ARTIFICIAL INTELLIGENCE IN CHEMISTRY EDUCATION: OPPORTUNITIES, ETHICAL CHALLENGES, AND THE DEVELOPMENT OF CRITICAL THINKING IN SECONDARY AND VOCATIONAL SETTINGS

Xafiza Rustamova
Chirchik City Technicum №. 1, Chirchik, Tashkent Region, Uzbekistan
Email: h.namazovna2020@gmail.com

ABSTRACT

The rapid integration of artificial intelligence (AI) tools is transforming chemistry education at secondary and vocational levels. This review synthesizes recent empirical studies and systematic reviews (primarily 2023–2025) on AI applications in chemistry teaching, including large language model (LLM) chatbots, adaptive learning platforms, and virtual laboratory simulations. Using the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) as the primary normative framework, the article examines the demonstrated benefits of AI—personalized feedback, enhanced molecular visualization, improved student self-efficacy, and scalable differentiated instruction—alongside persistent limitations, including computational errors in stoichiometric and redox calculations (with reported mechanistic accuracy as low as 28% for early models), hallucinated citations, and algorithmic bias. Six concrete ethical dilemmas specific to chemistry education are analyzed through detailed scenarios: data privacy and learner profiling, academic integrity in AI-assisted laboratory reports, algorithmic bias in automated assessment, the virtual–real competency gap in vocational training, the opacity of AI-driven assessment decisions, and digital inequality. The article argues that AI's pedagogical effectiveness is inseparable from its capacity to foster critical thinking in both students and educators. A practical four-pillar framework for responsible implementation is proposed, with implications for curriculum design, teacher professional development, and institutional policy. Directions for future empirical research in secondary and vocational contexts, particularly in resource-constrained settings, are identified.

Keywords: artificial intelligence, chemistry education, critical thinking, virtual laboratories, generative AI, large language models, ethics in education, secondary education, vocational education, AI literacy

INTRODUCTION

Artificial intelligence has emerged as one of the most consequential technologies in contemporary chemistry. The recognition of AI among IUPAC's *Top 10 Emerging Technologies in Chemistry* in 2023[1] and the awarding of the 2024 Nobel Prize in

Chemistry for the development of AlphaFold and its contributions to protein structure prediction[2] have demonstrated unequivocally that AI is no longer a peripheral tool but a central force in chemical research. These advances carry immediate implications for how chemistry is taught and learned.

The landscape of AI applications in chemistry education has expanded dramatically in recent years. A systematic review by Erümit and Sarıalioğlu, covering 142 publications from 2014 to 2024, documented a sharp increase in studies on AI in science and chemistry education after 2021, with ChatGPT and conversational agents emerging as the most frequently investigated tools and learning outcomes identified as the primary dependent variable.[3] A parallel critical review by Iyamuremye et al., synthesizing 62 articles from Scopus, Web of Science, ERIC, and Google Scholar, confirmed that while AI and machine learning remain in early stages of classroom implementation, they hold significant potential to transform chemistry teaching and learning through intelligent tutoring, automated assessment, and adaptive feedback systems.[4]

Chemistry occupies a distinctive position among the natural sciences in its demand for reasoning across multiple representational levels. Johnstone’s seminal framework identifies three domains—macroscopic (observable phenomena), submicroscopic (particulate models), and symbolic (formulae, equations)—that learners must navigate simultaneously.[5] This inherent complexity makes chemistry both particularly amenable to AI-assisted visualization and scaffolding, and particularly vulnerable to AI-generated errors. When a language model produces a plausible but chemically incorrect mechanistic explanation, the error can propagate through a student’s conceptual understanding in subtle and persistent ways, precisely because the AI’s response may exhibit high levels of explanation sophistication while being factually inaccurate.[6] Furthermore, the experimental character of chemistry introduces practical skills and safety considerations that cannot be fully replicated by AI, creating a distinctive tension between virtual and hands-on learning.

At the international policy level, the UNESCO Recommendation on the Ethics of Artificial Intelligence, adopted by 193 Member States in November 2021, established the first global normative framework for responsible AI development and use. The Recommendation identifies education as a priority domain, noting that “living in digitalizing societies requires new educational practices, ethical reflection, critical thinking, responsible design practices and new skills.”[7] This human-centric approach—grounded in the principles of transparency, fairness, privacy, accountability, and human autonomy—provides a valuable lens through which to evaluate the integration of AI into chemistry classrooms.

This article pursues three objectives: (1) to review the current evidence on AI applications in secondary (grades 9–11 or equivalent) and vocational chemistry education, including virtual laboratories, LLM chatbots, and adaptive platforms; (2) to critically examine the ethical challenges these applications present, analyzed through concrete classroom scenarios; and (3) to argue that the development of critical thinking—in both students and educators—is not merely a desirable outcome of AI integration but a prerequisite for its responsible implementation. The article concludes with a practical framework for embedding AI in chemistry curricula while safeguarding educational integrity and equity.

REVIEW METHODOLOGY

Search Strategy and Databases

This review synthesizes findings from peer-reviewed empirical studies, systematic reviews, and international policy documents. Literature searches were conducted in three databases—Web of Science Core Collection, Scopus, and Google Scholar—between January and March 2025, covering publications from January 2021 to March 2025. Search queries combined the terms "artificial intelligence" OR "generative AI" OR "ChatGPT" OR "large language model" with "chemistry education" OR "science education", further refined with "virtual laboratory," "ethics," or "critical thinking" where applicable.

Inclusion and Exclusion Criteria

Articles were included if they: (a) reported empirical data or conducted a systematic/scoping review on AI use in chemistry or science education; (b) were published in English in a peer-reviewed journal; and (c) addressed at least one of the following themes: learning outcomes, ethical considerations, or critical thinking development. Articles were excluded if they focused exclusively on AI in chemistry research (drug discovery, materials science) without educational applications, or if they were editorial commentaries shorter than three pages. The UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) was included as the primary normative framework despite not being a peer-reviewed study, given its unique international standing as the only global standard on AI ethics.

Selection Process and Corpus

Initial searches yielded 247 records. After removing duplicates (n = 58) and screening titles and abstracts for relevance, 112 articles were assessed for full-text eligibility. Of these, 43 met all inclusion criteria and form the primary corpus of this review. Priority was given to studies published in the *Journal of Chemical Education* (n = 22), *Discover Education* (n = 4), *Chemistry Education Research and Practice* (n = 3), and *Frontiers in Education* (n = 3). The remaining articles appeared in *Computers and Education: Artificial Intelligence*, *JMIR Formative Research*, and other relevant venues.

Scope and Limitations of Extrapolation

The target population of this review is secondary education (grades 9–12 or equivalent) and vocational/technical education (community colleges, technical schools, specialized chemistry training programs). Because the majority of published research originates from higher education settings—particularly undergraduate chemistry courses at research universities—findings are extrapolated to secondary and vocational contexts with explicit caution. Where studies involved high-school populations or vocational programs, this is noted. The author's professional experience in a vocational chemistry program (Chirchik City Technicum No. 1, Uzbekistan) informs the interpretation of findings and the formulation of recommendations, though systematic collection of original empirical data from this context was not undertaken in the present work.

AI APPLICATIONS IN CHEMISTRY EDUCATION

Virtual Laboratory Simulations

Virtual laboratory platforms represent the most mature application of AI-enhanced technology in chemistry education. Platforms such as Labster and PhET Interactive Simulations integrate adaptive feedback mechanisms that adjust simulation difficulty and scaffolding based on student performance in real time. The evidence base for their effectiveness has grown substantially.

A systematic literature review of virtual chemical laboratories found that virtual labs are more effective than passive teaching methods (lectures, textbooks, video) and show comparable or slightly superior effectiveness to hands-on laboratories, with the strongest learning outcomes observed when virtual and physical methods are combined.[8] Recent quasi-experimental work corroborates these findings. Teferi et al. (2024) compared three instructional conditions—real laboratory ($n = 20$), lecture-only ($n = 20$), and virtual laboratory ($n = 20$)—among undergraduate chemistry students at Dilla University (Ethiopia). Post-test mean scores were statistically comparable between real laboratory ($M = 67.7$, $SD = 11.1$) and virtual laboratory ($M = 66.1$, $SD = 12.6$) groups, with both significantly outperforming the lecture-only group ($M = 51.1$, $SD = 9.9$; $p < 0.05$). The 95% confidence intervals for the real and virtual laboratory groups overlapped substantially, supporting the conclusion that virtual laboratories can achieve outcomes equivalent to physical laboratories under controlled conditions.[9]

A pioneering study by Lizano-Sánchez et al. (2025) analyzed students' interactions with a GPT-4o–powered intelligent assistant integrated into a remote acid–base titration laboratory. Content analysis of student–AI interactions revealed four principal dimensions of use: experimental development (procedural guidance), data processing (calculation assistance), conceptual understanding (theoretical explanations), and report preparation. This study represents an emerging paradigm in which AI is not merely a simulation engine but an interactive pedagogical agent embedded within the experimental workflow, available around the clock to provide individualized support.[10]

For secondary and vocational education contexts—where laboratory equipment may be limited, reagents expensive, or safety constraints prohibitive—these findings carry particular significance. Virtual simulations can democratize access to experimental learning experiences, provided that they are implemented as complements to, rather than replacements for, hands-on work.

Large Language Model Chatbots

Generative AI chatbots, particularly OpenAI's ChatGPT, have become the most widely studied AI tools in chemistry education research since 2023.[3] The evidence reveals a nuanced picture of improving but still limited capabilities.

Tyson (2023) provided an early systematic assessment of ChatGPT's shortcomings, documenting that the model cannot perform mathematical operations reliably, makes conceptual errors in general education chemistry, and fabricates plausible-looking citations whose purported contents are only partly accurate.[11] Leon and Vidhani (2023) assessed ChatGPT's performance on organic chemistry tasks and concluded that the model itself requires a "chemistry tutor," struggling with mechanistic reasoning and reaction prediction.[12]

A study by Yik and Dood (2024) provided the most detailed analysis of AI's reasoning in organic chemistry to date. They found that only 28% of ChatGPT-3.5's explanations of organic reaction mechanisms were fully accurate, while the majority contained predominantly accurate descriptions of chemical phenomena accompanied by subtle errors—crucially, presented with high levels of explanation sophistication that could mislead students into accepting incorrect reasoning. Prompt engineering significantly enhanced explanation quality but did not improve mechanism accuracy, suggesting a fundamental limitation of current language models in chemical reasoning.[6]

Clark et al. (2023) compared the performance of college chemistry students and ChatGPT on acid–base calculations, finding that the model made characteristic computational errors distinct from those made by students.[13] Ruff et al. (2024) tasked upper-level analytical and inorganic chemistry students with using ChatGPT for method development and green chemistry analysis. In only 1 of 24 assignments did ChatGPT provide a specific, real, and useful literature reference; in other cases, it generated completely fabricated citations, including fake DOI numbers.[14]

Despite these limitations, constructive pedagogical applications have been demonstrated. Exintaris et al. (2023) developed the SMARTCHEMPer workshop, in which AI-generated responses to pharmaceutical chemistry problems served as prompts for student critique. Students analyzed AI outputs for errors, evaluated explanation quality, and produced improved responses, effectively transforming AI limitations into metacognitive learning opportunities.[15] Guo and Lee (2023) similarly showed that structured activities requiring students to evaluate and revise AI-generated chemistry content enhanced critical thinking skills.[16] At Penn State University, an intervention in Fall 2024 demonstrated that structured, ethically framed AI integration in honors general chemistry increased students' frequency of AI use, academic confidence, and perceived utility of AI while reducing anxiety about ethical implications.[17]

Adaptive Learning Platforms and Prompt Engineering

AI-powered adaptive learning systems offer personalized learning trajectories by adjusting task difficulty, feedback specificity, and content sequencing based on individual student performance. In chemistry, these systems can tailor problems in stoichiometry, thermodynamics, and organic reaction mechanisms to a student's demonstrated level of understanding, creating differentiated instruction at scale—a capability particularly valuable in vocational programs with heterogeneous student populations.[4,18]

An underexplored but critical dimension of effective AI use is prompt engineering—the skill of formulating effective queries for AI systems. Heller (2024) investigated prompting strategies among chemistry students and found a prevalent reliance on copy-pasting tactics in initial interactions with generative AI, with students requiring explicit guidance to improve their prompting abilities. After introducing the "Five S" prompting framework (Set the scene, Be specific, Simplify language, Structure the output, Share feedback), students reported significantly higher satisfaction with AI-generated responses.[19] This finding has direct pedagogical implications: teaching students to formulate precise, well-structured prompts is itself an exercise in scientific reasoning and metacognition. When a student learns to specify the level of detail, the representational format, and the reasoning framework they expect from an AI system, they must first clarify their own understanding of the chemical concept in question.

ETHICAL CHALLENGES: ANALYSIS OF CONCRETE DILEMMAS

The UNESCO Recommendation on the Ethics of AI (2021) establishes foundational principles for the responsible deployment of AI across all domains, including education: transparency and explainability, fairness and non-discrimination, privacy protection, accountability, and human autonomy.[7] In chemistry education, these abstract principles translate into concrete dilemmas that teachers, students, and administrators face daily. Six such dilemmas are analyzed below, each grounded in a realistic classroom scenario and mapped to the relevant UNESCO principle.

Dilemma 1: Data Privacy and Learner Profiling

Scenario. A ninth-grade student works through an acid–base titration simulation on an AI-enhanced platform. The system records response times, error patterns, the sequence of decisions at each procedural step, and behavioral indicators of frustration. The AI uses this data to recommend remedial exercises—a clear pedagogical benefit.

However, the accumulated data forms a detailed "learner profile" that persists in the platform's database. Who controls this profile? Can the platform developer use it for model training? Could it follow the student into future educational or employment contexts?

This dilemma arises from the tension between the UNESCO principle of proportionality (data collection should be limited to what is necessary for the stated purpose) and the business models of commercial educational technology platforms, which often depend on large-scale data aggregation. The risk is amplified in secondary education, where students are minors with limited capacity to provide informed consent. Practical resolution requires institutional data governance policies that specify data retention periods, access controls, and prohibitions on secondary use.

Dilemma 2: Academic Integrity and AI-Assisted Authorship

Scenario. After completing a virtual titration simulation, a student asks ChatGPT to generate a complete laboratory report—purpose, procedure, calculations, error analysis, and conclusions—and submits it with minor rephrasing. The teacher suspects AI involvement but cannot confirm it.

West et al. (2023) conducted a systematic analysis of AI-generated laboratory reports across the chemistry curriculum and found that student perceptions of AI-assisted writing as ethically problematic were declining: many students no longer viewed submission of lightly edited AI outputs without attribution as a form of academic dishonesty.[20] This normalization is compounded by the unreliability of AI detection tools, which produce elevated false-positive rates for non-native English speakers[21]—a finding of particular relevance in multilingual educational contexts such as those in Central Asia. Cooper (2026) has proposed that rather than relying on detection, educators should design assessments that require demonstrated understanding through explanation, revision, and reflection.[22]

Dilemma 3: Algorithmic Bias in Assessment

Scenario. An adaptive chemistry platform uses AI to evaluate student descriptions of molecular-level processes. Students from certain linguistic backgrounds consistently receive lower scores because their phrasing deviates from the "standard" patterns in the training data, despite demonstrating equivalent conceptual understanding. In a parallel scenario, the system's recommendation algorithm steers female students away from advanced chemistry modules based on correlations in historical enrollment data.

These scenarios instantiate the UNESCO principle of fairness and non-discrimination. Estrada et al. (2025) found that inclusivity and accessibility remain significantly underrepresented in the academic literature on AI in chemistry education, despite growing recognition of digital equity as a critical concern.[23] Algorithmic bias is particularly insidious because it operates invisibly, reinforcing existing inequalities under the guise of objective assessment.

Dilemma 4: The Virtual–Real Competency Gap

Scenario. Students in a vocational analytical chemistry program complete all titration exercises through virtual simulations with AI-guided feedback. Their assessment scores are excellent. However, when placed in a real laboratory for practicum, they cannot assemble physical apparatus, handle reagents safely, or interpret unexpected results without algorithmic prompting. The instructor faces a choice: continue the safe and convenient virtual model, or return students to physical laboratories for which they are now poorly prepared.

This dilemma is unique to chemistry and other laboratory sciences. The systematic review by Faulconer et al. noted that skill-based outcomes remain underresearched in the virtual laboratory literature[8]—a gap that is especially consequential for vocational programs, where graduates are expected to possess immediately applicable practical competencies. In the author's experience at a vocational chemistry program, the virtual–real gap is observable when students who perform flawlessly in simulations struggle with the unpredictability of real chemical systems—a reagent that has absorbed moisture, a burette with an air bubble, a color change that is less distinct than the simulation depicted.

Dilemma 5: Transparency and the "Black Box" Problem

Scenario. An AI system embedded in a virtual laboratory automatically adjusts the difficulty of a chemical kinetics simulation and assigns a grade. A student receives "unsatisfactory" without a comprehensible explanation. The teacher cannot reconstruct the assessment logic to respond to the student's appeal or the parents' inquiry.

The UNESCO Recommendation explicitly requires that AI systems be transparent and explainable. Yet many educational AI platforms fail to meet this standard. The opacity of algorithmic decision-making undermines the pedagogical principle that effective feedback must be comprehensible and actionable, and it transfers assessment authority from the teacher—who bears professional and legal responsibility—to an unaccountable algorithm.[7]

Dilemma 6: Digital Inequality

Scenario. An urban school integrates Labster, ChatGPT, and adaptive platforms into its chemistry curriculum, providing students with rich, personalized learning experiences. A rural school in the same country, with intermittent internet connectivity and outdated hardware, cannot access these tools. Students from the two schools sit the same national examination.

This dilemma reflects a structural reality: without deliberate equity-oriented policy, AI in education can widen rather than narrow access gaps. A combination of commercial and open-access platforms (e.g., Labster alongside PhET Interactive Simulations), provision of offline alternatives, and investment in digital infrastructure is needed to prevent AI from becoming a vector of inequality.[4,23]

AI AS A CATALYST FOR CRITICAL THINKING DEVELOPMENT

In Students

Paradoxically, the limitations of AI serve as powerful pedagogical tools. When a language model generates a chemically incorrect mechanism, fabricates a citation, or provides a stoichiometric calculation that yields an implausible result, these errors create authentic opportunities for students to practice verification, source evaluation, and scientific reasoning—skills central to chemical literacy.

The SMARTCHEMPer workshop by Exintaris et al. (2023) provides an empirically validated model. Students received AI-generated responses to pharmaceutical chemistry problems and were tasked with identifying errors in reasoning, evaluating the quality of explanations, and producing improved responses grounded in authoritative sources. The authors report that this approach effectively developed both metacognition and critical thinking skills.[15] Ruff et al. (2024) employed a complementary strategy: students were required to evaluate whether analytical methods suggested by ChatGPT actually existed, verify any provided references, and compare the greenness of AI-suggested protocols against established methods using MSDS data. In many cases, students discovered that the AI's references were fabricated and its methods vague—transforming the assignment into a rigorous exercise in scientific literature evaluation.[14]

Prompt engineering itself functions as a metacognitive exercise. When students learn to specify their queries precisely within the "Five S" framework[19]—setting the context, specifying the level of detail, simplifying language, structuring the expected output, and sharing feedback—they engage in a process that requires prior reflection on what they know, what they need to know, and how to evaluate the quality of the answer. In chemistry, formulating an effective prompt about the mechanism of nucleophilic substitution requires the student to first articulate their understanding of electrophilicity, leaving group ability, and solvent effects.

In Educators

The impact of AI on educator professional development has received less attention but is no less important. Recent qualitative research provides emerging evidence. Blonder and Feldman-Maggor (2024) published a study in *Chemistry Teacher International* examining responsible AI use and ethical considerations for chemistry teaching, arguing that teachers who engage critically with AI develop a more reflective and explicit understanding of expert reasoning.[24] A qualitative interview study with chemistry faculty at a U.S. university, conducted during Fall 2024 and published in *Discover Education*, found that systematic engagement with AI tools heightened instructors' awareness of algorithmic boundaries and encouraged reflexive pedagogical practice.[25]

Three dimensions of critical thinking development in educators emerge from this literature. *First, habitual verification:* regular interaction with LLMs exposes characteristic error patterns—incorrect equilibrium constants, confused reaction mechanisms, fabricated references—training educators to adopt systematic skepticism toward all information sources, not only AI. *Second, boundary awareness:* experienced chemistry teachers learn to identify the types of questions where AI performs reliably (standard stoichiometric calculations, general concept overviews, generation of practice problems) and those where it is unreliable (regioselectivity predictions, spectral data interpretation, non-standard kinetic situations, literature citation). This calibrated understanding of tool capabilities is a hallmark of professional expertise. *Third, reflexive practice:* the process of formulating effective prompts forces educators to articulate their knowledge structures explicitly, revealing implicit assumptions, gaps, and areas where their own understanding could be deeper.

PRACTICAL RECOMMENDATIONS

Guided by the UNESCO Recommendation on the Ethics of AI[7] and the empirical evidence reviewed above, the following framework offers actionable strategies for responsible AI integration in secondary and vocational chemistry education. Recommendations are organized into four complementary pillars, each linked to specific ethical principles and supported by evidence from the reviewed literature.

Pillar 1: Ensuring Transparency and Responsible Disclosure

Both students and teachers should explicitly attribute AI use in all submissions—laboratory reports, assignments, presentations, and teaching materials. A model attribution statement might read: "The initial draft of the procedure and calculations was generated using GPT-4o and subsequently revised and verified by the author." Transparency normalizes AI as a tool while preserving the expectation of intellectual honesty. This recommendation aligns with emerging international practice and addresses the normalization of unattributed AI use documented by West et al.[20]

Pillar 2: Developing Critical Verification Skills and AI Literacy

Every assignment involving AI-generated content should include a mandatory critical evaluation component. In chemistry, this might involve: comparing an AI's explanation of Le Chatelier's Principle or nucleophilic substitution with textbook sources and identifying discrepancies; cross-verifying virtual titration data against stoichiometric principles and propagated measurement uncertainty; or evaluating AI-suggested analytical methods against green chemistry criteria using MSDS data, following the model of Ruff et al.[14]

AI literacy should be embedded as a curricular module rather than treated as an ad hoc supplement. Topics should include: the nature and causes of LLM hallucinations, with chemistry-specific examples; algorithmic bias and its consequences for equity; data privacy and digital hygiene; the distinction between AI-assisted learning and AI-dependent learning; and prompt engineering as a metacognitive skill.[19]

Pillar 3: Hybrid Integration of Virtual and Hands-On Learning

The evidence consistently shows that the strongest learning outcomes arise from combining virtual and physical laboratory experiences.[8,9] Virtual simulations should prepare for, not replace, hands-on work. A recommended instructional sequence for a topic such as acid–base titration or gas law experiments is: (1) AI-supported pre-laboratory preparation (concept review, safety briefing, prediction of expected results); (2) adaptive virtual simulation (with AI feedback and scaffolding); (3) hands-on experiment in a physical laboratory (where feasible); and (4) AI-assisted data analysis with mandatory critical reflection on discrepancies between virtual and real results. This hybrid approach is especially important in vocational programs to prevent the virtual–real competency gap identified in Dilemma 4.

Pillar 4: Institutional Policies, Equity, and Human Oversight

Educational institutions should develop clear, differentiated policies that distinguish permissible AI uses (brainstorming, concept exploration, draft feedback, practice problem generation), conditional uses (with explicit attribution and instructor approval), and prohibited uses (unattributed submission of AI-generated work in summative assessments). These policies should be developed collaboratively with input from educators, students, and administrators, and revised regularly as AI capabilities evolve.

Equity requires a deliberate strategy: combining commercial platforms (Labster) with open-access tools (PhET Interactive Simulations), providing offline and low-bandwidth alternatives for resource-constrained settings, and investing in teacher training so that pedagogical quality does not depend on platform sophistication. Throughout, teachers must retain final responsibility for all assessment decisions and curricular content. AI-generated grades, recommendations, and feedback should always be subject to professional pedagogical review.

DIRECTIONS FOR FUTURE RESEARCH

Several significant gaps emerge from this review that warrant attention from the chemistry education research community.

First, the vast majority of published studies on AI in chemistry education originate from higher education settings in North America, Europe, and East Asia. Research in secondary and vocational contexts—and in the Global South more broadly—remains sparse. Empirical studies investigating the effectiveness of AI-enhanced chemistry instruction in resource-constrained settings, where the equity considerations are most acute, are urgently needed.

Second, longitudinal studies tracking the impact of AI integration on students' conceptual understanding, practical laboratory skills, and critical thinking dispositions over multiple semesters or academic years are virtually absent. Cross-sectional designs predominate, limiting our ability to assess lasting effects.

Third, the relationship between prompt engineering skills and chemistry learning outcomes has been insufficiently studied. While Heller (2024) demonstrated that structured prompting frameworks improve student satisfaction,[19] the question of whether prompt engineering training transfers to broader scientific reasoning skills remains open.

Fourth, comparative studies of different AI integration models—fully virtual vs. hybrid vs. AI-augmented traditional instruction—with attention to both cognitive and affective outcomes would provide the evidence base needed for informed policy decisions. Such studies should include vocational populations and measure practical competencies, not only conceptual understanding.

Finally, the development and validation of instruments for assessing AI literacy in chemistry education contexts would provide a needed measurement foundation for future research.

CONCLUSIONS

The integration of artificial intelligence into chemistry education at the secondary and vocational levels presents both unprecedented opportunities and significant challenges. The evidence reviewed in this article demonstrates that AI tools—from virtual laboratories to generative chatbots to adaptive learning platforms—can meaningfully enhance the teaching and learning of chemistry when implemented with pedagogical intentionality and ethical awareness.

However, the same evidence reveals persistent limitations: mechanistic accuracy as low as 28% for early LLM models,[6] systematic fabrication of citations,[11,14] algorithmic biases that may exacerbate educational inequity,[23] and a growing tension between virtual proficiency and practical competency. These are not merely technical problems awaiting engineering solutions; they are pedagogical and ethical challenges that require human judgment, institutional commitment, and continuous critical reflection.

Perhaps the most consequential insight from this review is that the pedagogical value of AI in chemistry education is inseparable from the development of critical thinking. AI's imperfections make it a uniquely effective tool for teaching verification, source evaluation, and scientific reasoning—provided that educators design learning experiences that foreground these skills rather than treating AI as an infallible oracle. For educators, systematic engagement with AI tools catalyzes professional development by making explicit the boundaries of expert knowledge and the structure of disciplinary reasoning.

The UNESCO Recommendation provides the normative compass: AI should amplify human capabilities without undermining human autonomy.[7] In chemistry education, this principle demands that AI serve as an instrument in the hands of a reflective teacher and a critically engaged student. The four-pillar framework proposed here—transparency, critical verification, hybrid integration, and institutional governance—offers a practical starting point. Its implementation will require sustained collaboration among educators, researchers, policymakers, and technology developers, guided by the conviction that the purpose of education is not the efficient delivery of information but the development of scientifically literate, ethically aware, and critically thinking individuals.

REFERENCES

(1) IUPAC. Top 10 Emerging Technologies in Chemistry 2023. Chem. Int. 2023, 45 (3), 14–21. DOI: 10.1515/ci-2023-0303.

(2) Jumper, J.; Evans, R.; Pritzel, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. Nature 2021, 596 (7873), 583–589. DOI: 10.1038/s41586-021-03819-2.

(3) Erümit, A. K.; Sarıalioğlu, R. Ö. Artificial Intelligence in Science and Chemistry Education: A Systematic Review. Discover Education 2025, 4 (1), 178. DOI: 10.1007/s44217-025-00622-3.

(4) Iyamuremye, A.; Niyonzima, F. N.; Mukiza, J.; Twagilimana, I.; Nyirahabimana, P.; Nsengimana, T.; Habiyaremye, J. D.; Habimana, O.; Nsabayezu, E. Utilization of Artificial Intelligence and Machine Learning in Chemistry Education: A Critical Review. Discover Education 2024, 3 (1), 95. DOI: 10.1007/s44217-024-00197-5.

(5) Johnstone, A. H. Why Is Science Difficult to Learn? Things Are Seldom What They Seem. J. Comput. Assisted Learn. 1991, 7 (2), 75–83. DOI: 10.1111/j.1365-2729.1991.tb00230.x.

(6) Yik, B. J.; Dood, A. J. ChatGPT Convincingly Explains Organic Chemistry Reaction Mechanisms Slightly Inaccurately with High Levels of Explanation Sophistication. J. Chem. Educ. 2024, 101 (5), 1836–1846. DOI: 10.1021/acs.jchemed.4c00235.

(7) UNESCO. Recommendation on the Ethics of Artificial Intelligence; UNESCO: Paris, 2022. Adopted on 23 November 2021.

(8) Faulconer, E.; Poore, J.; Fernandez, C.; Griffith, J. C. Virtual Chemical Laboratories: A Systematic Literature Review of Research, Technologies and Instructional Design. Comput. Educ. Open 2022, 2, 100053. DOI: 10.1016/j.caeo.2021.100053.

(9) Teferi, M. Y.; Negash, M. Y.; Diribsa, A. B. The Effect of Virtual Laboratories on the Academic Achievement of Undergraduate Chemistry Students: Quasi-Experimental Study. JMIR Form. Res. 2024, 8, e64476. DOI: 10.2196/64476.

(10) Lizano-Sánchez, F.; Arias-Navarro, A. M.; Orduña, P.; Dziabenko, O.; de La Cruz-García, J. S. Students' Interactions with an Artificial Intelligence Assistant in a

Remote Chemistry Laboratory. Front. Educ. 2025, 10, 1712743. DOI: 10.3389/feduc.2025.1712743.

(11) Tyson, J. Shortcomings of ChatGPT. J. Chem. Educ. 2023, 100 (8), 3098–3101. DOI: 10.1021/acs.jchemed.3c00361.

(12) Leon, A. J.; Vidhani, D. ChatGPT Needs a Chemistry Tutor Too. J. Chem. Educ. 2023, 100 (10), 3859–3865. DOI: 10.1021/acs.jchemed.3c00288.

(13) Clark, T. M.; Anderson, E.; Dickson-Karn, N. M.; Soltanirad, C.; Tafini, N. Comparing the Performance of College Chemistry Students with ChatGPT for Calculations Involving Acids and Bases. J. Chem. Educ. 2023, 100 (10), 3934–3944. DOI: 10.1021/acs.jchemed.3c00500.

(14) Ruff, E. F.; Franz, J. L.; West, J. K. Using ChatGPT for Method Development and Green Chemistry Education in Upper-Level Laboratory Courses. J. Chem. Educ. 2024, 101 (8), 3224–3232. DOI: 10.1021/acs.jchemed.4c00193.

(15) Exintaris, B.; Karunaratne, N.; Yuriev, E. Metacognition and Critical Thinking: Using ChatGPT-Generated Responses as Prompts for Critique in a Problem-Solving Workshop (SMARTCHEMPer). J. Chem. Educ. 2023, 100 (8), 2972–2980. DOI: 10.1021/acs.jchemed.3c00481.

(16) Guo, Y.; Lee, D. Leveraging ChatGPT for Enhancing Critical Thinking Skills. J. Chem. Educ. 2023, 100 (12), 4876–4883. DOI: 10.1021/acs.jchemed.3c00505.

(17) [Authors]. Using Exam Preparation and Reflection to Introduce Artificial Intelligence Tools in Honors General Chemistry. J. Chem. Educ. 2025, 102 (10), 4470–4478. DOI: 10.1021/acs.jchemed.5c00618.

(18) Alli, A. A. A Review of the Potential of Artificial Intelligence (AI) in Enhancing Chemistry Education. Int. J. Recent Innovations Acad. Res. 2025, 9, 31–43.

(19) Heller, S. Assessment of Students Use of Generative Artificial Intelligence: Prompting Strategies and Prompt Engineering in Chemistry Education. J. Chem. Educ. 2024, 101 (6), 2475–2482. DOI: 10.1021/acs.jchemed.4c00212.

(20) West, J. K.; Franz, J. L.; Hein, S. M.; Leverentz-Culp, H. R.; Mauser, J. F.; Ruff, E. F.; Zemke, J. M. An Analysis of AI-Generated Laboratory Reports across the Chemistry Curriculum and Student Perceptions of ChatGPT. J. Chem. Educ. 2023, 100 (11), 4351–4359. DOI: 10.1021/acs.jchemed.3c00581.

(21) Emenike, M. E.; Emenike, B. U. Was This Title Generated by ChatGPT? Considerations for Artificial Intelligence Text-Generation Software Programs for Chemists and Chemistry Educators. J. Chem. Educ. 2023, 100 (4), 1413–1418. DOI: 10.1021/acs.jchemed.3c00063.

(22) Cooper, M. M. A Preliminary Set of Principles to Support Learning in the Context of Generative AI. J. Chem. Educ. 2026, 103 (1), 36–42. DOI:

10.1021/acs.jchemed.5c01413.

(23) Estrada, L.; et al. Artificial Intelligence and Inclusive Education. Educ. Process: Int. J. 2025, 17 (380). DOI: 10.22521/edupij.2025.17.380.

(24) Blonder, R.; Feldman-Maggor, Y. AI for Chemistry Teaching: Responsible AI and Ethical Considerations. Chem. Teach. Int. 2024, 6, 385. DOI: 10.1515/cti-2024-0014.

(25) [Authors]. Chemistry Faculty Perspectives on Artificial Intelligence and Pedagogical Change in Higher Education. Discover Education 2026. DOI: 10.1007/s44217-026-01250-1.

(26) Alasadi, E. A.; Baiz, C. R. Generative AI in Education and Research: Opportunities, Concerns, and Solutions. J. Chem. Educ. 2023, 100 (8), 2965–2971. DOI: 10.1021/acs.jchemed.3c00323.

(27) Huwer, J.; et al. AI in Chemistry and Chemical Education. Int. J. Phys. Chem. Educ. 2025, 17 (1), 1–4. DOI: 10.51724/ijpce.v17i1.403.

(28) Kasneci, E.; Sessler, K.; Küchemann, S.; et al. ChatGPT for Good? On Opportunities and Challenges of Large Language Models for Education. Learn. Individ. Differ. 2023, 103, 102274. DOI: 10.1016/j.lindif.2023.102274.

(29) Farrokhnia, M.; Banihashem, S. K.; Noroozi, O.; Wals, A. A SWOT Analysis of ChatGPT: Implications for Educational Practice and Research. Innov. Educ. Teach. Int. 2024, 61, 460–474. DOI: 10.1080/14703297.2023.2195846.

(30) Mendez, J. D. Student Perceptions of Artificial Intelligence Utility in the Introductory Chemistry Classroom. J. Chem. Educ. 2024, 101 (8), 3547–3549. DOI: 10.1021/acs.jchemed.4c00075.

(31) Munawwarah, M.; Alqadri, Z.; Zubair, S. Exploring Students' Perspectives on the Integration of AI in Chemistry Learning Processes: Opportunities and Challenges. Asian J. Sci. Educ. 2025, 7 (1), 76–83. DOI: 10.24815/ajse.v7i1.43492.

(32) Watts, F. M.; Dood, A. J.; Shultz, G. V.; Rodriguez, J.-M. G. Comparing Student and Generative Artificial Intelligence Chatbot Responses to Organic Chemistry Writing-to-Learn Assignments. J. Chem. Educ. 2023, 100 (10), 3806–3817. DOI: 10.1021/acs.jchemed.3c00664.

(33) Zhai, C.; Wibowo, S.; Li, L. D. The Effects of Over-Reliance on AI Dialogue Systems on Students' Cognitive Abilities: A Systematic Review. Smart Learn. Environ. 2024, 11 (1), 28–64. DOI: 10.1186/s40561-024-00316-7.