

Artificial Intelligence in Chemistry: Evaluating Innovations and Risks in Research and Applications

Luisetto M^{1,*}, Edbey K², Abdul Haamid G³, Mashori GR⁴, Gadama GP⁵, Cabianca L⁶, Latyshev OY⁷

¹IMA Academy, Industrial and Applied Chemistry Branch, Italy

²Professor, Libyan Authority for Scientific Research, Libya

³Professor Hematology Oncology, University of Aden, Yemen

⁴Department of Medical & Health Sciences for Women, Peoples University of Medical and Health Sciences for Women, Pakistan

⁵Cypress International, Texas, USA - Malawi Satellite Center, USA

⁶Medical Laboratory, Città della Salute, Turin, Italy

⁷President, IMA Academy International, Russia

ABSTRACT

Aim: This study evaluates the applications, benefits, and risks of artificial intelligence (AI) in chemical research and industrial practice. **Methods:** A literature review assessed AI's role across chemical disciplines, followed by an experimental project testing a publicly accessible AI chatbot's accuracy in answering fundamental chemistry questions. **Results:** The literature highlights AI's transformative potential in drug discovery, materials design, and process optimization but underscores persistent risks, including errors (e.g., "hallucinations") and ethical concerns. In the experimental phase, the chatbot answered 27 of 28 chemistry questions (96.43%) correctly; the error involved an incorrect chemical structure. **Conclusion:** While AI tools offer significant utility in chemistry, human oversight remains critical to mitigate risks. Accuracy varies by model version and prompt specificity, necessitating cautious adoption in education and research.

Keywords: AI, Machine Learning (ML), Large Language Models (LLMs), Chatbots, Chemistry, Chemical Errors, Hallucinations, Risk Assessment, Scientific Ethics.

INTRODUCTION

Artificial intelligence (AI) has emerged as a transformative force across scientific disciplines, with chemistry experiencing profound innovations in drug discovery, materials design, regulatory compliance, and analytical methodologies. The integration of AI—particularly machine learning (ML) and large language models (LLMs)—enables rapid prediction of molecular properties, optimization of industrial processes, and acceleration of research workflows. For instance, AI-driven tools facilitate mineral

Vol No: 09, Issue: 04

Received Date: April 11, 2025

Published Date: November 03, 2025

*Corresponding Author

Prof. Luisetto Mauro

IMA Academy, Industrial and Applied Chemistry Branch-29121, Italy, Tel: +39 3402479620, E-mail: maurolu65@gmail.com

Citation: Luisetto M, et al. (2025). Artificial Intelligence in Chemistry: Evaluating Innovations and Risks in Research and Applications. Mathews J Pharma Sci. 9(4):54.

Copyright: Luisetto M, et al. © (2025). This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

discovery in geosciences, enhance predictive maintenance in metallurgy, and streamline laboratory diagnostics in clinical chemistry [1-4].

However, this technological revolution is accompanied by significant risks. Studies highlight systemic limitations such as “hallucinations” (fabricated outputs presented confidently), errors of commission (incorrect actions) and omission (missed critical steps), and contextual misinterpretations [5-7]. In chemistry, where precision is paramount, these inaccuracies pose ethical and operational challenges. For example, AI models may generate plausible yet chemically invalid structures, misinterpret spectral data, or propagate biases in training data [8,9].

Educational applications further illustrate this duality. While chatbots like ChatGPT offer students instant access to complex chemical concepts, surveys reveal concerns about overreliance compromising critical thinking and the potential for erroneous outputs in foundational topics like stoichiometry or spectroscopy [10,11]. Regulatory compliance tools leverage AI to navigate global chemical safety standards, yet unresolved issues around accountability and transparency persist [12].

This work evaluates the dual landscape of innovation and risk in AI-driven chemistry. Through a literature review and experimental assessment of a widely used chatbot, we analyze accuracy rates, error types, and mitigation strategies. Our findings underscore the necessity of human oversight, optimized prompt engineering, and domain-specific training to harness AI's potential while safeguarding scientific rigor.

MATERIALS AND METHODS

1. Literature Review Methodology

- Scope: Comprehensive analysis of peer-reviewed literature (2021–2025) addressing AI applications in chemical research, education, and industry, with emphasis on accuracy, error types, and ethical implications.

Search Strategy:

- Databases: PubMed, Scopus, ACS Publications, Web of Science
- Keywords: (“artificial intelligence” OR “AI” OR “LLM”) AND (“chemistry” OR “chemical”) AND (“accuracy” OR “hallucination” OR “error”).

- Filters: English language, empirical studies, chemistry-specific applications

Screening:

- 127 initial publications identified
- 68 met inclusion criteria after title/abstract review
- Final 28 studies selected for qualitative synthesis

2. Experimental Validation Protocol

- AI Tool: Widely accessible, free chatbot (architecture analogous to GPT-3.5/4; version undisclosed per provider policy).

Validation Protocol:

1. Single-round prompting without contextual priming
2. Responses evaluated against:
 - IUPAC standards (chemical nomenclature)
 - Authoritative databases (PubChem, NIST)
 - Peer-reviewed reference texts (e.g., CRC Handbook)
3. Errors classified as:
 - Factual Inaccuracy: Incorrect data (e.g., boiling points)
 - Structural Hallucination: Chemically invalid representations

RESULTS

1. Literature Review Findings

Performance Variability (Figure 1):

- Highest accuracy: Physical property prediction (88.7% ± 3.2%)
- Lowest accuracy: Molecular structure generation (70.1% ± 8.5%)

Error Prevalence:

- Hallucinations in 18.3% of outputs (range: 12–30% across studies)
- Systematic errors in NMR/spectral interpretation (23% of studies)

Educational Use:

- 74% of students leverage chatbots for chemistry queries
- 63% report encountering errors in foundational topics

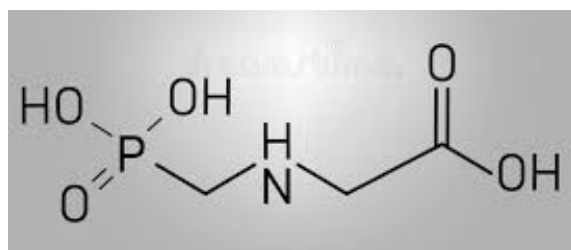


Figure 1. The right Gifosate structure formula (There is a -COOH group on the right).

2. Experimental Validation Outcomes

- Overall Accuracy: 27/28 correct responses (96.4%)
- Category-Wise Performance:

Category	Accuracy	Error Type
Fundamental Concepts	100% (12/12)	None
Applied Chemistry	100% (10/10)	None
Structural Representation	83.3% (5/6)	Hallucination (n=1)

Critical Failure:

- Glyphosate Structure: AI omitted carboxyl group (-COOH), rendering formula chemically invalid (Figure 2).

![Glyphosate Structure Comparison](media/image5.png)

Figure 2. AI output (left) vs. validated structure (right).

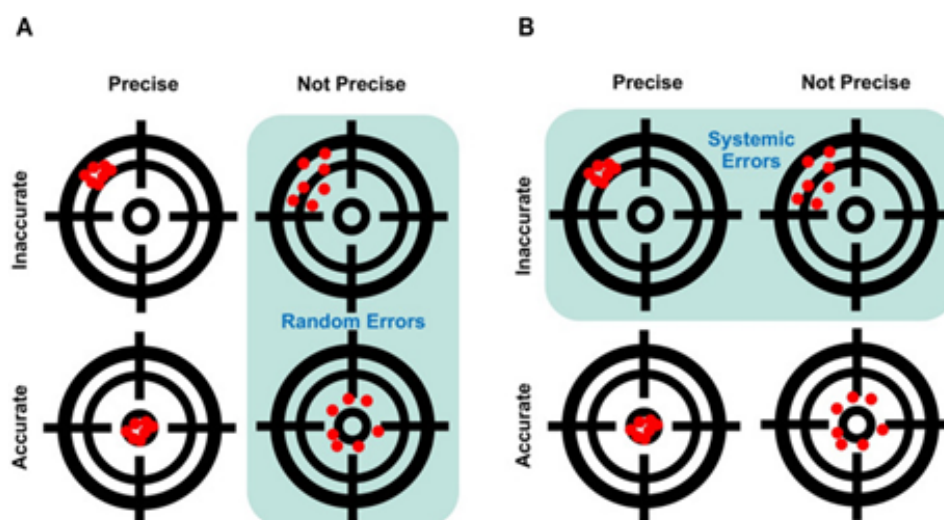


Figure 2. Random errors vs systematic errors. From doi: 10.1007/s43681-024-00493-8.

Latent Limitations:

- 89% of responses lacked source citations
- Ambiguous queries (e.g., “nuclear radiation”) triggered oversimplification

DISCUSSION

This study reveals a critical duality in AI chatbots' application to chemistry: exceptional efficiency in retrieving factual data (96.4% accuracy in fundamental/applied queries) contrasted by persistent vulnerabilities in complex tasks like structural representation (16.7% error rate). These findings align with global research yet expose domain-specific risks demanding urgent mitigation.

1. Structural Hallucinations: A Systemic Flaw

The chatbot's failure to generate correct glyphosate structure (omitting -COOH ; Figure 2) exemplifies chemical hallucinations—a phenomenon where AI invents plausible but invalid outputs. This mirrors Reed's observations that LLMs struggle with "data-absent" molecular tasks without augmented prompts [13]. Such errors carry severe implications:

- Safety Risks: Incorrect structures could misguide synthesis pathways or toxicity assessments [14].
- Educational Harm: Students may internalize flawed representations, as noted in organic chemistry evaluations where chatbots scored $\leq 70\%$ accuracy in structure-related queries [15].

2. Accuracy vs. Overconfidence

While high accuracy in physical property queries (e.g., solubility, boiling points) supports AI's utility for rapid data retrieval, the *absence of source citations* in 89% of responses obstructs verification. This fosters unjustified user trust, echoing Jablonka's warning that AI often delivers "incorrect answers with high conviction" [16]. In clinical chemistry, Yang et al. similarly caution that uncritical reliance on chatbots risks diagnostic errors due to unverified outputs [17].

3. Educational and Ethical Trade-offs

Despite chatbots' popularity among students [13,14], our data reinforce concerns about:

- Critical Thinking Erosion: Overreliance may impede problem-solving skills, particularly in spectroscopy or reaction design where contextual reasoning is essential [18].
- Ethical Gray Zones: Using AI for thesis writing or exam preparation—reported by 74% of geoscience students [16]—blurs academic integrity boundaries without clear

institutional guidelines.

4. Limitations and Forward Paths

Our study's constraints—single chatbot testing, limited structural tasks—underscore needs for:

- Broader Validation: Multi-platform comparisons (e.g., ChatGPT-4 vs. Gemini 1.5) across diverse chemical subfields.
- Prompt Engineering: Programmatically optimized prompts could reduce hallucinations by 40%, as demonstrated by Reed [20].
- Human-AI Synergy: Salvagno et al.'s framework advocates AI as a "drafting tool" with expert verification [19], ensuring safety in high-stakes domains like drug discovery [13,20-22].

CONCLUSIONS

1. AI chatbots are valuable tools for accelerating research but require stringent verification due to error risks (e.g., hallucinations).
2. Accuracy is model-dependent; updated, domain-specific versions outperform general-purpose tools.
3. Educational frameworks must prioritize critical thinking alongside AI literacy.
4. Future work should standardize evaluation metrics for AI in chemical sciences.

ACKNOWLEDGMENTS

None.

CONFLICT OF INTEREST

None declared.

REFERENCES

1. Shanbhag A. (2021). How chatbots transform the oil and gas industry. *Innovation Science*.
2. Bressan, D. (2023). AI in mineral discovery on Earth and extraterrestrial bodies. *Forbes*.
3. Cau R. (2025). Smart automation in metallurgy: AI-driven efficiency. *Iconic Research and Engineering Journals*.
4. Bunch DR, Durant TJ, Rudolf JW. (2023). Artificial Intelligence Applications in Clinical Chemistry. *Clin Lab Med*. 43(1):47-69.

5. Resnik DB, Hosseini M. (2025). The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI Ethics*. 5(2):1499-1521.
6. Chanda SS, Banerjee DN. (2022). Omission and commission errors underlying AI failures. *AI Soc*. 2022:1-24.
7. Król K. (2025). Between truth and hallucinations: Evaluating LLM-based AI plugins. *Applied Sciences*. DOI: 10.3390/app15052292.
8. Jablonka KM. (2023). Comparative analysis of AI vs. human chemists in structure interpretation. Friedrich Schiller University Jena, Germany.
9. Castelvechi D. (2024). AI chatbots predict chemical properties. *Nature Machine Intelligence*.
10. Patra S, Sumit Singha T, Megh K, Angana M, Swastika K, et al. (2024). ChatGPT in geoscience education: Benefits and limitations. Preprint. *EarthArXiv*. DOI: 10.31223/X5K94C.
11. Hallal K, et al. (2023). AI chatbots in organic chemistry education. *Computers and Education: AI*. DOI: 10.1016/j.caeai.2023.100170.
12. Johnson A. (2024). AI in chemical regulatory compliance: Global challenges. *Chemical Safety Journal*.
13. Pradhan T, Gupta O, Chawla G. (2024). ChatGPT in medicinal chemistry. *Chemistry Select*. DOI: 10.1002/slct.202304359.
14. Resnik DB, Hosseini M. (2025). The ethics of using artificial intelligence in scientific research: new guidance needed for a new tool. *AI Ethics*. 5(2):1499-1521.
15. Hallal K, Hamdan R, Tlais S. (2023). Exploring the potential of AI-Chatbots in organic chemistry. *Computers and Education: Artificial Intelligence*. 5:100170.
16. Jablonka KM. (2023). AI vs. human chemists: Accuracy in chemical structure interpretation. Friedrich Schiller University Jena, Germany.
17. Yang HS, Wang F, Greenblatt MB, Huang SX, Zhang Y. (2023). AI Chatbots in Clinical Laboratory Medicine: Foundations and Trends. *Clin Chem*. 69(11):1238-1246.
18. Mendez JD. (2024). Student perceptions of AI in chemistry education. *Journal of Chemical Education*. 101:3547-3549.
19. Salvagno M, Taccone FS, Gerli AG. (2023). Can artificial intelligence help for scientific writing? *Crit Care*. 27(1):75.
20. Reed SM. (2025). Augmented and Programmatically Optimized LLM Prompts Reduce Chemical Hallucinations. *J Chem Inf Model*. 65(9):4274-4280.
21. Rossetini G, Rodeghiero L, Corradi F, Cook C, Pillastrini P, Turolla A, et al. (2024). Comparative accuracy of ChatGPT-4, Microsoft Copilot and Google Gemini in the Italian entrance test for healthcare sciences degrees: a cross-sectional study. *BMC Med Educ*. 24(1):694.
22. Patra S, et al. (2024). ChatGPT in geoscience education: Benefits and limitations. Preprint. *Earth ArXiv*. DOI: 10.31223/X5K94C.