Section VI: References

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., . . . Amodei, D. (2020). Language models are few-shot learners (arXiv:2005.14165). arXiv. https://arxiv.org/abs/2005.14165

Dey, N., Gosal, G., Zhiming, Chen, Khachane, H., Marshall, W., Pathria, R., Tom, M., & Hestness, J. (2023). *Cerebras-GPT: Open compute-optimal language models trained on the Cerebras wafer-scale cluster* (arXiv:2304.03208). arXiv. https://arxiv.org/abs/2304.03208

Hoffmann, J., Borgeaud, S., Mensch, A., Buchatskaya, E., Cai, T., Rutherford, E., Casas, D. de L., Hendricks, L. A., Welbl, J., Clark, A., Hennigan, T., Noland, E., Millican, K., Driessche, G. van den, Damoc, B., Guy, A., Osindero, S., Simonyan, K., Elsen, E., … Sifre, L. (2022). *Training compute-optimal large language models* (arXiv:2203.15556). arXiv. https://arxiv.org/abs/2203.15556

Kreutzer, J., Caswell, I., Wang, L., Wahab, A., van Esch, D., Ulzii-Orshikh, N., Tapo, A., Subramani, N., Sokolov, A., Sikasote, C., Setyawan, M., Sarin, S., Samb, S., Sagot, B., Rivera, C., Rios, A., Papadimitriou, I., Osei, S., Suarez, P. O., … Adeyemi, M. (2021). *Quality at a glance: An audit of web-crawled multilingual datasets*. (arXiv:2103.12028). arXiv. https://arxiv.org/abs/2103.12028

Lee, K., Ippolito, D., Nystrom, A., Zhang, C., Eck, D., Callison-Burch, C., & Carlini, N. (2021). *Deduplicating training data makes language models better* (arXiv:2107.06499). arXiv. https://arxiv.org/abs/2107.06499

Longpre, S., Yauney, G., Reif, E., Lee, K., Roberts, A., Zoph, B., Zhou, D., Wei, J., Robinson, K., Mimno,

D., & Ippolito, D. (2023). *A pretrainer's guide to training data: Measuring the effects of data age, domain coverage, quality, & toxicity*. (arXiv:2305.13169). arXiv. https://arxiv.org/abs/2305.13169

Sun, H., Pei, J., Choi, M., & Jurgens, D. (2023). *Aligning with whom? Large language models have gender and racial biases in subjective NLP tasks* (arXiv:2311.09730). arXiv. https://arxiv.org/abs/2311.09730

Xue, L., Constant, N., Roberts, A., Kale, M., Siddhant, A., Barua, A., & Raffel, C. (2020). *MT5: A massively multilingual pre-trained text-to-text transformer* (arXiv:2010.11934). arXiv. https://arxiv.org/abs/2010.11934

Zhang, Y., Li, Y., Cui, L., Cai, D., Liu, L., Fu, T., Huang, X., Zhao, E., Zhang, Y., Chen, Y., Wang, L., Luu, A. T., Bi, W., Shi, F., & Shi, S. (2023). *Siren's song in the AI ocean: A survey on hallucination in large language models* (arXiv:2309.01219). arXiv. https://arxiv.org/abs/2309.01219

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2023). *A survey of large language models*. (arXiv:2303.18223). arXiv. https://arxiv.org/abs/2303.18223