READING FOR GENERATIVE MODELS AND LLMS

[2, 1, 4, 20, 13, 19, 27, 5, 10, 21, 23, 6, 26, 9, 18, 22, 11, 14, 25, 12, 3, 16, 15, 8, 17, 7, 24]

References

1. Emily M Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, *On the dangers of stochastic parrots: Can language models be too big?*, Proceedings of the 2021 ACM conference on fairness, accountability, and transparency, 2021, pp. 610–623.
2. Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al., *Language models are few-shot learners*, Advances in neural information processing systems **33** (2020), 1877–1901.
3. Deng Cai, Yan Wang, Lemao Liu, and Shuming Shi, *Recent advances in retrieval-augmented text generation*, Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 3417–3419.
4. Simon Frieder, Luca Pinchetti, Ryan-Rhys Griffiths, Tommaso Salvatori, Thomas Lukasiewicz, Philipp Petersen, and Julius Berner, *Mathematical capabilities of chatgpt*, Advances in Neural Information Processing Systems **36** (2024).
5. Karl J Friston, Maxwell JD Ramstead, Alex B Kiefer, Alexander Tschantz, Christopher L Buckley, Mahault Albarracin, Riddhi J Pitliya, Conor Heins, Brennan Klein, Beren Millidge, et al., *Designing ecosystems of intelligence from first principles*, Collective Intelligence **3** (2024), no. 1, 26339137231222481.
6. Leo Gao, Stella Biderman, Sid Black, Laurence Golding, Travis Hoppe, Charles Foster, Jason Phang, Horace He, Anish Thite, Noa Nabeshima, et al., *The pile: An 800gb dataset of diverse text for language modeling*, arXiv preprint arXiv:2101.00027 (2020).
7. Xiaoxin He, Yijun Tian, Yifei Sun, Nitesh V Chawla, Thomas Laurent, Yann LeCun, Xavier Bresson, and Bryan Hooi, *G-retriever: Retrieval-augmented generation for textual graph understanding and question answering*, arXiv preprint arXiv:2402.07630 (2024).
8. Geoffrey E Hinton, Peter Dayan, Brendan J Frey, and Radford M Neal, *The" wake-sleep" algorithm for unsupervised neural networks*, Science **268** (1995), no. 5214, 1158–1161.
9. Cheng-Yu Hsieh, Chun-Liang Li, Chih-Kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alexander Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister, *Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes*, arXiv preprint arXiv:2305.02301 (2023).
10. Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen, *Lora: Low-rank adaptation of large language models*, arXiv preprint arXiv:2106.09685 (2021).
11. Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al., *Mistral 7b*, arXiv preprint arXiv:2310.06825 (2023).
12. Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al.,

Date: February 20, 2024.

Retrieval-augmented generation for knowledge-intensive nlp tasks, Advances in Neural Information Processing Systems **33** (2020), 9459–9474.
13. Shayne Longpre, Le Hou, Tu Vu, Albert Webson, Hyung Won Chung, Yi Tay, Denny Zhou, Quoc V Le, Barret Zoph, Jason Wei, et al., *The flan collection: Designing data and methods for effective instruction tuning*, arXiv preprint arXiv:2301.13688 (2023).
14. Rithesh Murthy, Shelby Heinecke, Juan Carlos Niebles, Zhiwei Liu, Le Xue, Weiran Yao, Yihao Feng, Zeyuan Chen, Akash Gokul, Devansh Arpit, et al., *Rex: Rapid exploration and exploitation for ai agents*, arXiv preprint arXiv:2307.08962 (2023).
15. Dae Hoon Park, ChengXiang Zhai, and Lifan Guo, *Speclda: Modeling product reviews and specifications to generate augmented specifications*, Proceedings of the 2015 SIAM International Conference on Data Mining, SIAM, 2015, pp. 837–845.
16. Md Rizwan Parvez, Wasi Uddin Ahmad, Saikat Chakraborty, Baishakhi Ray, and Kai-Wei Chang, *Retrieval augmented code generation and summarization*, arXiv preprint arXiv:2108.11601 (2021).
17. Thomas Pouplin, Hao Sun, Samuel Holt, and Mihaela Van der Schaar, *Retrieval-augmented thought process as sequential decision making*, arXiv preprint arXiv:2402.07812 (2024).
18. Reid Pryzant, Dan Iter, Jerry Li, Yin Tat Lee, Chenguang Zhu, and Michael Zeng, *Automatic prompt optimization with" gradient descent" and beam search*, arXiv preprint arXiv:2305.03495 (2023).
19. Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al., *Language models are unsupervised multitask learners*, OpenAI blog **1** (2019), no. 8, 9.
20. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer, *High-resolution image synthesis with latent diffusion models*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 10684–10695.
21. Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman, *Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation*, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 22500–22510.
22. Simeng Sun, Yang Liu, Shuohang Wang, Chenguang Zhu, and Mohit Iyyer, *Pearl: Prompting large language models to plan and execute actions over long documents*, arXiv preprint arXiv:2305.14564 (2023).
23. Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al., *Llama: Open and efficient foundation language models*, arXiv preprint arXiv:2302.13971 (2023).
24. Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi, *Self-instruct: Aligning language model with self generated instructions*, arXiv preprint arXiv:2212.10560 (2022).
25. Chenshuang Zhang, Chaoning Zhang, Sheng Zheng, Mengchun Zhang, Maryam Qamar, Sung-Ho Bae, and In So Kweon, *A survey on audio diffusion models: Text to speech synthesis and enhancement in generative ai*, arXiv preprint arXiv:2303.13336 **2** (2023).
26. Lvmin Zhang and Maneesh Agrawala, *Adding conditional control to text-to-image diffusion models*, arXiv preprint arXiv:2302.05543 (2023).
27. Zhuosheng Zhang, Aston Zhang, Mu Li, Hai Zhao, George Karypis, and Alex Smola, *Multimodal chain-of-thought reasoning in language models*, arXiv preprint arXiv:2302.00923 (2023).