[7] Diwali, A., Saeedi, K., Dashtipour, K., Gogate, M., Cambria, E., and Hussain, A. Sentiment analysis meets explainable artificial intelligence: A survey on explainable sentiment analysis. In *IEEE Transactions on Affective Computing*, 2023.

[8] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., 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.

[9] Reynolds, L. and McDonell, K. Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. *arXiv preprint arXiv:2102.07350*, 2021.

[10] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35 (2022): 27730–27744.

[11] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Xia, F., Chi, E., Le, Q. V., and Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems*, 2022.

[12] Bender, E. M., Gebru, T., McMillan-Major, A., and Shmitchell, S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2021.

[13] Schick, T. and Schütze, H. Exploiting Cloze Questions for Few-Shot Text Classification and Natural Language Inference. *arXiv preprint arXiv:2101.06826*, 2021.

[14] Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., et al. Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*, 2021.

[15] Ribeiro, M. T., Wu, T., Guestrin, C., and Singh, S. Beyond accuracy: Behavioral testing of NLP models with CheckList. *arXiv preprint arXiv:2005.04118*, 2020.

[16] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., and Galstyan, A. A Survey on Bias and Fairness in Machine Learning. *arXiv preprint arXiv:2106.08982*, 2021.

[17] Sheng, E., Chang, K. W., Natarajan, P., and Peng, N. Towards Controllable Bias in Natural Language Generation. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021.