

DAFTAR PUSTAKA

- [1] J. Xiao and Z. Zhou, "Research progress of rnn language model," in *2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2020, pp. 1285–1288.
- [2] C. Gulcehre, O. Firat, K. Xu, K. Cho, and Y. Bengio, "On integrating a language model into neural machine translation," *Computer Speech & Language*, vol. 45, pp. 137–148, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0885230816301395>
- [3] J. Pilault, R. Li, S. Subramanian, and C. Pal, "On extractive and abstractive neural document summarization with transformer language models," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, B. Webber, T. Cohn, Y. He, and Y. Liu, Eds. Online: Association for Computational Linguistics, Nov. 2020, pp. 9308–9319. [Online]. Available: <https://aclanthology.org/2020.emnlp-main.748>
- [4] S. Kumawat, I. Yadav, N. Pahal, and D. Goel, "Sentiment analysis using language models: A study," in *2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2021, pp. 984–988.
- [5] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.
- [6] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proceedings of the 28th international conference on machine learning (ICML-11)*, 2011, pp. 1017–1024.
- [7] J. Nowak, A. Taspinar, and R. Scherer, "Lstm recurrent neural networks for short text and sentiment classification," in *Artificial Intelligence and Soft Computing*, L. Rutkowski, M. Korytkowski, R. Scherer, R. Tadeusiewicz, L. A. Zadeh, and J. M. Zurada, Eds. Cham: Springer International Publishing, 2017, pp. 553–562.
- [8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
- [9] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [10] K. Zeng and I. Paik, "Dynamic service recommendation using lightweight bert-based service embedding in edge computing," in *2021 IEEE 14th International*

Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoc), 2021, pp. 182–189.

- [11] D. Narayanan, M. Shoeybi, J. Casper, P. LeGresley, M. Patwary, V. Korthikanti, D. Vainbrand, P. Kashinkunti, J. Bernauer, B. Catanzaro, A. Phanishayee, and M. Zaharia, “Efficient large-scale language model training on gpu clusters using megatron-lm,” in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ser. SC '21. New York, NY, USA: Association for Computing Machinery, 2021. [Online]. Available: <https://doi.org/10.1145/3458817.3476209>
- [12] E. Ajao, “Google turns bard ai into gemini, launches gemini advanced,” *TechTarget*, 2024. [Online]. Available: <https://www.techtarget.com/searchenterpriseai/news/366569452/Google-turns-Bard-AI-into-Gemini-launches-Gemini-Advanced>
- [13] A. Izzak *et al.*, “Bilingualisme dalam perspektif pengembangan bahasa indonesia,” *Mabasan*, vol. 3, no. 1, pp. 16–29, 2009.
- [14] A. Setiawan and U. K. Luthfiyani, “Penggunaan chatgpt untuk pendidikan di era education 4.0: Usulan inovasi meningkatkan keterampilan menulis,” *JURNAL PETISI (Pendidikan Teknologi Informasi)*, vol. 4, no. 1, pp. 49–58, 2023.
- [15] M. G. An'ars, N. Hendrastuty, D. Damayanti, and A. D. Putra, “Perancangan teks promosi umkm sikop arrum batik menggunakan program berbasis ai chatgpt,” *Jurnal Pengabdian Kepada Masyarakat (JPKM) TABIKPUN*, vol. 4, no. 1, pp. 11–18, 2023.
- [16] R. Darman, “Peran chatgpt sebagai artificial intelligence dalam menyelesaikan masalah pertanahan dengan metode studi kasus dan black box testing,” *Tunas Agraria*, vol. 7, no. 1, pp. 18–46, 2024.
- [17] Rachmatunnisa, “Kominfo dukung pengembangan 'chat gpt' versi lokal,” *Detik*, 2023. [Online]. Available: <https://inet.detik.com/law-and-policy/d-7065641/kominfo-dukung-pengembangan-chat-gpt-versi-lokal>
- [18] L. Chen, M. Zaharia, and J. Zou, “How is chatgpt’s behavior changing over time?” *arXiv preprint arXiv:2307.09009*, 2023.
- [19] S. Cahyawijaya, G. I. Winata, B. Wilie, K. Vincentio, X. Li, A. Kuncoro, S. Ruder, Z. Y. Lim, S. Bahar, M. L. Khodra *et al.*, “Indonlg: Benchmark and resources for evaluating indonesian natural language generation,” *arXiv preprint arXiv:2104.08200*, 2021.
- [20] L. Owen, V. Tripathi, A. Kumar, and B. Ahmed, “Komodo: A linguistic expedition into indonesia’s regional languages,” *arXiv preprint arXiv:2403.09362*, 2024.
- [21] S. Cahyawijaya, H. Lovenia, F. Koto, R. A. Putri, E. Dave, J. Lee, N. Shadieq, W. Cenggoro, S. M. Akbar, M. I. Mahendra *et al.*, “Cendol: Open instruction-tuned generative large language models for indonesian languages,” *arXiv preprint arXiv:2404.06138*, 2024.

- [22] W.-L. Chiang, L. Zheng, Y. Sheng, A. N. Angelopoulos, T. Li, D. Li, H. Zhang, B. Zhu, M. Jordan, J. E. Gonzalez *et al.*, “Chatbot arena: An open platform for evaluating llms by human preference,” *arXiv preprint arXiv:2403.04132*, 2024.
- [23] M. Wortsman, G. Ilharco, S. Y. Gadre, R. Roelofs, R. Gontijo-Lopes, A. S. Morcos, H. Namkoong, A. Farhadi, Y. Carmon, S. Kornblith *et al.*, “Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time,” in *International conference on machine learning*. PMLR, 2022, pp. 23 965–23 998.
- [24] Y. Sheng, L. Zheng, B. Yuan, Z. Li, M. Ryabinin, B. Chen, P. Liang, C. Ré, I. Stoica, and C. Zhang, “Flexgen: High-throughput generative inference of large language models with a single gpu,” in *International Conference on Machine Learning*. PMLR, 2023, pp. 31 094–31 116.
- [25] A. Q. Jiang, A. Sablayrolles, A. Mensch, C. Bamford, D. S. Chaplot, D. d. I. Casas, F. Bressand, G. Lengyel, G. Lample, L. Saulnier *et al.*, “Mistral 7b,” *arXiv preprint arXiv:2310.06825*, 2023.
- [26] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale *et al.*, “Llama 2: Open foundation and fine-tuned chat models,” *arXiv preprint arXiv:2307.09288*, 2023.
- [27] S. Zhang, S. Roller, N. Goyal, M. Artetxe, M. Chen, S. Chen, C. Dewan, M. Diab, X. Li, X. V. Lin *et al.*, “Opt: Open pre-trained transformer language models,” *arXiv preprint arXiv:2205.01068*, 2022.
- [28] T. Le Scao, A. Fan, C. Akiki, E. Pavlick, S. Ilić, D. Hesslow, R. Castagné, A. S. Luccioni, F. Yvon, M. Gallé *et al.*, “Bloom: A 176b-parameter open-access multilingual language model,” *arXiv preprint arXiv:2211.05100*, 2023.
- [29] Y. Intrator, M. Halfon, R. Goldenberg, R. Tsarfaty, M. Eyal, E. Rivlin, Y. Matias, and N. Aizenberg, “Breaking the language barrier: Can direct inference outperform pre-translation in multilingual llm applications?” *arXiv preprint arXiv:2403.04792*, 2024.
- [30] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever *et al.*, “Language models are unsupervised multitask learners,” *OpenAI blog*, vol. 1, no. 8, p. 9, 2019.
- [31] R. H. Suraperwata and S. Suyanto, “Language modeling for journalistic robot based on generative pretrained transformer 2,” in *2020 8th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 2020, pp. 1–6.
- [32] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray *et al.*, “Training language models to follow instructions with human feedback,” *Advances in neural information processing systems*, vol. 35, pp. 27 730–27 744, 2022.
- [33] H. Li, F. Koto, M. Wu, A. F. Aji, and T. Baldwin, “Bactrian-x: A multilingual replicable instruction-following model with low-rank adaptation,” *arXiv preprint arXiv:2305.15011*, 2023.

- [34] Y. Labrak, A. Bazoge, E. Morin, P.-A. Gourraud, M. Rouvier, and R. Dufour, "Biomistral: A collection of open-source pretrained large language models for medical domains," *arXiv preprint arXiv:2402.10373*, 2024.
- [35] S. Kim, J. Suk, S. Longpre, B. Y. Lin, J. Shin, S. Welleck, G. Neubig, M. Lee, K. Lee, and M. Seo, "Prometheus 2: An open source language model specialized in evaluating other language models," *arXiv preprint arXiv:2405.01535*, 2024.
- [36] B. Mahesh, "Machine learning algorithms-a review," *International Journal of Science and Research (IJSR)*. [Internet], vol. 9, no. 1, pp. 381–386, 2020.
- [37] M. A. Nielsen, *Neural networks and deep learning*. San Francisco, CA, USA: Determination press, 2015.
- [38] B. Rokh, A. Azarpeyvand, and A. Khanteymoori, "A comprehensive survey on model quantization for deep neural networks in image classification," *ACM Trans. Intell. Syst. Technol.*, vol. 14, no. 6, nov 2023. [Online]. Available: <https://doi.org/10.1145/3623402>
- [39] I. Hubara, Y. Nahshan, Y. Hanani, R. Banner, and D. Soudry, "Accurate post training quantization with small calibration sets," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 4466–4475. [Online]. Available: <https://proceedings.mlr.press/v139/hubara21a.html>
- [40] E. Frantar, S. Ashkboos, T. Hoeffler, and D. Alistarh, "Gptq: Accurate post-training quantization for generative pre-trained transformers," 2022.
- [41] K. Chowdhary and K. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence*, pp. 603–649, 2020.
- [42] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: State of the art, current trends and challenges," *Multimedia tools and applications*, vol. 82, no. 3, pp. 3713–3744, 2023.
- [43] M. Zhou, N. Duan, S. Liu, and H.-Y. Shum, "Progress in neural nlp: modeling, learning, and reasoning," *Engineering*, vol. 6, no. 3, pp. 275–290, 2020.
- [44] M.-P. Hosseini, S. Lu, K. Kamaraj, A. Slowikowski, and H. C. Venkatesh, "Deep learning architectures," *Deep learning: concepts and architectures*, pp. 1–24, 2020.
- [45] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [46] R. Xiong, Y. Yang, D. He, K. Zheng, S. Zheng, C. Xing, H. Zhang, Y. Lan, L. Wang, and T. Liu, "On layer normalization in the transformer architecture," *International Conference on Machine Learning*, pp. 10 524–10 533, 2020.
- [47] N. Shazeer, "Glu variants improve transformer," *arXiv preprint arXiv:2002.05202*, 2020.

- [48] A. Fan, M. Lewis, and Y. Dauphin, “Hierarchical neural story generation,” in *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, I. Gurevych and Y. Miyao, Eds. Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 889–898. [Online]. Available: <https://aclanthology.org/P18-1082>
- [49] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, “Generalizing from a few examples: A survey on few-shot learning,” *ACM Comput. Surv.*, vol. 53, no. 3, jun 2020. [Online]. Available: <https://doi.org/10.1145/3386252>
- [50] W. Li, Y. Peng, M. Zhang, L. Ding, H. Hu, and L. Shen, “Deep model fusion: A survey,” *arXiv preprint arXiv:2309.15698*, 2023.
- [51] A. Miaschi, D. Brunato, F. Dell’Orletta, and G. Venturi, “What makes my model perplexed? a linguistic investigation on neural language models perplexity,” in *Proceedings of Deep Learning Inside Out (DeeLIO): The 2nd Workshop on Knowledge Extraction and Integration for Deep Learning Architectures*, 2021, pp. 40–47.
- [52] H. Li and W. Lu, “Mixed cross entropy loss for neural machine translation,” in *International Conference on Machine Learning*. PMLR, 2021, pp. 6425–6436.
- [53] J. Robinson, C. Rytting, and D. Wingate, “Leveraging large language models for multiple choice question answering,” *ArXiv*, vol. abs/2210.12353, 2022. [Online]. Available: <https://api.semanticscholar.org/CorpusID:253098700>
- [54] D. Lane, D. Scott, M. Hebl, R. Guerra, D. Osherson, and H. Zimmer, *Introduction to statistics*. Citeseer, 2003.
- [55] K. Cobbe, V. Kosaraju, M. Bavarian, M. Chen, H. Jun, L. Kaiser, M. Plappert, J. Tworek, J. Hilton, R. Nakano *et al.*, “Training verifiers to solve math word problems,” *arXiv preprint arXiv:2110.14168*, 2021.
- [56] S. Kudugunta, I. Caswell, B. Zhang, X. Garcia, D. Xin, A. Kusupati, R. Stella, A. Bapna, and O. Firat, “Madlad-400: A multilingual and document-level large audited dataset,” *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [57] R. Mahendra, A. F. Aji, S. Louvan, F. Rahman, and C. Vania, “Indonli: A natural language inference dataset for indonesian,” in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2021. [Online]. Available: <http://dx.doi.org/10.18653/v1/2021.emnlp-main.821>
- [58] G. I. Winata, A. F. Aji, S. Cahyawijaya, R. Mahendra, F. Koto, A. Romadhony, K. Kurniawan, D. Moeljadi, R. E. Prasajo, P. Fung *et al.*, “Nusax: Multilingual parallel sentiment dataset for 10 indonesian local languages,” *arXiv preprint arXiv:2205.15960*, 2022.
- [59] B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar *et al.*, “Indonlu: Benchmark and

resources for evaluating indonesian natural language understanding,” *arXiv preprint arXiv:2009.05387*, 2020.

- [60] D. Kim, C. Park, S. Kim, W. Lee, W. Song, Y. Kim, H. Kim, Y. Kim, H. Lee, J. Kim *et al.*, “Solar 10.7 b: Scaling large language models with simple yet effective depth up-scaling,” *arXiv preprint arXiv:2312.15166*, 2023.
- [61] X.-P. Nguyen, W. Zhang, X. Li, M. Aljunied, Q. Tan, L. Cheng, G. Chen, Y. Deng, S. Yang, C. Liu *et al.*, “Seallms—large language models for southeast asia,” *arXiv preprint arXiv:2312.00738*, 2023.
- [62] R. Group, A. Levine, C. Huang, C. Wang, E. Batista, E. Szymanska, H. Ding, H. W. Chou, J.-F. Pessiot, J. Effendi *et al.*, “Rakutenai-7b: Extending large language models for japanese,” *arXiv preprint arXiv:2403.15484*, 2024.
- [63] R. Taori, I. Gulrajani, T. Zhang, Y. Dubois, X. Li, C. Guestrin, P. Liang, and T. B. Hashimoto, “Stanford alpaca: An instruction-following llama model,” 2023.
- [64] S. Toshniwal, I. Moshkov, S. Narenthiran, D. Gitman, F. Jia, and I. Gitman, “Openmathinstruct-1: A 1.8 million math instruction tuning dataset,” *arXiv preprint arXiv:2402.10176*, 2024.
- [65] G. Wang, S. Cheng, X. Zhan, X. Li, S. Song, and Y. Liu, “Openchat: Advancing open-source language models with mixed-quality data,” *arXiv preprint arXiv:2309.11235*, 2023.
- [66] C. Goddard, S. Siriwardhana, M. Ehghaghi, L. Meyers, V. Karpukhin, B. Benedict, M. McQuade, and J. Solawetz, “Arcee’s mergekit: A toolkit for merging large language models,” *arXiv preprint arXiv:2403.13257*, 2024.
- [67] C. Wang, M. Li, and A. J. Smola, “Language models with transformers,” *arXiv preprint arXiv:1904.09408*, 2019.
- [68] B. Juarto *et al.*, “Indonesian news classification using indobert,” *International Journal of Intelligent Systems and Applications in Engineering*, vol. 11, no. 2, pp. 454–460, 2023.
- [69] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi, “The curious case of neural text degeneration,” *arXiv preprint arXiv:1904.09751*, 2019.
- [70] S. Welleck, I. Kulikov, S. Roller, E. Dinan, K. Cho, and J. Weston, “Neural text generation with unlikelihood training,” *arXiv preprint arXiv:1908.04319*, 2019.