References

[1] S. Zhu, X. Liu, and G. Khalighinejad, "Chemqa: a multimodal question-and-answering dataset on chemistry reasoning." https://huggingface.co/datasets/shangzhu/ChemQA, 2024.

[2] OpenAI, "Gpt-4 technical report," 2024.

[3] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," 2021.

[4] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. a. Persson, "The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials*, vol. 1, no. 1, p. 011002, 2013.

[5] D. G. Paul, H. Zhu, and I. Bayley, "Benchmarks and metrics for evaluations of code generation: A critical review," in *2024 IEEE International Conference on Artificial Intelligence Testing (AITest)*, pp. 87–94, IEEE, 2024.

[6] A. Mirza, N. Alampara, S. Kunchapu, M. Ríos-García, B. Emoekabu, A. Krishnan, T. Gupta, M. Schilling-Wilhelmi, M. Okereke, A. Aneesh, *et al.*, "Are large language models superhuman chemists?," *arXiv preprint arXiv:2404.01475*, 2024.

[7] J. Zhang, X. Xiang, B. Xu, S. Huang, Y. Xiong, S. Ma, H. Fu, Y. Ma, H. Chen, Z. Wu, and S. Zhao, "Rational design of high-entropy ceramics based on machine learning – a critical review," *Current Opinion in Solid State and Materials Science*, vol. 27, p. 101057, 4 2023.

[8] Clarivate Analytics, "Web of science." https://www.webofscience.com, 2025. Accessed: 2025-11-05.

[9] Mistral AI, "Mistral-large:123b-instruct-2407-q4_0." https://mistral.ai/news/mistral-large/, 2024. Large Language Model by Mistral AI.

[10] Open Source Science (OSS), "Gpt-oss:120b." https://huggingface.co/oss/gpt-oss-120b, 2024. Open large language model for scientific applications.

[11] M. Szymański, R. V. Vlasov, *et al.*, "Mendeleev – a python resource for properties of chemical elements, ions and isotopes," *Journal of Open Source Software*, vol. 3, no. 32, p. 1113, 2018.

[12] NOMAD Laboratory Consortium, "The nomad laboratory – fair data infrastructure for materials science." https://nomad-lab.eu, 2023. FAIR data platform for materials science.

[13] Thermfact/CRCT and GTT-Technologies, "Factsage thermochemical software and databases." https://www.factsage.com, 2022. Thermochemical calculation and database system.

[14] P. Dai, B. C. Chakoumakos, G. F. Sun, K. W. Wong, Y. Xin, and D. F. Lu, "Synthesis and neutron powder diffraction study of the superconductor HgBa2Ca2Cu3O8 + δ by Tl substitution," *Physica C: Superconductivity*, vol. 243, pp. 201–206, Mar. 1995.

[15] Gemini Team, Google, "Gemini: A Family of Highly Capable Multimodal Models," *arXiv preprint arXiv:2312.11805v5*, 2025.

[16] H. Park, A. Onwuli, and A. Walsh, "Exploration of crystal chemical space using text-guided generative artificial intelligence," *Nature Communications*, vol. 16, p. 4379, 2025.

[17] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, and et al., "The ai revolution in science," *Science*, vol. 359, no. 6379, p. eaao0185, 2018.

[18] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, and et al., "Language models are few-shot learners," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1877–1901, 2020.