Review Article

# Teaching Pharmacists to Challenge Artificial Intelligence: A Critical Review of Curricula, Competency Claims, and Assessment

**Lukas Weber[1]*, Sophie Janssen[2], Jonas Richter[1], Elena Fischer[3]**

[1]Department of Pharmacy Education and AI Literacy, Faculty of Pharmacy, University of Stuttgart, Stuttgart, Germany. [2]Department of AI Competency Assessment in Pharmacy, Faculty of Pharmaceutical Sciences, Eindhoven University of Technology, Eindhoven, Netherlands. [3]Department of Critical AI Thinking in Pharmacy, Faculty of Pharmacy, University of Basel, Basel, Switzerland.

## Abstract

Artificial intelligence is entering pharmacy education faster than evidence is establishing what pharmacists must know, demonstrate, and resist. This critical review examined whether curricula, competency claims, and assessments prepare pharmacists to challenge artificial intelligence outputs rather than merely operate tools. A transparent critical-review method with SANRA-informed quality controls was applied to peer-reviewed literature published from 2017 through 2025. Searches of PubMed/MEDLINE, targeted journal and publisher platforms, and backward and forward citation trails were executed on 2 August 2026 with publication-date censoring. Eligible reports addressed pharmacy curricula, competency frameworks, educational interventions, assessment, or directly transferable health-professions evidence relevant to verification, uncertainty, bias, hallucination, refusal, escalation, accountability, workplace performance, or patient communication. Records were deduplicated, screened, linked to study families, extracted using predefined fields, and appraised according to design. Heterogeneous findings were compared narratively without quantitative pooling. Fifty-nine retrievals yielded 38 unique records, 27 assessed full texts, and 18 included reports representing 18 studies: six pharmacy-specific and 12 transferable reports. Pharmacy-specific evidence chiefly described exposure, perceptions, task completion, item generation, or artificial-intelligence-assisted grading. Direct assessment of independent verification, justified refusal, escalation, documented accountability, or workplace transfer was scarce. Transferable frameworks widened the competency scope but did not validate pharmacy-specific performance. The review was limited by a single-reviewer workflow, source-interface constraints, methodological heterogeneity, and the later search execution date. Pharmacy education should distinguish tool use from challenge competence and align claims with observable, authentic performance across realistic medication decisions involving patients, teams, and governance constraints. Proposed domains require prospective validation before curricular standardization.

**Keywords:** Critical review, Digital pharmacy, Artificial intelligence, Pharmacy practice, Medication safety, Evidence synthesis

## Introduction

Artificial intelligence has moved from a specialist topic to a visible component of health-professions education, propelled by predictive systems, conversational models, automated assessment, and generative tools that can produce apparently authoritative explanations within seconds. Pharmacy programmes consequently face pressure to introduce learners to systems that may influence drug-information retrieval, counselling, documentation, assessment, and medication-related decisions. Yet curricular presence does not establish educational value. Early scholarship argued that artificial intelligence changes professional education because information access becomes less scarce while interpretation, judgement, and responsibility become more important; reviews also documented heterogeneous applications and unresolved implementation challenges [1-3].

The central educational problem is therefore not whether pharmacists will encounter artificial intelligence, but what kind of competence should govern that encounter.

**Address for correspondence:** Lukas Weber, Department of Pharmacy Education and AI Literacy, Faculty of Pharmacy, University of Stuttgart, Stuttgart, Germany.
lukas.weber@uni-stuttgart.de
**Received:** 27 August 2025; **Accepted:** 05 December 2025



**How to cite this article:** Weber L, Janssen S, Richter J, Fischer E. Teaching Pharmacists to Challenge Artificial Intelligence: A Critical Review of Curricula, Competency Claims, and Assessment. Arch Pharm Pract. 2025;16(4):82-91. https://doi.org/10.51847/Y20jhx7z1e

Three levels require separation. Using artificial intelligence means accessing a system, constructing prompts, navigating an interface, or completing a task with model assistance. Understanding it requires knowledge of intended purpose, data dependence, validation, performance measures, limitations, and ethical implications. Challenging it requires independent verification, recognition of uncertainty and bias, detection of fabricated or inconsistent content, refusal when evidential or safety conditions fail, escalation to appropriate expertise, and documentation of who remains accountable. This tripartite distinction adapts broader medical-education arguments that conceptual literacy and critical appraisal matter more than universal coding proficiency [4]. In pharmacy, the distinction is especially consequential because fluent output may concern doses, interactions, contraindications, evidence provenance, or patient-specific counselling.

Current educational claims often blur these levels. Learner satisfaction, willingness to use a tool, perceived preparedness, successful task completion, or improved efficiency may be valuable outcomes, but none alone demonstrates that a pharmacist can contest an unsafe recommendation. Conversely, resistance is frequently framed as an adoption barrier, although principled nonuse may represent appropriate professional judgement. Assessment design determines whether these distinctions remain rhetorical or become observable. An authentic test must reveal what a learner does when artificial intelligence is persuasive but wrong, incomplete, inequitable, or unsupported, not merely whether the learner can obtain a plausible response.

The distinction also changes faculty responsibility. Educators must be able to design adversarial cases, identify authoritative verification sources, explain escalation routes, and judge reasoning rather than reward polished output. Workplace constraints and patient perspectives matter because challenge competence is exercised through teams, documentation, communication, and governance. A classroom activity that omits these conditions may teach familiarity while leaving authority and contestability untested.

This review critically examines how artificial-intelligence competence is defined in pharmacy education; whether curricula prioritize operation, understanding, critical appraisal, or combinations of these aims; how verification, bias, uncertainty, hallucination, refusal, and escalation are taught; what assessment methods test authentic performance; which competency claims rest primarily on self-report; and what challenge-oriented domains and reforms are defensible from the evidence. Its purpose is argumentative rather than cataloguing: to compare claimed competence with teaching activity and assessment evidence, identify contradictions and omissions, and distinguish pharmacy-specific findings from transferable health-professions frameworks. The review does not infer clinical effectiveness, medication safety, workplace readiness, or universal curricular applicability from publication frequency, learner enthusiasm, model performance, or consensus language.

## Review Design, Questions, and Protocol

This review used a transparent critical-review design to evaluate the adequacy of educational claims rather than estimate a pooled intervention effect. SANRA-informed quality controls were used to make the review purpose, search description, referencing, scientific reasoning, and endpoint explicit, without assigning a formal SANRA score [5]. Methodological decisions aligned the questions, evidence boundaries, appraisal, and synthesis with the review's critical purpose; comprehensiveness was pursued through multiple sources, but publication volume was not treated as evidential strength and systematic-review exhaustiveness was not claimed [6].

The protocol was frozen before final record capture. It specified the 2017–2025 publication window, English-language peer-reviewed journal reports, Q1 journal boundary, eligible populations and settings, article-specific questions, source-type rules, extraction fields, full-text exclusion hierarchy, study-family linking, and distinction between pharmacy-specific and transferable evidence. The search was executed on 2 August 2026 and date-censored through 31 December 2025. This later execution date is reported because presenting it as a 2025 search would misstate the review chronology.

The scope included pharmacy learners, pharmacists, faculty, curricula, competency frameworks, educational interventions, and assessment studies. Directly relevant health-professions, workplace, patient, and psychometric literature was eligible only when it informed pharmacists' capacity to evaluate or challenge artificial intelligence. PRISMA terminology supported transparent record accounting and the flow visualization, but did not redefine the work as a systematic review [7]. The six analytical questions addressed competency definitions, curricular emphasis, teaching of challenge behaviours, authenticity of assessment, reliance on self-report, and defensible reform.

## Eligibility, Definitions, and Information Sources

Eligible evidence had to provide detail to appraise an educational objective, competency claim, learning design, assessment method, or transferable decision-work implication. Curricula and interventions were included when they described what learners were expected to know, do, judge, verify, communicate, refuse, or escalate. Competency frameworks were included when domains and behaviours could be extracted. Assessment studies were included when the evaluated construct, task, comparator, scoring approach, or validity evidence was identifiable. Studies of artificial intelligence as an assessor were retained only to examine assessment design and human oversight; their model performance was not interpreted as learner competence.

Exclusions comprised tool demonstrations without educational objectives or assessment, nonempirical opinion presented as evidence of attainment, perception or readiness studies when competence was claimed without observed performance, nontransferable education literature, publications outside the date window, and journals outside the prespecified Q1 boundary. The critical analytical lens separated use, foundational understanding, evidence verification, uncertainty appraisal, bias recognition, hallucination detection, professional refusal, escalation, authentic assessment, workplace competence, and patient perspective. These were review operationalizations, not validated competency standards.

Information sources were PubMed/MEDLINE, targeted journal and publisher platforms, and backward and forward citation searching from anchor reviews and frameworks. Search concepts combined pharmacy or health-professions populations, artificial-intelligence technologies, educational phenomena, and evaluation terms. Platforms, exact queries, filters, dates, capture counts, and completion notes were recorded. Eligibility mapping, reproducible search reporting, and concept-block construction followed relevant reporting and search-method guidance [8-10].

**Table 1** organizes the evidence, constructs, relationships, uncertainties, and boundary conditions required for eligibility, definitions, and evidence-extraction framework for the review Teaching Pharmacists to Challenge Artificial Intelligence: A Critical Review of Curricula, Competency Claims, and Assessment.

**Table 1.** Eligibility, definitions, and evidence-extraction framework for the review Teaching Pharmacists to Challenge Artificial Intelligence: A Critical Review of Curricula, Competency Claims, and Assessment.

| Review element | Operational definition | Eligibility or decision rule | Data field | Rationale | Bias or ambiguity risk | Planned synthesis use |
|---|---|---|---|---|---|---|
| Review boundary | Pharmacy-specific evidence and directly transferable health-professions, workplace, patient, or assessment evidence relevant to challenging artificial intelligence | Include sources with extractable learning, competency, assessment, workflow, or patient relevance; exclude generic technology promotion | Profession; setting; source type; relevance pathway | Keeps the review centred on medication-decision work | Transferable evidence may be mistaken for pharmacy-specific validation | Separate direct pharmacy findings from transferable interpretation |
| Competence claim | An observable claim concerning knowledge, skill, judgement, communication, or performance | Retain claims only when the source identifies what learners or practitioners can know, do, judge, or demonstrate | Claim wording; construct; evidence source; outcome level | Prevents broad competency language from exceeding supporting evidence | Confidence, satisfaction, exposure, or willingness may be relabelled as competence | Compare claimed competence with its evidential basis |
| Artificial-intelligence use or operation | Ability to access, prompt, navigate, or complete a task with an artificial-intelligence tool | Include when operation is an explicit objective or assessed activity; do not classify operation alone as challenge competence | Learning activity; task; tool; assessment method | Distinguishes functional familiarity from critical capability | Successful task completion may conceal unverified output | Classify curricula by operational emphasis |
| Foundational understanding | Knowledge of model purpose, data dependence, validation, performance measures, limitations, and appropriate task boundaries | Include conceptual learning objectives; coding is not required unless specified by the curriculum | Content domain; learning objective; assessment | Establishes the knowledge needed to interpret system output | Technical vocabulary may be mistaken for applied understanding | Identify the depth of artificial-intelligence literacy |
| Evidence verification | Independent checking of claims, calculations, citations, guidelines, patient data, and provenance | Verification must use an external authoritative source or reproducible calculation | Verification task; source consulted; detected error; response | Tests whether learners can challenge apparently credible content | Plausibility checking may be reported as verification | Evaluate direct preparation for evidence checking |
| Uncertainty, bias, and hallucination appraisal | Recognition of missing information, calibration limits, representation or measurement bias, fabricated content, inconsistency, and contextual mismatch | Include explicit teaching or assessment of detection and response; awareness without action is insufficient | Error type; uncertainty cue; bias domain; learner response | Groups related failure modes while preserving distinct response requirements | General ethics teaching may be interpreted as demonstrated error detection | Compare coverage of critical-risk domains |
| Professional refusal and escalation | Ability to reject, stop, qualify, override, or refer an artificial-intelligence-mediated task when evidential or safety conditions fail | Include explicit refusal or escalation triggers, responsible recipient, urgency, and documentation expectations | Trigger; action; recipient; rationale; documentation | Treats principled nonuse as a professional capability | Refusal may be misclassified as resistance to innovation | Assess whether curricula preserve professional authority |

| | | | | | | |
|---|---|---|---|---|---|---|
| Authentic assessment | Performance in a realistic medication-related task requiring verification, challenge, documentation, communication, or escalation | Self-reported confidence or satisfaction alone does not demonstrate authentic performance | Scenario; comparator; scoring rubric; assessor; validity evidence | Aligns competency claims with observable behaviour | Simulated performance may still overstate workplace transfer | Rank assessment evidence by authenticity |
| Workplace and patient relevance | Performance under workflow constraints, team dependencies, local governance, patient communication, and contestability | Course completion is not assumed to transfer to practice; patient acceptance is not equated with validity or safety | Setting; stakeholders; workflow constraint; patient outcome or perspective | Extends competence beyond isolated classroom interaction | Professional and patient evidence may originate outside pharmacy | Identify transfer gaps and contextual requirements |

Table note: The definitions, classifications, and decision rules are review operationalizations derived from the frozen protocol. They are not validated competency standards, clinical recommendations, or deployment criteria.

## Search, Selection, Extraction, Appraisal, and Synthesis

All captured citations received record identifiers. Deduplication prioritized DOI, then normalized title, first author, year, and study details; each retrieval was linked to a canonical record. One reviewer screened titles and abstracts, sought full texts, assigned one primary exclusion reason to every excluded report, linked multiple reports through study-family identifiers, and completed extraction fields. No independent duplicate screening or inter-rater agreement was claimed.

Appraisal and synthesis were design-sensitive and nonquantitative [11-13]. Review-level sources were examined for search transparency, selection logic, appraisal, and synthesis credibility. Qualitative, quantitative, mixed-methods, survey, curriculum, consensus, and psychometric reports were judged against design-appropriate questions rather than collapsed into one universal score. Extraction captured setting, participants or units, artificial-intelligence focus, educational objective, learning activity, assessment method, outcome level, competency claim, challenge-domain coverage, applicability, and limitations.

Comparison proceeded across source type, profession, claimed competence, teaching activity, assessment evidence, and validation level. Findings were retained only at method-supported levels. Contradictions were examined by distinguishing attitudes from performance, feasibility from validity, model accuracy from learner capability, classroom outcomes from workplace transfer, and consensus from empirical demonstration. No meta-analysis was attempted because interventions, constructs, populations, and outcomes were not commensurable.

Reflexivity was addressed through frozen definitions, a prespecified exclusion hierarchy, positive record-level traceability, and explicit labelling of critical interpretation and original synthesis. Nevertheless, single-reviewer decisions, English-language restriction, interface-dependent capture, Q1 eligibility, and reliance on transferable evidence may have shaped the corpus. Appraisal moderated each resulting interpretation; it did not produce a universal certainty grade.

## Search and Selection Results

The search identified 59 retrievals: 31 from PubMed/MEDLINE, 17 from targeted journal and publisher searches, and 11 from backward and forward citation searching. Removing 21 duplicate retrievals left 38 unique records for title and abstract screening. Eleven records were excluded at that stage, leaving 27 reports sought and successfully retrieved. All 27 full texts were assessed. Nine were excluded: one nonempirical opinion without competency-attainment evidence, three perception or readiness reports without demonstrated competence, three reports from journals outside the prespecified Q1 boundary, and two reports with insufficient challenge-oriented extraction detail or redundant review evidence. Eighteen reports representing 18 distinct studies entered the critical synthesis, comprising six pharmacy-specific and 12 transferable reports. The earlier pharmacy evidence map similarly characterized the field as small and heterogeneous, although its boundaries and search period differed from this review [14]. These values describe a finite captured corpus, not a systematic census of every educational publication.

**Figure 1** presents the completed review-selection pathway in a black-and-white prisma-style flow diagram for evidence selection, documenting identification, deduplication, screening, eligibility, exclusions with reasons, and final inclusion using the values approved in the Part 1 PRISMA Record-Accounting Audit.

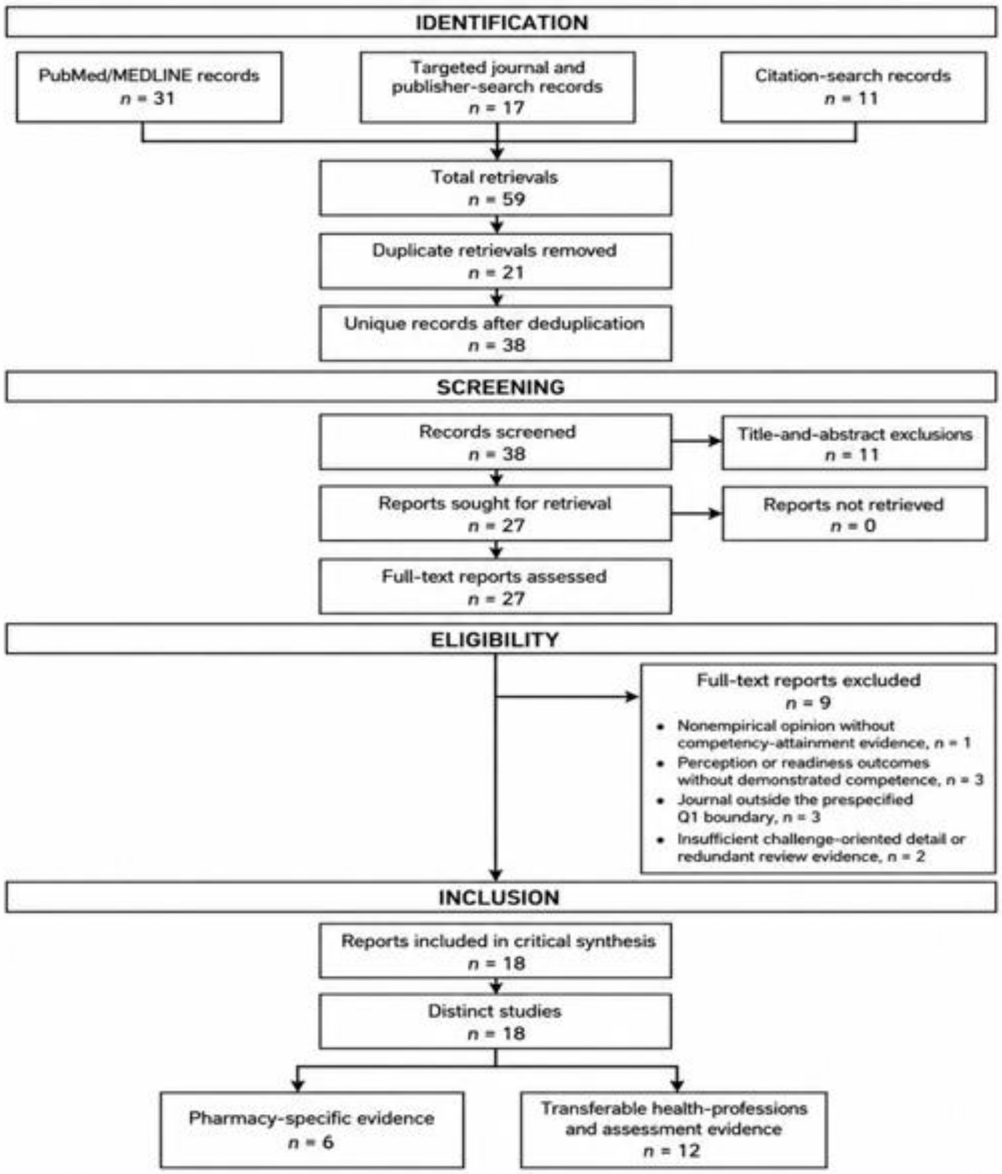


**Figure 1.** Black-and-White PRISMA-Style Flow Diagram for Evidence Selection

Across the included evidence, "competence" was used inconsistently. In pharmacy-specific work, it most often referred to exposure, perceived preparedness, task completion, or the successful incorporation of an artificial-intelligence tool. A medicinal-chemistry project demonstrated that learners could use artificial intelligence within a curricular assignment, but its educational design did not directly test whether they could independently verify, reject, or escalate unsafe output [15].

A student-pharmacist survey documented interest, anticipated relevance, and perceived curricular need; those findings described attitudes rather than observed competence [16]. Consequently, publication of a course or positive learner response could not be treated as evidence that challenge-oriented capability had been achieved.

Curricular emphasis was predominantly operational. Learners prompted systems, generated content, explored applications, or completed disciplinary tasks with artificial-intelligence assistance. Some reports discussed ethics, limitations, or critical appraisal, but the actions required when output was wrong remained less explicit. The corpus rarely specified an authoritative verification source, a minimum evidential threshold, a refusal trigger, an escalation recipient, or an accountability record. This gap matters because a learner may know that hallucinations occur yet still fail to check a fabricated reference, correct a dose, communicate uncertainty, or stop inappropriate use.

## Evidence-Base Characteristics

The evidence base contained six pharmacy-specific reports and 12 transferable reports, but these categories offered different warrants. A pharmacogenetic-counselling skills laboratory placed artificial intelligence within a realistic pharmacy communication task and provided experiential evidence; it did not establish durable workplace transfer or independent challenge across cases [17].

An assessment-item study showed that GPT-4 could generate potentially usable pharmacy questions, yet expert review and revision remained necessary, illustrating that fluent output cannot substitute for content validation [18].

A pilot of artificial-intelligence-assisted objective structured clinical examination grading examined agreement, efficiency, and consistency of an assessor system rather than learners' capacity to contest artificial intelligence [19]. Thus, even apparently performance-oriented pharmacy studies frequently evaluated tool feasibility or assessor behaviour instead of challenge competence.

Coverage of evidence verification was partial. Some activities encouraged comparison with course materials or professional sources, but explicit verification procedures were not consistently described or scored. Bias and uncertainty were more commonly named as topics than translated into observable decisions. Hallucination was often presented as a known limitation, yet few educational tasks deliberately planted fabricated citations, contradictory recommendations, missing patient variables, or confident but unsupported claims. The review therefore distinguished awareness from response: recognizing that a model can err is not equivalent to detecting the error, establishing its consequence, and acting appropriately.

Methodological strength varied within both branches. Pharmacy studies frequently offered concrete educational or assessment settings, but they were usually bounded by one course, one cohort, exploratory design, or outcomes close to the intervention. Transferable reviews and consensus frameworks broadened the construct map, yet their heterogeneity and professional distance limited direct applicability. No synthesis member combined pharmacy-specific curriculum design, objective adversarial assessment, workplace observation, patient involvement, and longitudinal follow-up. Accordingly, the strongest conclusion concerns a mismatch between broad competency language and narrow evidence, not the comparative effectiveness of particular teaching models.

## Teaching Professional Resistance and Refusal

Professional resistance requires careful definition. In this review, resistance did not mean generalized hostility to technology. It meant the capacity to withhold reliance, override output, or stop an artificial-intelligence-mediated task when evidential, ethical, or medication-safety conditions were not met. Transferable competency work extended professional capability beyond operation to evidence appraisal, workflow integration, ethics, communication, and ongoing evaluation [20]. That breadth supports a challenge-oriented interpretation, but the framework was not validated specifically for pharmacists and did not itself establish an assessment standard for refusal.

Implementation-focused competency mapping further implied that professionals need to recognize system limits, identify affected stakeholders, evaluate workflow consequences, and participate in monitoring [21]. These responsibilities logically create escalation duties when a problem exceeds the user's authority or expertise. However, escalation was seldom operationalized as a teachable sequence specifying the trigger, urgency, recipient, interim action, and documentation. Ethical education can also support principled refusal by asking learners to reason from obligations and contextual consequences rather than memorize abstract principles [22]. Still, a proposed ethics framework does not demonstrate that learners will refuse a persuasive but unsafe recommendation under time pressure.

The pharmacy-specific corpus offered little direct evidence that refusal or escalation was taught and assessed. No included study established a validated threshold for overriding an artificial-intelligence output, demonstrated entrustment for independent challenge, or observed longitudinal workplace behaviour. This absence does not prove that existing curricula never address such actions; it indicates that the published evidence did not make them visible enough to support competency claims. Refusal and escalation are therefore retained here as original review-derived domains requiring prospective pharmacy validation.

## Assessment Methods and Authenticity

Assessment methods ranged from surveys and course evaluations to project outputs, expert item review, simulated counselling, objective knowledge testing, and artificial-intelligence-assisted grading. Reviews of health-professions programmes found heterogeneous educational designs and inconsistent outcome evaluation, with many studies emphasizing knowledge, attitudes, or satisfaction [23]. Such outcomes may identify acceptability or learning needs, but they cannot independently support claims about safe performance. Curriculum proposals have appropriately argued for longitudinal teaching of limitations, transparency, ethics, and liability, yet proposed content remains distinct from validated assessment [24].

Self-reported confidence was particularly weak as a stand-alone measure. Confidence may rise because a learner becomes familiar with an interface, receives rapid output, or overestimates the system's reliability. It may also fall after education exposes complexity, despite better judgement. An objective generative-artificial-intelligence literacy test showed that directly measured literacy could contribute information beyond self-reported proficiency when predicting task performance [25]. Nevertheless, a psychometric knowledge test in general higher education is not equivalent to pharmacy challenge performance. It does not show whether a learner can identify a medication error, select an authoritative source, justify refusal, communicate with a patient, and escalate within a real workflow.

Authentic assessment should therefore create consequential disagreement between the learner and the system. A credible case would include incomplete patient data, plausible but fabricated evidence, subgroup bias, conflicting recommendations, or pressure to accept an efficient answer. Scoring should reward verification strategy, detection accuracy, uncertainty qualification, refusal or override when warranted, escalation, communication, and audit-ready documentation. The review proposes this assessment architecture as a defensible synthesis, not as a validated instrument or high-stakes standard.

Faculty capability was another implicit boundary. Challenge-oriented teaching requires educators to select authoritative comparators, construct realistic failure cases, distinguish uncertainty from error, and judge whether escalation is proportionate. It also requires policies for permitted tool use, disclosure, documentation, and remediation when learners accept unsafe output. The included pharmacy reports demonstrated faculty experimentation with projects, counselling, item review, and grading, but did not establish a reproducible faculty-development model. Without such preparation, curricula may assess prompt fluency because it is easy to observe while leaving the harder work of evidential challenge, relational communication, and accountable refusal unmeasured. This limitation applies to classroom, simulation, and workplace assessment, although its consequences differ across settings.

## Missing Workplace and Patient Perspectives

Workplace and patient perspectives were markedly underrepresented. Clinicians in primary care anticipated changes in professional roles, trust, deskilling, and accountability, indicating that artificial-intelligence competence is exercised within organizational relationships rather than isolated individual cognition [26]. These findings are transferable design signals, not direct evidence about pharmacists. They suggest that education should test whether learners preserve independent reasoning, recognize when automation redistributes responsibility, and act when organizational incentives favour speed over scrutiny.

Patient evidence also complicates the assumption that resistance is simply a deficit in literacy. Experimental research found that resistance to medical artificial intelligence can reflect concerns that systems will not recognize personal uniqueness or provide relational care [27]. Survey evidence on an artificial-intelligence-assisted symptom checker further showed that perceived usefulness is shaped at the interface between technical output, user expectations, and communication [28]. Neither context is identical to pharmacy practice, but both indicate that patient-facing competence includes explaining the system's role, limitations, alternatives, and routes for contesting a decision.

The review consequently proposes ten challenge-oriented domains: foundational artificial-intelligence literacy; independent evidence-source verification; uncertainty and calibration appraisal; bias and equity response; hallucination and inconsistency detection; professional refusal and override; escalation and consultation; documentation and accountability; patient communication and contestability; and workflow monitoring and learning. Each domain requires observable action, not only awareness. Their inclusion is an original synthesis derived from convergent gaps across pharmacy-specific and transferable evidence. They should guide hypothesis generation, curricular design, and prospective assessment research, but they should not be represented as a validated universal framework.

## Discussion: Principal Findings and Interpretation

The principal finding is not that pharmacy education has ignored artificial intelligence, but that its published claims often exceed the outcomes used to support them. Existing models correctly recognize that pharmacists need exposure to emerging tools, conceptual vocabulary, ethical awareness, and opportunities to practise with disciplinary tasks. The pharmacy studies also show constructive experimentation: artificial intelligence was incorporated into medicinal chemistry, pharmacogenetic counselling, item development, and objective structured clinical examination grading. These activities make technology concrete and can reveal practical limitations. However, the broader medical-education literature remains heterogeneous, mixes artificial intelligence used to teach with education about artificial intelligence, and provides limited high-quality evidence for durable educational outcomes [29]. Growth in activity therefore demonstrates curricular attention, not challenge competence.

A recurring problem is construct inflation. "Competence" may be inferred from familiarity, willingness, satisfaction, confidence, completion, or agreement with an artificial-intelligence system. Those indicators answer legitimate but narrower questions. Familiarity concerns exposure; confidence concerns self-perception; agreement concerns concordance; efficiency concerns workflow; and knowledge tests concern conceptual recognition. None alone warrants the conclusion that a pharmacist can safely adjudicate a contested medication decision. Competency statements derived through Delphi processes can clarify important content, but consensus descriptors still require translation into observable tasks, scoring rules, performance standards, and validity evidence [30]. Otherwise, the curriculum may name verification, ethics, or accountability without determining whether learners can enact them.

This distinction has direct consequences for pharmacy education. Artificial-intelligence teaching should not be isolated as a short technology module concerned mainly with terminology and prompting. It should be integrated across drug information, therapeutics, calculations, pharmacogenomics, communication, professional practice, and medication safety. The relevant question is how learners preserve pharmaceutical reasoning when a model intervenes in those activities. Broader digital-health competency work supports linking artificial intelligence to person-centredness, data stewardship, systems thinking, implementation, and interprofessional practice rather than treating it as a stand-alone tool [31]. Pharmacy programmes can use that systems orientation while retaining medicine-specific responsibilities for evidence provenance, dose reasoning, interaction assessment, counselling, and escalation.

Assessment reform should begin with explicit claim alignment. Before describing a learner as competent, educators should specify the target action, conditions, comparator, acceptable error boundary, and evidential warrant. A foundational assessment may test whether the learner can explain intended use and validation limitations. A verification assessment should require an external source and reproducible reasoning. An uncertainty assessment should require qualification or an information request. A refusal assessment should present pressure to accept fluent output despite a failed evidential threshold. An escalation assessment should require selection of the correct recipient, urgency, interim protection, and documentation. A patient-facing assessment should test understandable disclosure, alternatives, teach-back, and contestability.

Authenticity also requires cases in which artificial intelligence is neither consistently wrong nor consistently right. If every output contains an obvious error, learners can succeed through generalized distrust. If every output is correct, the assessment rewards compliance. Cases should vary error type, clinical consequence, uncertainty, patient characteristics, and workflow pressure. Scoring should distinguish detection from explanation, correction, communication, and follow-through. Multiple cases are needed because one successful challenge may reflect chance or topic familiarity. Consequences attached to the simulated decision should be visible so that learners understand why verification and refusal matter.

Current educational models also need a stronger account of professional resistance. Adoption frameworks often position resistance as an obstacle to be reduced, but medication practice sometimes requires deliberate nonuse. The educational objective is calibrated reliance: neither reflexive acceptance nor categorical rejection. Learners should be able to state when the system is useful, what evidence would justify reliance, what signals should lower confidence, and when professional authority must interrupt the process. Such judgement is relational and organizational. A pharmacist may need to challenge a prescriber, informatics team, vendor claim, local protocol, or colleague, not merely an on-screen answer.

Faculty development is therefore central. Educators need shared definitions, access to authoritative comparators, expertise in adversarial case design, and rubrics that capture reasoning rather than stylistic polish. They also need governance for tool access, privacy, disclosure, version changes, and retention of learner work. Evaluation should record model and version because educational performance may shift when systems change. Institutions should avoid investing high-stakes authority in an assessment tool merely because it improves speed or agreement during a pilot.

### Implications, Strengths, and Limitations

The review supports a staged research agenda. First, pharmacy-specific challenge domains should undergo content validation with educators, practising pharmacists, learners, patients, medication-safety specialists, informaticians, and regulators as stakeholders, without treating stakeholder agreement as outcome validation. Second, assessment tasks should be cognitively interviewed and piloted for realism, scoring clarity, subgroup fairness, and susceptibility to coaching. Third, prospective studies should compare self-report, objective knowledge, simulation performance, and workplace observation. Fourth, longitudinal research should examine retention, transfer across medication tasks, incident recognition, escalation quality, and patient communication. Fifth, multicentre studies should evaluate contextual variation rather than assuming one programme or health system represents the profession.

Educational content should preserve three boundaries. Explanation displays are not automatic justification, because interpretable outputs may still rest on unreliable relationships [32]. Challenge competence must extend beyond the encounter to lifecycle evaluation, dataset shift, incident reporting, and postdeployment monitoring [33]. Generative systems can combine impressive capability with hallucination, inconsistency, and contextual failure; teaching must therefore evaluate the task and version rather than confer trust [34].

Review strengths were the frozen operational definitions, record-level traceability, design-sensitive appraisal, explicit separation of pharmacy-specific from transferable evidence, and preservation of distinctions among attitudes, performance, validity, and workplace transfer. Limitations were single-reviewer screening and extraction, English-language and Q1 restrictions, interface-dependent source capture, methodological heterogeneity, and a search executed in 2026 but date-censored through 2025. These choices may have omitted relevant evidence or influenced interpretation. Accordingly, the framework remains evidence-bounded and cannot establish educational effectiveness, medication-safety benefit, or universal applicability across all pharmacy settings.

## Conclusion

Pharmacy education is beginning to engage seriously with artificial intelligence, but the published evidence does not yet show that pharmacists are being prepared consistently to challenge it. Pharmacy-specific studies demonstrate curricular experimentation, learner interest, disciplinary applications, and emerging uses in assessment. Their outcomes, however, primarily concern exposure, perceptions, task completion, feasibility, agreement, or efficiency. These outcomes should not be relabelled as verified competence in evidence checking, uncertainty appraisal, bias response, hallucination detection, refusal, escalation, accountability, workplace performance, or patient communication.

The critical distinction is between operating a system and exercising professional authority over its use. Challenge competence requires pharmacists to compare output with authoritative evidence, identify missing or distorted information, reject unsupported recommendations, communicate residual uncertainty, escalate unresolved risk, and document responsibility. It also requires attention to patients, teams, workflow pressures, and system monitoring. Transferable health-professions frameworks support this broader direction, but they do not validate pharmacy-specific curricula or assessment standards.

The review therefore proposes challenge-oriented domains and an assessment reform agenda centred on observable performance in realistic medication cases. These proposals are original synthesis, not a validated framework, clinical guideline, regulatory requirement, or deployment-ready standard. Future research should establish content validity, scoring reliability, fairness, longitudinal retention, workplace transfer, patient relevance, and relationships with medication-safety outcomes. Until that evidence exists, educators should make competency claims proportionate to what was actually taught and measured. Teaching pharmacists to use artificial intelligence is insufficient when professional responsibility requires knowing when, why, and how to challenge it. Such restraint remains an essential educational safeguard.

ACKNOWLEDGMENTS: None
CONFLICT OF INTEREST: None
FINANCIAL SUPPORT: None
ETHICS STATEMENT: None

## References

1. Wartman SA, Combs CD. Medical education must move from the information age to the age of artificial intelligence. Acad Med. 2018;93(8):1107-9. doi:10.1097/ACM.0000000000002044
2. Masters K. Artificial intelligence in medical education. Med Teach. 2019;41(9):976-80. doi:10.1080/0142159X.2019.1595557
3. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: Integrative review. JMIR Med Educ. 2019;5(1). doi:10.2196/13930
4. McCoy LG, Nagaraj S, Morgado F, Harish V, Das S, Celi LA. What do medical students actually need to know about artificial intelligence? NPJ Digit Med. 2020;3:86. doi:10.1038/s41746-020-0294-7
5. Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019;4:5. doi:10.1186/s41073-019-0064-8
6. Snyder H. Literature review as a research methodology: An overview and guidelines. J Bus Res. 2019;104:333-9. doi:10.1016/j.jbusres.2019.07.039
7. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ. 2021;372. doi:10.1136/bmj.n71
8. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med. 2018;169(7):467-73. doi:10.7326/M18-0850
9. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, et al. PRISMA-S: An extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev. 2021;10(1):39. doi:10.1186/s13643-020-01542-z
10. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: An efficient and complete method to develop literature searches. J Med Libr Assoc. 2018;106(4):531-41. doi:10.5195/jmla.2018.283
11. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358. doi:10.1136/bmj.j4008
12. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. BMJ. 2020;368. doi:10.1136/bmj.l6890
13. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: A modified e-Delphi study. J Clin Epidemiol. 2019;111:49-59.e1. doi:10.1016/j.jclinepi.2019.03.008
14. 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
15. Culp ML, Mahmoud S, Liu D, Haworth IS. An artificial intelligence-supported medicinal chemistry project: An example for incorporating artificial intelligence within the pharmacy curriculum. Am J Pharm Educ. 2024;88(5):100696. doi:10.1016/j.ajpe.2024.100696
16. 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
17. Basalelah L, Caldas LM, Crawford AN, Wijesinghe D, Donohoe KL. Student pharmacists' experience with using artificial intelligence for pharmacogenetic counseling in a skills laboratory course. Am J Pharm Educ. 2025;89(8):101436. doi:10.1016/j.ajpe.2025.101436
18. 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
19. Sourial M, Hagler JC. A pilot study using artificial intelligence to enhance efficiency, accuracy, and objectivity in grading pharmacy objective structured clinical examinations. Am J Pharm Educ. 2025;89(8):101455. doi:10.1016/j.ajpe.2025.101455
20. Russell RG, Lovett Novak L, Patel M, Garvey KV, Craig KJT, Jackson GP, et al. Competencies for the use of artificial intelligence-based tools by health care professionals. Acad Med. 2023;98(3):348-56. doi:10.1097/ACM.0000000000004963
21. Garvey KV, Thomas Craig KJ, Russell R, Novak LL, Moore D, Miller BM. Considering clinician competencies for the implementation of artificial intelligence-based tools in health care: Findings from a scoping review. JMIR Med Inform. 2022;10(11). doi:10.2196/37478
22. Weidener L, Fischer M. Proposing a principle-based approach for teaching AI ethics in medical education. JMIR Med Educ. 2024;10. doi:10.2196/55368
23. Charow R, Jeyakumar T, Younus S, Dolatabadi E, Salhia M, Al-Mouaswas D, et al. Artificial intelligence education programs for health care professionals: Scoping review. JMIR Med Educ. 2021;7(4). doi:10.2196/31043
24. Paranjape K, Schinkel M, Nannan Panday R, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. JMIR Med Educ. 2019;5(2). doi:10.2196/16048
25. Jin Y, Martinez-Maldonado R, Gašević D, Yan L. GLAT: The generative AI literacy assessment test. Comput Educ Artif Intell. 2025;9:100436. doi:10.1016/j.caeai.2025.100436
26. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: Exploratory qualitative study of UK general practitioners' views. J Med Internet Res. 2019;21(3). doi:10.2196/12802

27. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. J Consum Res. 2019;46(4):629-50. doi:10.1093/jcr/ucz013
28. Meyer AND, Giardina TD, Spitzmueller C, Shahid U, Scott TMT, Singh H. Patient perspectives on the usefulness of an artificial intelligence-assisted symptom checker: Cross-sectional survey study. J Med Internet Res. 2020;22(1). doi:10.2196/14679
29. 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
30. Lee YM, Kim S, Lee YH, Kim HS, Seo SW, Kim H, et al. Defining medical artificial intelligence competencies for medical school graduates: Outcomes of a Delphi survey and medical student/educator questionnaire of South Korean medical schools. Acad Med. 2024;99(5):524-33. doi:10.1097/ACM.0000000000005618
31. Car J, Ong QC, Erlikh Fox T, Leightley D, Kemp SJ, Švab I, et al. The Digital Health Competencies in Medical Education Framework: An international consensus statement based on a Delphi study. JAMA Netw Open. 2025;8(1). doi:10.1001/jamanetworkopen.2024.53131
32. Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11). doi:10.1016/S2589-7500(21)00208-9
33. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: A roadmap for responsible machine learning for health care. Nat Med. 2019;25(9):1337-40. doi:10.1038/s41591-019-0548-6
34. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med. 2023;388(13):1233-9. doi:10.1056/NEJMsr2214184