authentic evidence
Claim–evidence matching	Summarize study or guideline statements	Determine whether the source actually supports the generated claim	Can identify overstatement, missing qualification, or population mismatch	Treats citation presence as proof of evidentiary support
Causal pharmacology	Generate candidate mechanisms, interactions, adverse effects, or dosing considerations	Explain how drug properties, exposure, mechanism, and clinical state produce the treatment consequence	Can reconstruct the mechanistic chain without relying on generated wording	Produces the correct recommendation but cannot explain why it is pharmacologically appropriate
Patient-specific integration	List relevant patient variables or alternative options	Decide which patient variables materially change the recommendation	Can revise treatment appropriately when clinically meaningful patient information changes	Repeats a generic recommendation despite altered renal function, interactions, adherence, monitoring, or goals

Uncertainty recognition	Generate possible limitations or alternative explanations	State what remains unknown and which missing information could change the decision	Can distinguish uncertainty from factual error and identify discriminating information	Accepts unjustified certainty or adds generic disclaimers without decision relevance
Accountable therapeutic recommendation	Draft or format a final plan	Select, justify, defend, and revise the final recommendation	Can reconstruct the evidence-to-decision pathway under questioning	Responsibility collapses to "the AI recommended it"

Assessment tasks that expose independent understanding

Assessment design should follow the same distinction between product quality and reasoning attribution. Generative AI can assist educators as well as students. In pharmacy education, GPT-4 has been used to generate examination items, but expert review and revision remained necessary before many items were considered suitable for use[19]. This experience is instructive because it separates efficiency from validity. AI may accelerate construction of an assessment artifact without determining what that artifact validly measures. The same principle applies when evaluating learners: a polished AI-assisted response may be educationally useful while remaining insufficient evidence of independent clinical reasoning.

A multimethod approach is therefore preferable when high-stakes claims are made about learner competence. Oral defense can require the learner to explain why a recommendation follows from evidence and pharmacology rather than merely restate the submitted answer. A source audit can require retrieval of the evidence underlying selected claims and identification of unsupported or fabricated references. Traceable-reasoning tasks can ask the learner to show which patient facts changed the therapeutic decision and which information was considered but rejected. These approaches are consistent with established clinical-reasoning assessment principles, although their specific use for detecting GenAI-dependent performance requires prospective validation[3].

Case perturbation may be particularly informative. After a learner submits an AI-assisted treatment plan, the assessor can alter one clinically meaningful variable: renal function, concentration timing, interacting therapy, pregnancy status, microbiological susceptibility, adverse-effect chronology, therapeutic target, or patient preference. Independent understanding predicts that the learner can identify whether the new information matters, explain why, and revise the decision accordingly. Changing an irrelevant detail provides a complementary test of robustness. A learner who substantially alters the recommendation without a pharmacological reason may be reproducing generated output rather than applying a stable causal model of the case.

Supervised AI assessment provides another option. Instead of banning the technology, learners can be observed using it while being required to identify uncertainties, verify sources, reject incorrect suggestions, and justify final decisions. Such assessments measure a different but professionally relevant capability: safe augmentation. They should not replace opportunities to demonstrate unaided foundational competence, because the two assessment claims differ. Together, supervised and independent conditions can reveal whether AI is extending an existing reasoning capability or substituting for one that has not yet been established.

These approaches differ in the type of evidence they obtain about understanding and therefore should not be treated as interchangeable safeguards.

Table 2. Assessment designs for distinguishing generated performance from independently understood pharmacotherapy reasoning

Assessment design	Primary signal obtained	Required learner performance	AI dependence the task can expose	Principal limitation	Validation priority
Oral defense	Reconstructability of reasoning	Explain evidence, mechanism, patient fit, alternatives, and uncertainty in real time	Memorized or copied final answer without underlying understanding	Examiner variability; communication skill may influence performance	Standardized prompts and scoring criteria
Traceable reasoning record	Visibility of the evidence-to-decision path	Identify decisive patient facts, evidence sources, discarded alternatives, and decision changes	Unexplained jumps from generated output to recommendation	Can become performative if documentation is completed retrospectively	Prospective capture and authenticity procedures

Source audit	Independent evidence verification	Retrieve cited evidence and demonstrate claim–source correspondence	Fabricated, irrelevant, obsolete, or misrepresented sources	Tests verification more directly than broader clinical reasoning	Reliable sampling of claims and source quality
Case perturbation	Adaptability of causal reasoning	Revise or defend a recommendation after a clinically meaningful variable changes	Dependence on fixed wording, memorized output, or shallow pattern matching	Requires carefully designed perturbations with defensible expected effects	Item design and inter-rater agreement
Supervised AI use	Quality of human–AI collaboration	Query, challenge, verify, reject, and appropriately incorporate generated suggestions	Passive acceptance and failure to detect plausible errors	Does not establish unaided competence	Observation protocols and performance standards

Educational implications

The evidence does not currently justify the claim that generative AI inherently degrades clinical reasoning. An updated review of undergraduate medical education found no included studies directly evaluating AI's effects on medical students' critical thinking or clinical reasoning, and broader medical-education reviews continue to describe important methodological and implementation gaps[20, 21]. Pharmacy education should therefore avoid building assessment policy around an unproven deterioration narrative. The more defensible concern is evidentiary: traditional products no longer reveal learner contribution as clearly when capable generative tools are available.

Constructive uses of AI also argue against prohibition as the default educational response. A pharmacy virtual-tutor proof of concept demonstrated high learner uptake and favorable perceptions, including increased confidence and perceived understanding, although those outcomes should not be mistaken for controlled evidence of durable learning[22]. Student pharmacists have likewise described both anticipated benefits and concerns involving errors, ethics, limited knowledge, and implications for patient-centered care[23]. These findings suggest that learners are already negotiating appropriate and inappropriate uses of AI before education has established mature assessment standards.

Curriculum design should consequently make verification and evaluative AI literacy explicit competencies. Earlier medical-education reviews emphasized AI literacy, ethics, oversight, and the ability to evaluate technology rather than merely use it[24]. Surveys of clinicians and students have also found limited familiarity and formal preparation for AI while identifying concerns about professional skills and implementation[25]. Pharmacy curricula can make these general priorities more discipline-specific by requiring learners to verify drug-information sources, reconstruct causal pharmacology, evaluate patient applicability, recognize uncertainty, and document responsibility for the final therapeutic decision.

Faculty development is equally important. Educators need assessment designs that specify the claim being made before choosing whether AI is permitted. A task intended to assess independent pharmacological reasoning requires conditions that expose that reasoning. A task intended to assess safe professional augmentation should permit AI and evaluate how well the learner verifies and governs its contribution. Programs should avoid treating these as the same construct simply because both culminate in a therapeutic recommendation.

The most consequential curricular change may therefore be a shift from policing AI presence to specifying evidence of competence. Where independent knowledge and reasoning are prerequisites for safe practice, learners should demonstrate them under conditions in which machine output cannot substitute for explanation. Where professional practice is expected to involve AI, learners should demonstrate that they can use it without transferring verification or responsibility to the tool. These complementary conditions preserve both educational integrity and realistic preparation for technology-enabled pharmacy practice.

Conclusion

Generative AI makes fluent therapeutic answers easier to produce, but it does not eliminate the need to determine who understood the reasoning behind them. Pharmacy education should distinguish the quality of a generated product from evidence that the learner can verify sources, explain causal pharmacology, integrate patient-specific information, recognize uncertainty, and defend an accountable therapeutic decision. Current evidence supports neither unrestricted delegation nor a categorical assumption that AI use damages learning. It supports a more

precise educational response: permit augmentation where it adds value, while designing assessments that make consequential reasoning reconstructable and attributable.

A proposed test of independent understanding goes beyond reproducing a correct answer after AI assistance. It asks whether the learner can explain why the answer follows from trustworthy evidence and pharmacology, determine when patient information changes the decision, detect unsupported or misleading output, and revise the recommendation when the clinical conditions change. These capabilities provide a defensible boundary between assistance and outsourcing. Future educational research should test whether oral defense, source audit, traceable reasoning, case perturbation, and supervised AI assessment can reliably discriminate genuine understanding from successful output generation.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Abdel Aziz MH, Rowe C, Southwood R, Nogid A, Berman S, Gustafson K. A scoping review of artificial intelligence within pharmacy education. Am J Pharm Educ. 2024;88(1):100615. doi:10.1016/j.ajpe.2023.100615
2. Young M, Thomas A, Gordon D, Gruppen L, Lubarsky S, Rencic J, et al. The terminology of clinical reasoning in health professions education: Implications and considerations. Med Teach. 2019;41(11):1277-84. doi:10.1080/0142159X.2019.1635686
3. Daniel M, Rencic J, Durning SJ, Holmboe E, Santen SA, Lang V, et al. Clinical reasoning assessment methods: A scoping review and practical guidance. Acad Med. 2019;94(6):902-12. doi:10.1097/ACM.0000000000002618
4. Kanmaz TJ, Culhane NS, Berenbrok LA, Jarrett J, Johanson EL, Ruehter VL, et al. A curriculum crosswalk of the Core Entrustable Professional Activities for new pharmacy graduates. Am J Pharm Educ. 2020;84(11):8077. doi:10.5688/ajpe8077
5. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. 2023;9. doi:10.2196/45312
6. Singhal K, Azizi S, Tu T, Mahdavi SS, Wei J, Chung HW, et al. Large language models encode clinical knowledge. Nature. 2023;620(7972):172-80. doi:10.1038/s41586-023-06291-2
7. Ehlert A, Ehlert B, Cao B, Morbitzer K. Large language models and the North American Pharmacist Licensure Examination (NAPLEX) practice questions. Am J Pharm Educ. 2024;88(11):101294. doi:10.1016/j.ajpe.2024.101294
8. Hager P, Jungmann F, Holland R, Bhagat K, Hubrecht I, Knauer M, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med. 2024;30(9):2613-22. doi:10.1038/s41591-024-03097-1
9. Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, et al. Large language model influence on diagnostic reasoning: A randomized clinical trial. JAMA Netw Open. 2024;7(10). doi:10.1001/jamanetworkopen.2024.40969
10. Al-Dujaili Z, Omari S, Pillai J, Al Faraj A. Assessing the accuracy and consistency of ChatGPT in clinical pharmacy management: A preliminary analysis with clinical pharmacy experts worldwide. Res Social Adm Pharm. 2023;19(12):1590-4. doi:10.1016/j.sapharm.2023.08.012
11. Hsu HY, Hsu KC, Hou SY, Wu CL, Hsieh YW, Cheng YD. Examining real-world medication consultations and drug-herb interactions: ChatGPT performance evaluation. JMIR Med Educ. 2023;9. doi:10.2196/48433

12. Huang X, Estau D, Liu X, Yu Y, Qin J, Li Z. Evaluating the performance of ChatGPT in clinical pharmacy: A comparative study of ChatGPT and clinical pharmacists. Br J Clin Pharmacol. 2024;90(1):232-8. doi:10.1111/bcp.15896
13. Grossman S, Zerilli T, Nathan JP. Appropriateness of ChatGPT as a resource for medication-related questions. Br J Clin Pharmacol. 2024;90(10):2691-5. doi:10.1111/bcp.16212
14. Shin E, Hartman M, Ramanathan M. Performance of the ChatGPT large language model for decision support in community pharmacy. Br J Clin Pharmacol. 2024;90(12):3320-33. doi:10.1111/bcp.16215
15. Carou-Senra P, Delgado-Taboada I, Alvarez-Lorenzo C, Diaz-Rodriguez P, Goyanes A. Exploring the role of LLMs like ChatGPT in pharmacy education for supporting students' therapeutic decision-making. Am J Pharm Educ. 2025;89(8):101462. doi:10.1016/j.ajpe.2025.101462
16. Lyell D, Coiera E. Automation bias and verification complexity: A systematic review. J Am Med Inform Assoc. 2017;24(2):423-31. doi:10.1093/jamia/ocw105
17. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Soc Sci Med. 2020;260:113172. doi:10.1016/j.socscimed.2020.113172
18. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7
19. Shultz B, DiDomenico RJ, Goliak K, Mucksavage J. Exploratory assessment of GPT-4's effectiveness in generating valid exam items in pharmacy education. Am J Pharm Educ. 2025;89(5):101405. doi:10.1016/j.ajpe.2025.101405
20. Simoni J, Urtubia-Fernandez J, Mengual E, Simoni DA, Royo M, Egaña-Yin D, et al. Artificial intelligence in undergraduate medical education: An updated scoping review. BMC Med Educ. 2025;25(1):1609. doi:10.1186/s12909-025-08188-2
21. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME Guide No. 84. Med Teach. 2024;46(4):446-70. doi:10.1080/0142159X.2024.2314198
22. Cain J, Rajan AS. Proof of concept of ChatGPT as a virtual tutor. Am J Pharm Educ. 2024;88(12):101333. doi:10.1016/j.ajpe.2024.101333
23. Zhang X, Tsang CCS, Ford DD, Wang J. Student pharmacists' perceptions of artificial intelligence and machine learning in pharmacy practice and pharmacy education. Am J Pharm Educ. 2024;88(12):101309. doi:10.1016/j.ajpe.2024.101309
24. Lee J, Wu AS, Li D, Kulasegaram KM. Artificial intelligence in undergraduate medical education: A scoping review. Acad Med. 2021;96(11S). doi:10.1097/ACM.0000000000004291
25. Boillat T, Nawaz FA, Rivas H. Readiness to embrace artificial intelligence among medical doctors and students: Questionnaire-based study. JMIR Med Educ. 2022;8(2). doi:10.2196/34973