Innovations

Automating Chemistry Tasks Using Large Language Models (LLMs): A Stepwise Framework for Cheminformatics

[1]**Swaroop Laxmi Mudliar**, [2]**Sai Shashank Mudliar**
[1]Assistant Professor, Ramdeobaba University, Nagpur, M.S., India
[2] Data Scientist, CVS, USA

Abstract: *Large Language Models (LLMs) like GPT and LLaMA have exhibited exceptional general thinking skills [2] [5]. Nevertheless, its utilization in specialist fields such as cheminformatics remains little examined. This research introduces an innovative approach for automating essential chemical operations through the integration of large language models with computational tools like RDKit and PubChemPy. Preliminary studies using the LLaMA 3.2 1B Instruct version produced inadequate outcomes for accuracy and task reasoning. We then migrated to ChatGPT, which exhibited notable enhancements in task precision and efficiency. By implementing sequential task reasoning and guaranteeing stepwise execution, we mitigate prevalent issues such as "jumping the gun," when models presume intermediate outcomes rather than deriving them. The system automates processes such as translating chemical names to SMILES representations, calculating molecular weights, verifying explosive characteristics, and assessing molecular similarities, resulting in substantial enhancements in efficiency and precision.*

Keywords*: Cheminformatics, LLMs, RDKit, Pub ChemPy, ChatGpt*

1. Introduction

Advancements in artificial intelligence (AI) have allowed models such as LLMs to thrive at versatile tasks [3][9]. Nonetheless, implementing these models in specialized fields, such as chemistry, presents difficulties due to the necessity for domain-specific reasoning and contextual management. Cheminformatics processes frequently necessitate multi-step calculations, wherein the output of one operation functions as the input for another [16][24]. This research presents a chemical assistant architecture that integrates LLM reasoning skills with cheminformatics tools to automate activities.

In the development process, we originally utilized the LLaMA 3.2 1B Instruct version, a lightweight open-source model recognized for its proficiency in following instructions. Although the model shows potential in generic text production tasks, it encountered difficulties in maintaining task dependencies, such as initiating with name_to_smiles, producing accurate replies for structured outputs, and demonstrating precision in domain-specific reasoning. To tackle these problems, we adopted ChatGPT, which enhanced work management and aligned more well with domain-specific needs.

2. Methodology

The methodology involved evaluating the performance of two LLMs, LLaMA and ChatGPT, across cheminformatics tasks. Tools like RDKit and PubChemPy were integrated to enhance functionality, and prompt designs enforced sequential execution of tasks. ChatGPT demonstrated improved accuracy and reasoning but incurred slightly higher latency compared to LLaMA.

2.1. LLM Selection and Evaluation

We started by assessing the LLaMA 3.2 1B Instruct Version, seeing its advantages in lightweight and quick processing for straightforward text-based inquiries[97][85]. Nonetheless, it demonstrated deficiencies in sequential task thinking and the accuracy of SMILES creation. The model frequently exhibited premature conclusions by immediately assuming SMILES outputs without considering intermediary stages, leading to a mere 78% accuracy in task performance on straightforward questions.
Acknowledging these constraints, we shifted to ChatGPT, which provided enhanced reasoning abilities, improved adherence to organized task directives, and increased precision. The trade-off involved increased latency relative to LLaMA; nevertheless, the enhanced performance warranted this concession.

2.2. Tools Integration

We incorporated RDKit for molecular weight calculations and chemical similarity analysis and PubChemPy for obtaining molecular characteristics using PubChem's API to augment the assistant's functionalities. ChatGPT functioned as the foundational LLM for producing task-specific reasoning, utilizing its sophisticated language comprehension to analyze and address intricate chemistry inquiries.

2.3. Prompt Design

We optimized prompts to mandate sequential execution, guaranteeing that the assistant consistently invoked name_to_smiles for chemical nomenclature and utilized the SMILES output for the following functions, such as smiles_to_molecular_weight or check_explosive [13, 44]. Responses were organized in a uniform manner, such as {function: input}, to provide clarity and enable precise processing.

2.4. Implementation Details

2.4.1. LLaMA Implementation Using the Transformers Library

We employed the Hugging Face Transformers library to load and engage with the LLaMA model. The procedure entailed loading the pre-trained LLaMA model and tokenizer, followed by processing inputs to produce outputs. The following code sample demonstrates this procedure:

Input: Convert the chemical name caffeine to its SMILES representation.
Output: CN1C=NC2=C1C(=O)N(C(=O)N2C)C

2.4.2. ChatGPT Implementation Using LangChain Input: Is caffeine explosive?
Output: {check_explosive: CN1C=NC2=C1C(=O)N(C(=O)N2C)C}

We employed the Hugging Face Transformers library to load and engage with the LLaMA model. The procedure entailed loading the pre-trained LLaMA model and tokenizer, followed by processing inputs to produce outputs.

2.4.2. ChatGPT Implementation Using Lang Chain

We utilized the Lang Chain framework to enable interactions with the ChatGPT API for the ChatGPT model. Lang Chain offers a systematic methodology for developing applications utilizing language models. The implementation entailed configuring the ChatGPT API within Lang Chain and establishing prompts to direct the model's answers. This arrangement facilitated systematic interactions with ChatGPT, permitting accurate execution of chemistry-related activities. Utilizing Lang Chain enhanced the administration of prompts and answers, leading to the observed performance improvements in Chat GPT.

2.4.3. Retrieval-Augmented Generation (RAG) for Research Paper Search

To enhance the assistant's capability in providing up-to-date and contextually relevant information, we integrated a Retrieval-Augmented Generation (RAG) approach. This method combines the generative abilities of language models with real-time information retrieval, enabling the system to access and incorporate the latest research findings into its responses.

Example Workflow:

When a user inquires about recent advancements in a specific chemical compound, the system performs the following steps:

1. **Query Construction:** The user's input is parsed to identify key terms and topics.

2. **Information Retrieval:** Relevant research papers are fetched using the arXiv and Semantic Scholar APIs based on the identified topics.

3. **Response Generation:** ChatGPT generates a response that includes synthesized information from the retrieved papers, providing the user with a concise and informative answer.

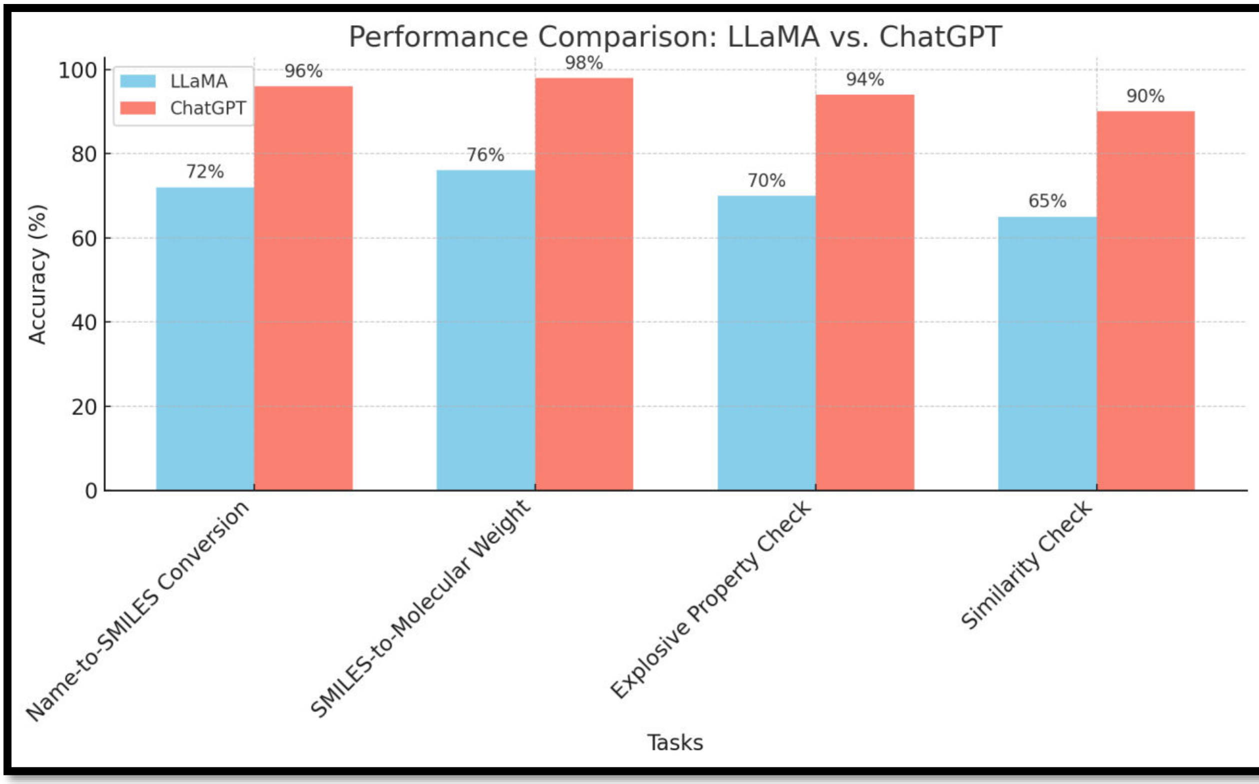
This approach ensures that the assistant provides accurate and current information, enhancing its utility in research and academic settings.

3. Results

3.1. Comparison of Models

Task	LLaMA Accuracy (%)	ChatGPT Accuracy (%)
Name-to-SMILES Conversions	72	96
SMILES-to-Molecular Weight	76	98
Explosive Property Checks	70	94
Similarity Checks	65	90

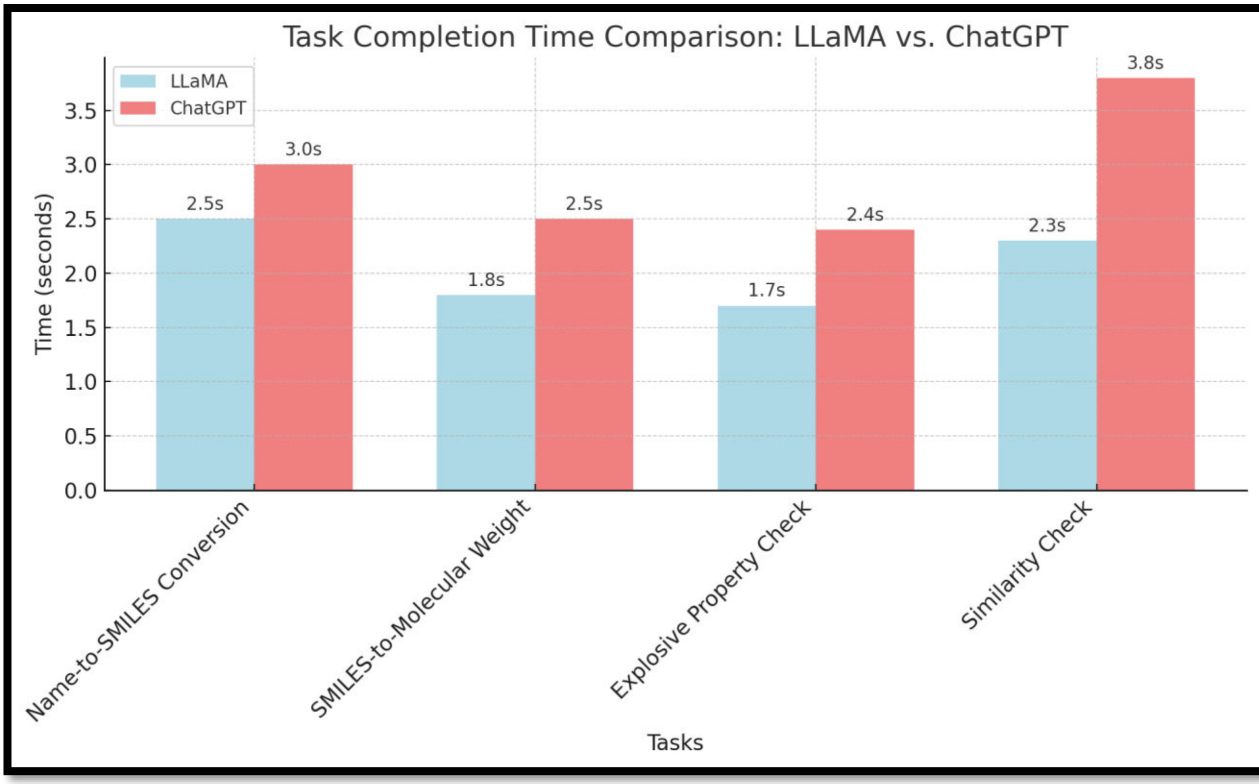
Figure 1 illustrates the comparative performance of LLaMA and ChatGPT across four key cheminformatics tasks: Name-to-SMILES conversion, SMILES-to-Molecular Weight calculation, Explosive Property Check, and Similarity Check. The chart highlights the significantly higher accuracy achieved by ChatGPT in all tasks, validating its suitability for complex cheminformatics applications despite the computational trade-offs..



3.2. Task Completion Time

Task	LLaMA Time (seconds)	ChatGPT Time (seconds)
Name-to-SMILES Conversions	2.5	3.0
SMILES-to-Molecular Weight	1.8	2.5
Explosive Property Checks	1.7	2.4
Similarity Checks	2.3	3.8

Figure 1: Task Completion Time Comparison



Visualization: ChatGPT exhibits slower task times due to additional reasoning steps but achieves higher accuracy

4. Discussion

4.1. Transition from LLaMA to ChatGPT

The decision to switch from LLaMA to ChatGPT was driven by the need to address reasoning gaps, as LLaMA often bypassed intermediate steps, such as directly generating molecular weights without resolving SMILES[41,43]. ChatGPT adhered to sequential reasoning, producing more reliable outputs. Additionally, the accuracy demands of cheminformatics tasks, which require precise computations, were better met by ChatGPT due to its fine-tuned instruction-following behavior.

4.2. Strengths of ChatGPT

ChatGPT demonstrated significant improvements in accuracy across all tasks, better understanding of domain-specific prompts, and adherence to the required output format without deviation.

4.3. Limitations

Despite its strengths, ChatGPT's higher computational requirements resulted in slower task completion. Additionally, some tasks required real-time API access, such as PubChem, introducing a dependency on internet connectivity.

5. Conclusion

This research demonstrates the potential of Large Language Models (LLMs) in automating cheminformatics tasks. While open-source models like LLaMA offer speed and efficiency, ChatGPT provides superior accuracy and reasoning capabilities, making it more suitable for complex chemical queries[18,20]. By integrating domain-specific tools and refining task prompts, we ensure reliable and scalable automation in computational chemistry.
Future work will focus on fine-tuning LLaMA models to align closer with domain-specific reasoning, reducing ChatGPT's latency through optimized query handling, and expanding task coverage to include retrosynthesis and drug discovery.

6. References

- *Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 2018.*
- *Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A., et al. Language models are few-shot learners. Advances in neural information processing systems 2020, 33, 1877–1901.*
- *Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M. S.; Bohg, J.; Bosselut, A.; Brunskill, E., et al. On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258 2021,*
- *Chowdhery, A.; Narang, S.; Devlin, J.; Bosma, M.; Mishra, G.; Roberts, A.; Barham, P.; Chung, H. W.; Sutton, C.; Gehrmann, S., et al. Palm: Scaling language modeling with pathways. arXiv preprint arXiv:2204.02311 2022,*
- *Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y. T.; Li, Y.; Lundberg, S., et al. Sparks of artificial general intelligence: Early experiments with gpt-4. arXiv preprint arXiv:2303.12712 2023,*
- *GitHub Copilot: Your AI pair programmer. copilot.github.com.*
- *Li, R. et al. StarCoder: may the source be with you! 2023.*
- *Ziegler, A.; Kalliamvakou, E.; Li, X. A.; Rice, A.; Rifkin, D.; Simister, S.; Sittampalam, G.; Aftandilian, E. Productivity assessment of neural code completion. 2022, 21–29.*
- *Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. Advances in neural information processing systems 2017, 30.*
- *Schick, T.; Dwivedi-Yu, J.; Dessì, R.; Raileanu, R.; Lomeli, M.; Zettlemoyer, L.; Cancedda, N.; Scialom, T. Toolformer: Language models can teach themselves to use tools. arXiv preprint arXiv:2302.04761 2023,*

- *Castro Nascimento, C. M.; Pimentel, A. S. Do Large Language Models Understand Chemistry? A Conversation with ChatGPT. Journal of Chemical Information and Modeling 2023, 63, 1649–1655.*
- *OpenAI, GPT-4 Technical Report. 2023.*
- *Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A., et al. Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems 2022, 35, 27730–27744.*
- *White, A. D.; Hocky, G. M.; Gandhi, H. A.; Ansari, M.; Cox, S.; Wellawatte, G. P.; Sasmal, S.; Yang, Z.; Liu, K.; Singh, Y., et al. Assessment of chemistry knowledge in large language models that generate code. Digital Discovery 2023.*
- *Lowe, D. M.; Corbett, P. T.; Murray-Rust, P.; Glen, R. C. Chemical Name to Structure: OPSIN, an Open Source Solution. Journal of Chemical Information and Modeling 2011, 51, 739–753, PMID: 21384929.*
- *Coley, C. W.; Barzilay, R.; Jaakkola, T. S.; Green, W. H.; Jensen, K. F. Prediction of organic reaction outcomes using machine learning. ACS central science 2017, 3, 434–443. 12.*
- *Coley, C. W.; Jin, W.; Rogers, L.; Jamison, T. F.; Jaakkola, T. S.; Green, W. H.; Barzilay, R.; Jensen, K. F. A graph-convolutional neural network model for the prediction of chemical reactivity. Chem. Sci. 2019, 10, 370–377.*
- *Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. ACS central science 2019, 5, 1572–1583.*
- *Pesciullesi, G.; Schwaller, P.; Laino, T.; Reymond, J.-L. Transfer learning enables the molecular transformer to predict regio-and stereoselective reactions on carbohydrates. Nat. Commun. 2020, 11, 1–8.*
- *Irwin, R.; Dimitriadis, S.; He, J.; Bjerrum, E. J. Chemformer: a pre-trained transformer for computational chemistry. Machine Learning: Science and Technology 2022, 3, 015022.*
- *Szymku´c, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-assisted synthetic planning: the end of the beginning. Angew. Chem.Int. Ed. 2016, 55, 5904–5937.*
- *Segler, M. H.; Preuss, M.; Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. Nature 2018, 555, 604–610.*
- *Coley, C. W.; Thomas, D. A.; Lummiss, J. A.; Jaworski, J. N.; Breen, C. P.; Schultz, V.; Hart, T.; Fishman, J. S.; Rogers, L.; Gao, H., et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. Science 2019, 365.*
- *Schwaller, P.; Petraglia, R.; Zullo, V.; Nair, V. H.; Haeuselmann, R. A.; Pisoni, R.; Bekas, C.; Iuliano, A.; Laino, T. Predicting retrosynthetic pathways using*

transformer-based models and a hyper-graph exploration strategy. Chemical science 2020, 11, 3316–3325.

- *Genheden, S.; Thakkar, A.; Chadimová, V.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning. J. Cheminf. 2020, 12, 1–9.*
- *Molga, K.; Szymku´ c, S.; Grzybowski, B. A. Chemist Ex Machina: Advanced Synthesis Planning by Computers. Acc. Chem. Res. 2021, 54, 1094–1106.*
- *Schwaller, P.; Vaucher, A. C.; Laplaza, R.; Bunne, C.; Krause, A.; Corminboeuf, C.; Laino, T. Machine intelligence for chemical reaction space. Wiley Interdisciplinary Reviews: Computational Molecular Science 2022, 12, e1604.*
- *Mayr, A.; Klambauer, G.; Unterthiner, T.; Hochreiter, S. DeepTox: toxicity prediction using deep learning. Frontiers in Environmental Science 2016, 3, 80.*
- *Yang, K.; Swanson, K.; Jin, W.; Coley, C.; Eiden, P.; Gao, H.; Guzman-Perez, A.; Hopper, T.; Kelley, B.; Mathea, M., et al. Analyzing learned molecular representations for property prediction. Journal of chemical information and modeling 2019, 59, 3370–3388.*
- *Chithrananda, S.; Grand, G.; Ramsundar, B. Chemberta: Large-scale self-supervised pretraining for molecular property prediction. arXiv preprint arXiv:2010.09885 2020.*
- *van Tilborg, D.; Alenicheva, A.; Grisoni, F. Exposing the limitations of molecular machine learning with activity cliffs. Journal of Chemical Information and Modeling 2022, 62, 5938–5951.*
- *Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Is GPT-3 all you need for low-data discovery in chemistry? 2023.*
- *Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. ACS Cent. Sci. 2018, 4, 268–276, PMID: 29532027.*
- *Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REINVENT 2.0: an AI tool for de novo drug design. Journal of chemical information and modeling 2020, 60, 5918–5922. 13.*
- *Tao, Q.; Xu, P.; Li, M.; Lu, W. Machine learning for perovskite materials design and discovery. npj Computational Materials 2021, 7, 1–18, Number: 1 Publisher: Nature Publishing Group.*
- *Gómez-Bombarelli, R. et al. Design of efficient molecular organic light-emitting diodes by a highthroughput virtual screening and experimental approach. Nature Materials 2016, 15, 1120–1127, Number: 10 Publisher: Nature Publishing Group.*

- *Shields, B. J.; Stevens, J.; Li, J.; Parasram, M.; Damani, F.; Alvarado, J. I. M.; Janey, J. M.; Adams, R. P.; Doyle, A. G. Bayesian reaction optimization as a tool for chemical synthesis. Nature 2021, 590, 89–96.*
- *Torres, J. A. G.; Lau, S. H.; Anchuri, P.; Stevens, J. M.; Tabora, J. E.; Li, J.; Borovika, A.; Adams, R. P.; Doyle, A. G. A Multi-Objective Active Learning Platform and Web App for Reaction Optimization. Journal of the American Chemical Society 2022, 144, 19999–20007.*
- *Ramos, M. C.; Michtavy, S. S.; Porosoff, M. D.; White, A. D. Bayesian Optimization of Catalysts With In-context Learning. arXiv preprint arXiv:2304.05341 2023.*
- *Marra, G.; Giannini, F.; Diligenti, M.; Gori, M. Integrating learning and reasoning with deep logic models. 2020, 517–532.*
- *Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Chi, E.; Le, Q.; Zhou, D. Chain of thought prompting elicits reasoning in large language models. arXiv preprint arXiv:2201.11903 2022,*
- *Ho, N.; Schmid, L.; Yun, S.-Y. Large Language Models Are Reasoning Teachers. arXiv preprint arXiv:2212.10071 2022,*
- *Yao, S.; Zhao, J.; Yu, D.; Du, N.; Shafran, I.; Narasimhan, K.; Cao, Y. React: Synergizing reasoning and acting in language models. arXiv preprint arXiv:2210.03629 2022,*
- *Zelikman, E.; Wu, Y.; Mu, J.; Goodman, N. Star: Bootstrapping reasoning with reasoning. Advances in Neural Information Processing Systems 2022, 35, 15476–15488.*
- *Zhao, Z.-W.; del Cueto, M.; Troisi, A. Limitations of machine learning models when predicting compounds with completely new chemistries: possible improvements applied to the discovery of new non-fullerene acceptors. Digital Discovery 2022, 1, 266–276.*
- *Vaucher, A. C.; Schwaller, P.; Geluykens, J.; Nair, V. H.; Iuliano, A.; Laino, T. Inferring experimental procedures from text-based representations of chemical reactions. Nature communications 2021, 12, 2573.*
- *Schwaller, P.; Probst, D.; Vaucher, A. C.; Nair, V. H.; Kreutter, D.; Laino, T.; Reymond, J.-L. Mapping the space of chemical reactions using attention-based neural networks. Nature machine intelligence 2021, 3, 144–152.*
- *rxn4Chemistry, rxn4Chemistry. github.com, 2020; Accessed: April 2023.*
- *Thakkar, A.; Kogej, T.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. J. Datasets and their influence on the development of computer assisted synthesis planning tools in the pharmaceutical domain. Chemical science 2020, 11, 154–168.*
- *Thakkar, A.; Selmi, N.; Reymond, J.-L.; Engkvist, O.; Bjerrum, E. J. "Ring breaker": neural network driven synthesis prediction of the ring system chemical space. Journal of medicinal chemistry 2020, 63, 8791–8808.*

- *Yang, Z.; Li, L.; Wang, J.; Lin, K.; Azarnasab, E.; Ahmed, F.; Liu, Z.; Liu, C.; Zeng, M.; Wang, L. MM-REACT: Prompting ChatGPT for Multimodal Reasoning and Action. arXiv preprint arXiv:2303.11381 2023,*
- *Shen, Y.; Song, K.; Tan, X.; Li, D.; Lu, W.; Zhuang, Y. Hugging GPT: Solving AI Tasks with ChatGPT and its Friends in Hugging Face. 2023. 14*
- *Karpas, E.; Abend, O.; Belinkov, Y.; Lenz, B.; Lieber, O.; Ratner, N.; Shoham, Y.; Bata, H.; Levine, Y.; Leyton-Brown, K., et al. MRKL Systems: A modular, neuro-symbolic architecture that combines large language models, external knowledge sources and discrete reasoning. arXiv preprint arXiv:2205.00445 2022.*
- *Boiko, D. A.; MacKnight, R.; Gomes, G. Emergent autonomous scientific research capabilities of large language models. arXiv preprint 2023.*
- *IBM RoboRXN | Science | IBM Research — research.ibm.com. research.ibm.com [Accessed 12-May-2023].*
- *Wittkopp, A.; Schreiner, P. R. Metal-Free, Noncovalent Catalysis of Diels–Alder Reactions by Neutral Hydrogen Bond Donors in Organic Solvents and in Water. Chemistry– A European Journal 2003, 9, 407–414.*
- *Schreiner, P. R.; Wittkopp, A. H-Bonding Additives Act Like Lewis Acid Catalysts. Organic Letters 2002, 4, 217–220, Publisher: American Chemical Society.*
- *Herrera, R. P.; Sgarzani, V.; Bernardi, L.; Ricci, A. Catalytic Enantioselective Friedel–Crafts Alkylation of Indoles with Nitroalkenes by Using a Simple Thiourea Organocatalyst. AngewandteChemie International Edition 2005, 44, 6576–6579.*
- *Okino, T.; Hoashi, Y.; Takemoto, Y. Enantioselective Michael Reaction of Malonates to Nitroolefins Catalyzed by Bifunctional Organocatalysts. Journal of the American Chemical Society 2003, 125, 12672–12673, Publisher: American Chemical Society.*
- *Lowe, D. M. Extraction of chemical structures and reactions from the literature. Ph.D. thesis, University of Cambridge, 2012.*
- *Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. Chemical science 2018, 9, 513–530.*
- *Liu, Y.; Iter, D.; Xu, Y.; Wang, S.; Xu, R.; Zhu, C. GPTEval: NLG Evaluation using GPT-4 with Better Human Alignment. arXiv preprint arXiv:2303.16634 2023,*
- *Eloundou, T.; Manning, S.; Mishkin, P.; Rock, D. Gpts are gpts: An early look at the labor market impact potential of large language models. arXiv preprint arXiv:2303.10130 2023,*

- *Grzybowski, B. A.; Badowski, T.; Molga, K.; Szymku´ c, S. Network search algorithms and scoring functions for advanced-level computerized synthesis planning. WIREs Computational Molecular Science 2023, 13, e1630.*
- *Thakkar, A.; Johansson, S.; Jorner, K.; Buttar, D.; Reymond, J.-L.; Engkvist, O. Artificial intelligence and automation in computer aided synthesis planning. Reaction chemistry & engineering 2021, 6, 27–51.*
- *Urbina, F.; Lentzos, F.; Invernizzi, C.; Ekins, S. Dual use of artificial-intelligence-powered drug discovery. Nature Machine Intelligence 2022, 4, 189–191.*
- *Urbina, F.; Lentzos, F.; Invernizzi, C.; Ekins, S. A teachable moment for dual-use. Nature machine intelligence 2022, 4, 607–607.*
- *Campbell, Q. L.; Herington, J.; White, A. D. Censoring chemical data to mitigate dual use risk. arXiv preprint arXiv:2304.10510 2023,*
- *Gao, L.; Schulman, J.; Hilton, J. Scaling Laws for Reward Model Overoptimization. arXiv preprint arXiv:2210.10760 2022,*
- *Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I., et al. Improving language understanding by generative pre-training. 2018,*
- *Li, B.; Qi, P.; Liu, B.; Di, S.; Liu, J.; Pei, J.; Yi, J.; Zhou, B. Trustworthy AI: From Principles to Practices. ACM Computing Surveys 2021, 55, 1– 46. 15*
- *Hocky, G. M.; White, A. D. Natural language processing models that automate programming will transform chemistry research and teaching. Digital Discovery 2022, 1, 79–83.*
- *Henderson, P.; Li, X.; Jurafsky, D.; Hashimoto, T.; Lemley, M. A.; Liang, P. Foundation Models and Fair Use. arXiv preprint arXiv:2303.15715 2023.*
- *Askell, A.; Brundage, M.; Hadfield, G. The Role of Cooperation in Responsible AI Development. 2019.*
- *Neufville, R. d.; Baum, S. D. Collective action on artificial intelligence: A primer and review. Technology in Society 2021, 66, 101649.*
- *Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.-A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; Rodriguez, A.; Joulin, A.; Grave, E.; Lample, G. LLaMA: Open and Efficient Foundation Language Models. 2023.*
- *Chiang, W.-L.; Li, Z.; Lin, Z.; Sheng, Y.; Wu, Z.; Zhang, H.; Zheng, L.; Zhuang, S.; Zhuang, Y.; Gonzalez, J. E.; Stoica, I.; Xing, E. P. Vicuna: An Open-Source Chatbot Impressing GPT-4 with 90%* ChatGPT Quality. 2023; lmsys.org.*
- *Mukherjee, S.; Mitra, A.; Jawahar, G.; Agarwal, S.; Palangi, H.; Awadallah, A. Orca: Progressive Learning from Complex Explanation Traces of GPT-4. 2023.*
- *Chase, H. LangChain. 2022; github.com.*
- *Press, O.; Zhang, M.; Min, S.; Schmidt, L.; Smith, N. A.; Lewis, M. Measuring and Narrowing the Compositionality Gap in Language Models. arXiv preprint arXiv:2210.03350 2022,*

- *SerpAPI, SerpAPI- Google Search Results API. 2023; serpapi.com.*
- *Neelakantan, A.; Xu, T.; Puri, R.; Radford, A.; Han, J. M.; Tworek, J.; Yuan, Q.; Tezak, N.; Kim, J. W.; Hallacy, C., et al. Text and code embeddings by contrastive pre-training. arXiv preprint arXiv:2201.10005 2022,*
- *Johnson, J.; Douze, M.; Jégou, H. Billion-scale similarity search with GPUs. IEEE Transactions on Big Data 2019, 7, 535–547.*
- *ChemSpace, ChemSpace- Provider of Chemical Building Blocks, Fragment Libraries, and Screening Compounds. chem-space.com, 2023.*
- *National Center for Biotechnology Information, PubChem. pubchem. ncbi. nlm. nih. gov, 2023.*
- *Lowe, D. M.; Corbett, P. T.; Murray-Rust, P.; Glen, R. C. Chemical name to structure: OPSIN, an open source solution. 2011.*
- *Medina, J.; White, A. D. Bloom filters for molecules. arXiv preprint arXiv:2304.05386 2023.*
- *Irwin, J. J.; Tang, K. G.; Young, J.; Dandarchuluun, C.; Wong, B. R.; Khurelbaatar, M.; Moroz, Y. S.; Mayfield, J.; Sayle, R. A. ZINC20—a free ultralarge-scale chemical database for ligand discovery. Journal of chemical information and modeling 2020, 60, 6065–6073.*
- *Chemical Abstracts Service, CAS Registry Number. www.cas.org, Accessed: April 2023.*
- *TT, T. An elementary mathematical theory of classification and prediction; 1958.*
- *Rogers, D.; Hahn, M. Extended-connectivity fingerprints. Journal of chemical information and modeling 2010, 50, 742–754.*
- *White, Andrew D, SynSpace. github.com, Accessed: April 2023; GitHub repository.*
- *Wellawatte, G. P.; Seshadri, A.; White, A. D. Model agnostic generation of counterfactual explanations for molecules. Chemical science 2022, 13, 3697–3705. 16*
- *Med Chem Comm. Accessed: 5 April 2023,*
- *Yang, Q.; Sresht, V.; Bolgar, P.; Hou, X.; Klug-McLeod, J. L.; Butler, C. R., et al. Molecular transformer unifies reaction prediction and retrosynthesis across pharma chemical space. Chemical communications 2019, 55, 12152–12155.*
- *Mcule, Purchasable Mcule. purchasable.mcule.com, Accessed: April 2023.*
- *RDKit: Open-source cheminformatics. www.rdkit.org, [Online; accessed April 2023].*
- *Chemical Weapons Convention, Annex on Chemicals, B. Schedules of Chemicals. www. opcw.org, Accessed: May 2023.*
- *Group, A. Export Control List: Chemical Weapons Precursors. australiagroup.net, Accessed: May 2023.*

- *Next Move Software, Name Rxn. www.nextmovesoftware.com, last accessed 2020/11/07; Accessed: 5 April 2023.*
- *Carey, J. S.; Laffan, D.; Thomson, C.; Williams, M. T. Analysis of the reactions used for the preparation of drug candidate molecules. Organic & biomolecular chemistry 2006, 4, 2337–2347.*
- *Vaucher, A. C.; Schwaller, P.; Geluykens, J.; Nair, V. H.; Iuliano, A.; Laino, T. Inferring experimental procedures from text-based representations of chemical reactions. Nature Communications 2021, 12, 2573, Number: 1 Publisher: Nature Publishing Group.*
- *Edwards, C.; Lai, T.; Ros, K.; Honke, G.; Cho, K.; Ji, H. Translation between Molecules and Natural Language. 2022.*
- *Christofidellis, D.; Giannone, G.; Born, J.; Winther, O.; Laino, T.; Manica, M. Unifying Molecular and Textual Representations via Multi-task Language Modelling. arXiv preprint arXiv:2301.12586 2023,*
- *Azamfirei, R.; Kudchadkar, S. R.; Fackler, J. Large language models and the perils of their hallucinations. Critical Care 2023, 27, 1–2.*
- *Khapra, M. M.; Sai, A. B. A tutorial on evaluation metrics used in natural language generation. 2021, 15–19.*
- *Melis, G.; Dyer, C.; Blunsom, P. On the state of the art of evaluation in neural language models. arXiv preprint arXiv:1707.05589 2017,*
- *Flam-Shepherd, D.; Zhu, K.; Aspuru-Guzik, A. Language models can learn complex molecular distributions. Nature Communications 2022, 13, 3293.*