experts," *arXiv preprint arXiv:2304.04370*, 2023.

[18] H. Chase, "Langchain," https://github.com/langchain-ai/langchain.

[19] Y. Zhu, H. Yuan, S. Wang, J. Liu, W. Liu, C. Deng, Z. Dou, and J.-R. Wen, "Large language models for information retrieval: A survey," *arXiv preprint arXiv:2308.07107*, 2023.

[20] V. Karpukhin, B. Oğuz, S. Min, P. Lewis, L. Wu, S. Edunov, D. Chen, and W.-t. Yih, "Dense passage retrieval for open-domain question answering," *arXiv preprint arXiv:2004.04906*, 2020.

[21] W. Wang, F. Wei, L. Dong, H. Bao, N. Yang, and M. Zhou, "Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers," *CoRR*, vol. abs/2002.10957, 2020.

[22] W. Zhong, L. Guo, Q. Gao, and Y. Wang, "Memorybank: Enhancing large language models with long-term memory," *arXiv preprint arXiv:2305.10250*, 2023.

[23] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, no. 1, p. 107–113, jan 2008.

[24] N. F. Liu, K. Lin, J. Hewitt, A. Paranjape, M. Bevilacqua, F. Petroni, and P. Liang, "Lost in the middle: How language models use long contexts," *arXiv preprint arXiv:2307.03172*, 2023.

[25] "Vosk speech recognition toolbox," https://github.com/alphacep/vosk-api, accessed: 2023-10-27.

[26] "Langchain openapi toolkit," https://python.langchain.com/docs/integrations/toolkits/openapi, accessed: 2023-10-26.

[27] "Smartthings api," https://developer.smartthings.com/docs/api/public, accessed: 2023-10-26.

[28] "Sourcegraph cody ai," https://github.com/sourcegraph/cody, accessed: 2023-10-26.

[29] "Gpt engineer," https://github.com/AntonOsika/gpt-engineer, accessed: 2023-10-26.