[1]        A. Toosi, A. Bottino, B. Saboury, E. Siegel, and A. Rahmim, “A brief history of AI: how to prevent another winter (a critical review),” *PET Clinics*, vol. 16, no. 4, pp. 449–469, Oct. 2021, doi: 10.1016/j.cpet.2021.07.001.

[2]        SITNFlash, “The History of Artificial Intelligence,” Science in the News. Accessed: Apr. 15, 2024. [Online]. Available: <https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>

[3]        S. Bubeck *et al.*, “Sparks of Artificial General Intelligence: Early experiments with GPT-4.” arXiv, Apr. 13, 2023. Accessed: Apr. 14, 2024. [Online]. Available: <http://arxiv.org/abs/2303.12712>

[4]        J. E. T. Taylor and G. W. Taylor, “Artificial cognition: How experimental psychology can help generate explainable artificial intelligence,” *Psychon Bull Rev*, vol. 28, no. 2, pp. 454–475, Apr. 2021, doi: 10.3758/s13423-020-01825-5.

[5]        H. Zhao *et al.*, “Explainability for Large Language Models: A Survey,” *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 2, pp. 1–38, Apr. 2024, doi: 10.1145/3639372.

[6]        A. Notovich, H. Chalutz-Ben Gal, and I. Ben-Gal, “Explainable Artificial Intelligence (XAI): Motivation, Terminology, and Taxonomy,” in *Machine Learning for Data Science Handbook: Data Mining and Knowledge Discovery Handbook*, L. Rokach, O. Maimon, and E. Shmueli, Eds., Cham: Springer International Publishing, 2023, pp. 971–985. doi: 10.1007/978-3-031-24628-9\_41.

[7]        A. Vaswani *et al.*, “Attention is All you Need,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Apr. 16, 2024. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>

[8]        A. Pagnoni, V. Balachandran, and Y. Tsvetkov, “Understanding Factuality in Abstractive Summarization with FRANK: A Benchmark for Factuality Metrics.” arXiv, Jul. 23, 2021. Accessed: Apr. 19, 2024. [Online]. Available: <http://arxiv.org/abs/2104.13346>

[9]        B. Peng *et al.*, “Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback.” arXiv, Mar. 08, 2023. Accessed: Feb. 03, 2024. [Online]. Available: <http://arxiv.org/abs/2302.12813>

[10]      A. Aynetdinov and A. Akbik, “SemScore: Automated Evaluation of Instruction-Tuned LLMs based on Semantic Textual Similarity.” arXiv, Feb. 05, 2024. Accessed: Apr. 19, 2024. [Online]. Available: <http://arxiv.org/abs/2401.17072>

[11]      W. Wang, B. Haddow, A. Birch, and W. Peng, “Assessing the Reliability of Large Language Model Knowledge.” arXiv, Oct. 15, 2023. Accessed: Jan. 30, 2024. [Online]. Available: <http://arxiv.org/abs/2310.09820>

[12]      J.-Y. Yao, K.-P. Ning, Z.-H. Liu, M.-N. Ning, and L. Yuan, “LLM Lies: Hallucinations are not Bugs, but Features as Adversarial Examples.” arXiv, Oct. 04, 2023. Accessed: Apr. 14, 2024. [Online]. Available: <http://arxiv.org/abs/2310.01469>

[13]      C. Zhou *et al.*, “Detecting Hallucinated Content in Conditional Neural Sequence Generation,” arXiv.org. Accessed: Jan. 30, 2024. [Online]. Available: <https://arxiv.org/abs/2011.02593v3>

[14]      K. Zhu *et al.*, “PromptBench: Towards Evaluating the Robustness of Large Language Models on Adversarial Prompts.” arXiv, Oct. 18, 2023. Accessed: Apr. 19, 2024. [Online]. Available: <http://arxiv.org/abs/2306.04528>