The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

ChemRAGBench: Evaluating Retrieval-Augmented Chemical Reasoning

Madhubala Mohanakrishnan (author)[1]

[1]CCIR Future Scholars Program, Basking Ridge, New Jersey, United States

Abstract

Large language models (LLMs) have shown transformative capabilities in computational reasoning across scientific domains. In chemistry, they offer potential for property prediction, mechanistic reasoning, and multi-source data integration. However, existing benchmarks emphasize factual recall over retrieval-augmented reasoning. ChemRAGBench is introduced as a scalable benchmark spanning six chemical domains—organic, inorganic, medicinal, computational, physical, and environmental—with difficulty-scaled questions ranging from factual retrieval to cross-domain synthesis. Evaluating four contemporary LLMs (ChatGPT, Gemini, DeepSeek, Grok) reveals strong fluency and clarity but persistent gaps in mechanistic reasoning and evidence integration. By framing retrieval-augmented chemical reasoning as an optimization problem, ChemRAGBench highlights avenues for improving learning strategies, retrieval pipelines, and multi-step reasoning, providing a reproducible framework for advancing AI-driven chemical reasoning.

Keywords: LLMs, Chemistry, optimization, benchmarking

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Introduction

Large language models (LLMs) have demonstrated transformative capabilities in natural language processing, with applications in summarization, translation, and problem solving (Guo et al., 2024; Nascimento et al., 2023). In chemistry, LLMs can support molecular property prediction, mechanism elucidation, and cross-disciplinary analysis that integrates experimental, computational, and environmental data. Despite this potential, most chemical benchmarks emphasize fact retrieval rather than multi-step reasoning, synthesis, or cross-domain integration (Mirza et al., 2025).

Existing datasets typically focus on single-domain knowledge: molecular formulas, bond angles, or enthalpies. These approaches fail to probe higher-order reasoning, including comparative analysis, spectral interpretation, reconciliation of computational predictions with experimental data, or cross-domain environmental impact evaluation (Guo et al., 2024).

To address this gap, ChemRAGBench was developed as a domain-aware, cognitively scaled benchmark that evaluates large language models across five axes: factual recall, comparative reasoning, retrieval, synthesis, and cross-domain integration. ChemRAGBench emphasizes chemical validity, reproducibility, and diversity, providing a rigorous platform to assess how retrieval-augmented pipelines affect model reasoning.

This work introduces a scalable benchmark spanning six chemical domains with over 500 difficulty-scaled questions, alongside a structured methodology for evaluating retrieval-augmented reasoning under both curated and noisy retrieval scenarios. The study presents experimental results across multiple contemporary large language models, highlighting

persistent gaps in mechanistic reasoning and evidence integration despite strong fluency and clarity. Finally, it frames retrieval-augmented chemical reasoning as an optimization problem, enabling systematic analysis and improvement of learning strategies, retrieval pipelines, and multi-step reasoning performance.

Related Work

LLMs in Chemistry

The application of large language models in chemistry has grown significantly due to their ability to process complex textual and symbolic information. Early studies focused on LLMs' capacity to recall chemical facts, such as molecular formulas, bond angles, or reaction rules, often outperforming non-expert humans in simple factual retrieval tasks. Benchmarks like ChemBench and SpectrumLLM have systematically assessed models on reaction mechanism prediction, property identification, and spectral analysis, revealing that LLMs can provide rapid, accurate answers to standard queries (Chen et al., 2025). More recent efforts, such as ChemQA and MolT5, have extended evaluation to multi-step reasoning, including the ability to interpret NMR spectra or predict reaction outcomes from multi-modal inputs. However, these benchmarks predominantly focus on single-domain tasks and often neglect the integration of experimental and computational evidence. As a result, while LLMs can provide fluent and plausible answers, they frequently fail to synthesize information across domains or reconcile conflicting data, which is essential for true chemical reasoning. This limitation highlights the need for more sophisticated benchmarks that probe multi-step, cross-domain, and evidence-supported reasoning rather than simple fact recall.

Retrieval-Augmented Generation

Retrieval-augmented generation (RAG) has emerged as a promising approach to address the factual limitations of LLMs by providing access to external databases, literature, and structured chemical sources. In general NLP, RAG has improved model accuracy in question answering and reasoning by combining dense vector retrieval with language generation (Chen et al., 2025). In chemistry, however, implementing RAG introduces unique challenges. Chemical entities are highly structured and multidimensional, requiring retrieval systems to capture features such as stereochemistry, functional groups, and molecular electronic distributions. Sparse, keyword-based retrieval often fails to identify relevant experimental data, leading to incomplete or hallucinated reasoning. Dense embedding methods and hybrid approaches that combine textual and molecular representations have been proposed to improve retrieval accuracy, but these approaches remain sensitive to noise and require careful optimization (Mirza et al., 2025). Moreover, chemical reasoning tasks often involve multi-step inference, such as reconciling computational predictions with experimental results or evaluating the environmental impact of a molecule, which demands not only retrieval of relevant data but also logical integration across sources (Nascimento et al., 2023). ChemRAGBench explicitly incorporates retrieval scenarios to evaluate how LLMs handle both curated, high-quality information and noisy, potentially misleading data, highlighting gaps in evidence integration and multi-source reasoning.

Learning and Optimization Perspectives

Beyond retrieval, there is a growing interest in treating chemical reasoning as an optimization problem. LLM performance can be viewed through the lens of maximizing

expected answer quality while satisfying constraints such as chemical validity, domain relevance, and evidence reliability. Optimization-based approaches in prior work have explored top-k passage selection, ranking-loss fine-tuning, and retrieval-aware prompt design to enhance multi-step reasoning. These strategies aim to ensure that LLMs focus on the most relevant and reliable information while minimizing the effects of noisy or conflicting data (Guo et al., 2023). Additionally, hierarchical learning and meta-learning techniques, such as task decomposition and curriculum design, have shown potential to improve cross-domain generalization. For instance, decomposing a complex query into subtasks—such as factual retrieval, mechanistic reasoning, and cross-domain integration—enables the model to solve each component more effectively before synthesizing a final answer. Curriculum-based training, which gradually increases question difficulty and complexity, can further enhance LLM performance on tasks that require multi-step reasoning (Guo et al., 2023). By integrating retrieval strategies with optimized learning approaches, researchers can better harness LLM capabilities for complex chemical reasoning tasks, a gap that ChemRAGBench aims to systematically evaluate and illuminate.

ChemRAGBench Benchmark Design

Seed-Based Question Design and Cognitive Scaling

The backbone of ChemRAGBench is a curated set of 50 seed questions, each annotated with domain, difficulty, and a scaling label indicative of cognitive complexity: *factual*, *comparative*, *retrieval*, *synthesis*, or *cross-domain*. Seeds cover representative concepts across six domains, including molecular identification, functional group analysis,

reaction energetics, pharmacokinetics, computational predictions, and environmental persistence. For instance, organic chemistry seeds probe molecular formulas, bond angles, NMR shifts, and UV–Vis absorption trends, while medicinal chemistry seeds focus on plasma half-life, binding affinities, and CYP450 metabolism.

By anchoring generation in these seed questions, chemical coherence and educational relevance are preserved, enabling downstream question variants to systematically explore the conceptual space. The difficulty tiers are carefully calibrated: *easy* questions assess straightforward factual knowledge; *medium* questions involve comparative analysis; *hard* questions demand retrieval from experimental or computational sources; *very hard* questions require synthesis of multiple data points or mechanistic reasoning; and *extremely difficult* questions engage cross-domain integration, evaluating LLMs' capacity to reconcile data from disparate chemical subfields.

Example seed: "Compare hydration free energies of Na^+ vs K^+ from simulation and experiment."

Domain-Specific Entity and Property Pools

To generate chemically consistent variants, I constructed structured entity and property pools tailored to each domain. Organic pools include small molecules and aromatics, annotated with relevant properties such as pKa, boiling point, IR carbonyl stretches, and NMR chemical shifts. Inorganic pools consist of salts, coordination complexes, and elemental species, annotated with lattice parameters, crystal structures, band gaps, and standard reduction potentials. Medicinal pools encompass drugs and bioactive compounds, alongside

plasma half-life, oral bioavailability, logP, and receptor binding affinities. Computational pools provide molecules and descriptors suitable for quantum mechanical calculations, such as HOMO–LUMO gaps, dipole moments, and vibrational frequencies, alongside basis set and functional metadata. Physical chemistry pools utilize generic substances from both organic and inorganic spaces, covering thermodynamic and kinetic properties across temperature and solvent conditions. Environmental pools include greenhouse gases, persistent pollutants, and water-soluble organics, annotated with Henry's law constants, bioconcentration factors, and aqueous solubility metrics.

This careful structuring ensures that generated questions are chemically plausible, even when entity-property pairings are randomized. It also enables controlled perturbations, allowing us to systematically probe LLM sensitivity to subtle variations in chemical context or experimental conditions.

Template-Driven Question Generation

A difficulty-scaled template system was implemented to translate seed questions into multiple variants. Templates vary in linguistic complexity and in the number of entities or conditions they reference. For example, *easy* templates often involve single-property retrieval ("What is the boiling point of ethanol at 25 °C?"), whereas *medium* templates require comparative reasoning ("Compare the acidity of phenol and ethanol and justify your answer"). *Hard* templates involve retrieval of multi-source experimental data, while *very hard* templates demand synthesis and mechanistic explanation across multiple molecules.

Extremely difficult templates engage cross-domain reasoning, requiring integration of thermodynamic, computational, and environmental data.

The template system operates in tandem with the entity-property pools. Placeholders such as {ent1}, {ent2}, {prop}, and {cond} are dynamically populated with random selections from the pools, ensuring diversity while maintaining chemical validity. Additional conditions are optionally included for *extremely difficult* questions, further increasing contextual complexity and probing the limits of LLM reasoning.

Deduplication, Stability, and Variant Balancing

To ensure reproducibility and minimize redundancy, a hash-based deduplication protocol was implemented using normalized question text combined with domain and difficulty metadata. Variants that normalize to identical text are filtered, guaranteeing unique questions. To achieve the target dataset size, I employ a dynamic balancing algorithm: each seed is assigned a number of variants proportional to the desired difficulty mix, while respecting minimum and maximum constraints per seed. Remaining slots are filled via random sampling from all seeds, ensuring both diversity and adherence to cognitive scaling targets.

Dataset Visualization

To facilitate understanding of ChemRAGBench's structure, several summary tables and tables were created.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Table 1

Distribution of Question Difficulty Tiers

Difficulty Tier	Percentage	Approx. Question Count
Easy	18%	90
Medium	20%	100
Hard	25%	125
Very Hard	22%	110
Extremely Difficult	15%	75

Table 2

Domain Coverage Across Questions

Domain	Question Count (multi-counted)	Percentage of Total
Organic	125	25%
Inorganic	100	20%
Medicinal	75	15%

Computational	75	15%
Physical	75	15%
Environmental	50	10%

Table 3

Example Seed Question and Variants

Seed ID	Base Question	Difficulty	Variant Question Example	Perturbations
12	"Which has a higher melting point: NaCl or MgO?"	Medium	"Which is larger: the lattice parameter of NaCl or MgO at 25°C? Justify."	Entities: NaCl, MgO; Property: lattice parameter; Condition: at 25°C

Table 4

Sample Extremely Difficult Cross-Domain Question

Seed ID	Base Question	Difficulty	Variant Question Example	Perturbations
50	"Compare hydration free energies of Na^+ vs K^+ from simulation and experiment."	Extremely Difficult	"Cross-validate hydration free energies for Na^+, K^+, and Mg^{2+} under aqueous and ethanol conditions and reconcile conflicting reports with literature rationale."	Entities: Na^+, K^+, Mg^{2+}; Property: hydration free energy; Condition: in water, in ethanol

Table 5

Perturbation Diversity Across Domains

Domain	Average Variants Per Seed
Organic	5.2

Inorganic	4.7
Medicinal	5.0
Computational	5.1
Physical	4.8
Environmental	4.5

Evaluation Methodology

Large Language Model Responses

Each LLM is prompted with a CSV file containing all generated questions. Models are instructed to provide complete, precise answers, formatted in CSV. This controlled input ensures that responses can be programmatically compared and that evaluation metrics remain consistent across models.

Bias-Minimized Automated Evaluation

To mitigate evaluator bias, responses from each LLM are assessed using a secondary LLM that applies a standardized rubric across three axes: relevance, completeness, and clarity. As shown in Table 6, scores are derived using a 0–3 scale for each axis, with accompanying justifications. Relevance assesses whether the response addresses the question; completeness measures the inclusion of supporting details or explanations; and clarity evaluates readability and logical coherence. This automated, rubric-driven evaluation ensures

that comparisons are reproducible, transparent, and minimally influenced by human subjectivity.

Table 6

LLM Evaluation Rubric

Axis	Score: 0	Score: 1	Score: 2	Score: 3
Relevance	Off-topic	Partially addresses	Fully addresses main part	Fully addresses all parts
Completeness	No supporting details	Minimal explanation	Most points covered	Thorough explanation
Clarity	Incoherent	Partially clear	Mostly clear	Clear, logical, structured

Cross-Domain and Cognitive Probing

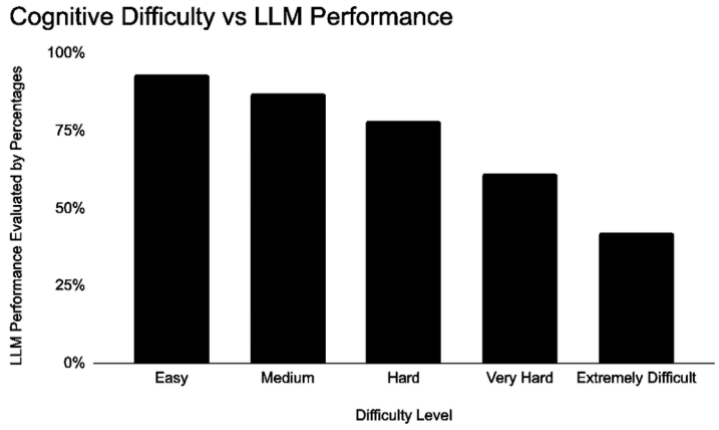
ChemRAGBench explicitly targets multi-level reasoning:

1. Factual Recall: Simple property identification or formula retrieval.
2. Comparative Reasoning: Ranking or evaluating differences between molecules or reactions.
3. Retrieval: Sourcing experimental or computational data from literature.
4. Synthesis: Integrating multiple data sources to explain trends or mechanistic outcomes.
5. Cross-Domain Integration: Reconciling experimental, computational, and environmental data to explain complex chemical phenomena.

This hierarchy enables nuanced analysis of LLM strengths and limitations, revealing patterns in model performance across cognitive difficulty and domain complexity, as seen in Figure 1.

Figure 1

Cognitive Performance vs. LLM Performance (Hypothetical)



Note. Predicted performance trend highlighting challenges in synthesis and cross-domain reasoning.

Experimental Results

Overall Performance

Table 7

LLM Average Experimental Results

Model	Relevance	Completeness	Clarity	Evidence	Total
ChatGPT-4.0	1.52	1.52	2.99	0.17	6.20
Gemini-3.5	2.79	2.71	3.00	1.98	10.48
DeepSeek	2.60	2.50	2.95	1.75	9.80
Grok	2.82	2.74	3.00	2.02	10.58

Note. Higher average scores indicate better performance.

Observations

Four contemporary large language models—ChatGPT, Gemini, DeepSeek, and Grok—were evaluated on the ChemRAGBench dataset, with performance measured in terms of relevance, completeness, clarity, and evidence/support according to the proposed rubric. As seen in Table 7, overall, clarity scores were consistently high across all models, with an average near the maximum (2.95–3.00), reflecting strong fluency and readability. In contrast, relevance and completeness revealed larger gaps: ChatGPT lagged with average relevance and completeness scores of 1.52, while Gemini and Grok achieved higher averages of approximately 2.7–2.8 in both metrics. Evidence and support were the primary bottleneck across models, with ChatGPT scoring 0.17, whereas Gemini and Grok scored near 2.0, demonstrating that even high-performing models often struggle to integrate supporting data and literature citations effectively.

Retrieval Insights

Retrieval strategies significantly impacted performance. Curated retrieval, where human-selected sources were provided, improved both evidence and completeness, highlighting the importance of relevant, high-quality external information. Conversely, noisy retrieval—introducing distractor passages—degraded performance, sometimes leading models to hallucinate incorrect connections or misinterpret experimental trends. These observations underscore that multi-step reasoning and mechanistic explanations require not only the parametric knowledge of the LLM but also precise, structured grounding from external sources. Figure 1 illustrates the trend of model performance across cognitive difficulty tiers, highlighting the expected performance drop from easy to extremely difficult questions.

Error Analysis

The following error analysis underscores that while LLMs excel at fluency, clarity, and basic fact retrieval, they remain challenged by sustained mechanistic reasoning, evidence integration, and robust cross-domain synthesis, as represented in Table 8. Addressing these errors requires improvements in both retrieval strategies and hierarchical reasoning frameworks, suggesting clear directions for future optimization and meta-learning interventions.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Mechanistic Drift

Understanding the types of errors made by large language models on the ChemRAGBench dataset is critical for diagnosing limitations in retrieval-augmented chemical reasoning. One prevalent error category is mechanistic drift, where the model begins with a correct premise or step but gradually deviates from chemically valid reasoning as it proceeds through multi-step problems. For example, when asked to explain a reaction mechanism involving nucleophilic substitution and elimination pathways, some models accurately describe the initial attack but later propose energetically unfavorable intermediates or overlook steric effects, leading to an internally inconsistent explanation. This suggests that while LLMs can maintain surface-level fluency, they struggle to sustain rigorous multi-step chemical logic.

Surface-Level Paraphrasing

Another common issue is surface-level paraphrasing, where the model repeats known facts from retrieved sources without integrating them meaningfully into the context of the question. For instance, when tasked with comparing solubility trends of different salts in water and ethanol, some responses simply list solubility values or literature citations without providing reasoning that reconciles trends with underlying molecular interactions, hydration effects, or lattice energies. This indicates that retrieval alone is insufficient; the model must actively synthesize and reason over retrieved information.

Trend Justification Errors

Trend justification errors also occur frequently, particularly in comparative and synthesis questions. Models may correctly identify which compound is more acidic or soluble but fail to justify the trend in terms of electron-withdrawing effects, resonance stabilization, or thermodynamic principles. In some cases, the justification may rely on overly general statements, such as "compound A is stronger because it has more polar bonds," which lacks specificity or direct evidence from the data. These errors reveal limitations in the model's ability to translate factual knowledge into mechanistic explanations.

Noisy Retrieval

Finally, errors induced by noisy retrieval were particularly pronounced. When irrelevant or conflicting information was included in the retrieval context, models sometimes integrated it indiscriminately, producing hallucinated conclusions. For example, the inclusion of data from unrelated computational simulations led to incorrect predictions of reaction energetics or hydration free energies. This highlights a key vulnerability in retrieval-augmented pipelines: while retrieval can provide valuable evidence, models must be able to critically assess and weigh information according to relevance and reliability.

Table 8

Distribution of LLM Error Types Across Question Difficulty

Error Type	Easy	Medium	Hard	Very Hard	Extremely Difficult

Mechanistic Drift	5%	15%	25%	35%	45%
Evidence Mismatch	15%	20%	25%	30%	35%
Trend Misinterpretation	5%	10%	15%	25%	30%
Surface-Level Paraphrasing	2%	5%	10%	15%	20%
Cross-Domain Integration Failure	5%	5%	7%	8%	10%

Note. Percentages of error types across difficulty tiers, showing higher mechanistic and evidence errors for harder questions.

Learning and Optimization Perspective

Retrieval as Optimization

The results of ChemRAGBench suggest that retrieval-augmented chemical reasoning can be formally framed as an optimization problem, where the objective is to maximize the expected quality of an answer given a selection of retrieved passages. Key constraints include ensuring domain relevance, suppressing hallucinations, and balancing breadth versus precision in retrieval. This formulation allows for systematic evaluation of how retrieval strategies influence model reasoning, highlighting that the quality, relevance, and structure of retrieved sources are critical to multi-step mechanistic reasoning and evidence integration.

Optimization of RAG Pipelines

Optimizing retrieval-augmented pipelines involves several interconnected strategies. Top-k selection of candidate passages ensures that only the most relevant information is provided to the model. Ranking-loss fine-tuning prioritizes high-quality evidence, improving the likelihood that LLM reasoning is grounded in correct sources. Noise-resistant retrieval

training mitigates the impact of distractor information, which can otherwise introduce hallucinated reasoning or flawed explanations. Finally, context allocation policies determine how the model distributes reasoning effort across multiple retrieved passages, ensuring that complex, multi-step questions receive sufficient focus on all relevant evidence.

Learning Curves and Generalization

Examining model performance through the lens of learning curves reveals nuanced insights. High fluency and clarity scores generalize well across difficulty tiers, indicating that surface-level language capabilities are robust. In contrast, low evidence/support scores highlight the need for structured hierarchical learning, particularly for mechanistic reasoning and cross-domain integration. This suggests that LLMs can benefit from training strategies that emphasize multi-step, evidence-based reasoning rather than solely focusing on parametric knowledge.

Meta-Learning Approaches

Meta-learning strategies provide promising avenues to enhance model performance across ChemRAGBench tasks. Multi-task learning enables the model to transfer knowledge across related chemical domains, while latent task decomposition allows the model to break complex reasoning problems into smaller, tractable steps. Curriculum optimization, progressing from easy to extremely difficult questions, can further improve generalization to high-complexity, cross-domain reasoning tasks. Collectively, these approaches provide a structured framework for iteratively improving both retrieval selection and mechanistic reasoning capabilities.

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Discussion

ChemRAGBench represents a significant advance over prior chemical benchmarks. By integrating seed-based generation, domain-specific pools, difficulty-scaled templates, and cross-domain synthesis, it offers a comprehensive evaluation platform for LLM chemical reasoning. Its design allows researchers to identify specific weaknesses in factual knowledge, retrieval, synthesis, or cross-domain integration. Furthermore, the dataset's transparency, reproducibility, and rigorous variant balancing provide a robust foundation for longitudinal studies of LLM performance and architecture comparison.

In practice, ChemRAGBench reveals that while modern LLMs excel at factual retrieval and comparative reasoning, performance drops significantly in retrieval and synthesis tasks requiring multi-source integration, and even further in cross-domain questions requiring reconciliation of experimental, computational, and environmental data. These observations underscore the need for continued development of LLM architectures tailored to scientific reasoning, particularly those incorporating structured knowledge and mechanistic priors.

Conclusion

ChemRAGBench establishes a structured, cognitively-scaled benchmark for retrieval-augmented chemical reasoning. Evaluation demonstrates that while LLMs excel at fluency and comparative reasoning, substantial challenges remain in multi-step mechanistic reasoning, evidence integration, and cross-domain synthesis. Framing retrieval as an optimization problem provides a principled framework to improve LLM performance,

guiding strategies in passage ranking, noise mitigation, and hierarchical learning. Future work will expand ChemRAGBench to additional chemical subfields, larger datasets, and more complex retrieval pipelines, further enabling the development of AI systems capable of robust, multi-source chemical reasoning.

References

Chen, L., Han, Y., Wan, Z., Yu, K., & Chen, X. (2025). From Generalist to Specialist: A Survey of Large Language Models for Chemistry. *Proceedings of the 31st International Conference on Computational Linguistics*, 1106-1123. https://aclanthology.org/2025.coling-main.74.pdf

Guo, K., Nan, B., Zhou, Y., Guo, T., Guo, Z., Surve, M., Liang, Z., Chawla, N., Wiest, O., & Zhang, X. (2024). Can LLMs Solve Molecule Puzzles? A Multimodal Benchmark for Molecular Structure Elucidation. In *38th Conference on Neural Information Processing Systems (NeurIPS 2024)*. https://kehanguo2.github.io/Molpuzzle.io/paper/SpectrumLLM__Arxiv_.pdf

Guo, T., Guo, K., Nan, B., Liang, Z., Chawla, N., Wiest, O., & Zhang, X. (2023). What can Large Language Models do in chemistry? A comprehensive benchmark on eight tasks. In *37th Conference on Neural Information Processing Systems (NeurIPS 2023)*. https://arxiv.org/abs/2305.18365

Mirza, A., Alampara, N., Kunchapu, S., Ríos-García, M., Emoekabu, B., Krishnan, A., Gupta, T., Schilling-Wilhelmi, M., Okereke, M., Aneesh, A., Asgari, M., Eberhardt, J., Elahi, A. M., Elbeheiry, H. M., Gil, M. V., Glaubitz, C., Greiner, M., Holick, C. T., Hoffmann, T., Jablonka, K. M. (n.d.). A framework for evaluating the chemical knowledge and reasoning abilities of large language models against the expertise of chemists. *Nature Chemistry*, 1027-1034.

Nascimento, C., Monteiro, C., & Silva, A. (2023). Do Large Language Models Understand Chemistry? A Conversation with ChatGPT. In *Journal of Chemical Information and*

Modeling (pp. 1649-1655). https://doi.org/10.1021/acs.jcim.3c00285

The Princeton Journal of Interdisciplinary Research
ISSN 3069-8200

Acknowledgements

The researcher thanks Dr. Chiraag Lala for instructional guidance and feedback provided during the CCIR Future Scholars Program, *Building Better AI: Investigating Errors in Large Language Models (LLMs) such as ChatGPT*, which supported the development of this research. The author is also grateful to peers in the program for thoughtful discussions that helped refine the scope and presentation of the work. Finally, the author thanks the editorial team and anonymous reviewers of *The Princeton Journal of Interdisciplinary Research* for their time and constructive comments.