

DAFTAR PUSTAKA

- [1] B. Clavié, A. Ciceu, F. Naylor, G. Soulié, and T. Brightwell, “Large language models in the workplace: A case study on prompt engineering for job type classification,” 2023.
- [2] J. Zhang, T. Bui, S. Yoon, X. Chen, Z. Liu, C. Xia, Q. H. Tran, W. Chang, and P. Yu, “Few-shot intent detection via contrastive pre-training and fine-tuning,” 2021. [Online]. Available: <https://arxiv.org/abs/2109.06349>
- [3] B. Mahesh, “Machine learning algorithms - a review,” *International Journal of Science and Research (IJSR)*, vol. 9, no. 1, p. 381386, Jan. 2020. [Online]. Available: <http://dx.doi.org/10.21275/ART20203995>
- [4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [5] J. Ainslie, J. Lee-Thorp, M. de Jong, Y. Zemlyanskiy, F. Lebrón, and S. Sanghai, “Gqa: Training generalized multi-query transformer models from multi-head checkpoints,” 2023. [Online]. Available: <https://arxiv.org/abs/2305.13245>
- [6] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen, “Lora: Low-rank adaptation of large language models,” 2021. [Online]. Available: <https://arxiv.org/abs/2106.09685>
- [7] T. Dettmers, A. Pagnoni, A. Holtzman, and L. Zettlemoyer, “Qlora: Efficient fine-tuning of quantized llms,” 2023.
- [8] M. Castelli, L. Manzoni, and A. Popovi, “An artificial intelligence system to predict quality of service in banking organizations,” *Computational Intelligence and Neuroscience*, vol. 2016, 2016.
- [9] V. Rajnak and T. Puschmann, “The impact of blockchain on business models in banking,” *Information Systems and e-Business Management*, pp. 1–53, 2020.
- [10] P. Parne, “Cloud computing strategy and impact in banking/financial services,” *Computer Science and Information Technology Trends*, 2021.
- [11] A. Abubakar and R. Tasmin, “The impact of information and communication technology on banks performance and customer service delivery in the banking industry,” *The International Journal of Latest Trends in Finance and Economic Sciences*, vol. 2, 2012.
- [12] S. J. Ho and S. Mallick, “The impact of information technology on the banking industry,” *Journal of the Operational Research Society*, vol. 61, pp. 211–221, 2010.
- [13] S. Mehroliya, S. Alagarsamy, V. Moorthy, and J. S., “Will users continue using banking chatbots? the moderating role of perceived risk,” *FIIIB Business Review*, 2023.
- [14] D. Biswas, “Privacy preserving chatbot conversations,” in *2020 IEEE Third International Conference on Artificial Intelligence and Knowledge Engineering (AIKE)*, 2020, pp. 179–182.

- [15] A. N. E Mogaji, J Balakrishnan, “Emerging-market consumers’ interactions with banking chatbots,” *Telematics and Informatics*, 2021. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0736585321001507>
- [16] J. Cahn, “Chatbot: Architecture, design, & development,” *University of Pennsylvania School of Engineering and Applied Science*, 2017. [Online]. Available: <https://www.academia.edu/download/57035006/CHATBOTthesisfinal.pdf>
- [17] F. Johannsen, S. Leist, D. Konadl, and M. Basche, “Comparison of commercial chatbot solutions for supporting customer interaction,” 2018.
- [18] A. Elcholiqi, “Chatbot berbahasa indonesia menggunakan natural language processing untuk menyediakan informasi produk perbankan,” Ph.D. dissertation, Universitas Gadjah Mada, 2018.
- [19] W. Maeng and J. Lee, “Designing a chatbot for survivors of sexual violence: Exploratory study for hybrid approach combining rule-based chatbot and ml-based chatbot,” in *Proceedings of the Asian CHI Symposium 2021*, ser. Asian CHI ’21. New York, NY, USA: Association for Computing Machinery, 2021, p. 160166. [Online]. Available: <https://doi.org/10.1145/3429360.3468203>
- [20] N. Bhartiya, N. Jangid, S. Jannu, P. Shukla, and R. Chapaneri, “Artificial neural network based university chatbot system,” in *2019 IEEE Bombay Section Signature Conference (IBSSC)*, 2019, pp. 1–6.
- [21] 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.
- [22] K. Kopyl, “Generative AI in Finance: Real-World Examples of LLMs in Banking — tovie.ai,” <https://tovie.ai/blog/generative-ai-in-finance-real-world-examples-of-llms-in-banking-in-2023>, 2023, [Accessed 16-06-2024].
- [23] D. R. So, W. Ma’nke, H. Liu, Z. Dai, N. M. Shazeer, and Q. V. Le, “Primer: Searching for efficient transformers for language modeling,” *ArXiv*, vol. abs/2109.08668, 2021.
- [24] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Kuttler, M. Lewis, W. tau Yih, T. Rocktäschel, S. Riedel, and D. Kiela, “Retrieval-augmented generation for knowledge-intensive nlp tasks,” *ArXiv*, vol. abs/2005.11401, 2020.
- [25] N. Ding, Y. Qin, G. Yang, F. Wei, Z. Yang, Y. Su, S. Hu, Y. Chen, C.-M. Chan, W. Chen, J. Yi, W. Zhao, X. Wang, Z. Liu, H.-T. Zheng, J. Chen, Y. Liu, J. Tang, J. Li, and M. Sun, “Parameter-efficient fine-tuning of large-scale pre-trained language models,” *Nat. Mach. Intell.*, vol. 5, no. 3, pp. 220–235, Mar. 2023.
- [26] T. Dettmers, M. Lewis, S. Shleifer, and L. Zettlemoyer, “8-bit optimizers via block-wise quantization,” 2022. [Online]. Available: <https://arxiv.org/abs/2110.02861>

- [27] C. Costello, R. Lin, V. Mruthyunjaya, B. Bolla, and C. Jankowski, “Multi-layer ensembling techniques for multilingual intent classification,” 2018. [Online]. Available: <https://arxiv.org/abs/1806.07914>
- [28] I. Sheikh, I. Illina, D. Fohr, and G. Linares, “Learning Word Importance with the Neural Bag-of-Words Model,” in *ACL, Representation Learning for NLP (Repl4NLP) workshop*, ser. Proceedings of ACL 2016, Berlin, Germany, Aug. 2016. [Online]. Available: <https://hal.science/hal-01331720>
- [29] H. Purohit, G. Dong, V. Shalin, K. Thirunarayan, and A. Sheth, “Intent classification of short-text on social media,” in *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, 2015, pp. 222–228.
- [30] J. Schuurmans and F. Frasincar, “Intent classification for dialogue utterances,” *IEEE Intelligent Systems*, vol. 35, no. 1, pp. 82–88, 2019.
- [31] Y. Shi, K. Yao, L. Tian, and D. Jiang, “Deep LSTM based feature mapping for query classification,” in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, K. Knight, A. Nenkova, and O. Rambow, Eds. San Diego, California: Association for Computational Linguistics, Jun. 2016, pp. 1501–1511. [Online]. Available: <https://aclanthology.org/N16-1176>
- [32] Q. Chen, Z. Zhuo, and W. Wang, “Bert for joint intent classification and slot filling,” 2019. [Online]. Available: <https://arxiv.org/abs/1902.10909>
- [33] D. Braun, A. Hernandez Mendez, F. Matthes, and M. Langen, “Evaluating natural language understanding services for conversational question answering systems,” in *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, K. Jokinen, M. Stede, D. DeVault, and A. Louis, Eds. Saarbrücken, Germany: Association for Computational Linguistics, Aug. 2017, pp. 174–185. [Online]. Available: <https://aclanthology.org/W17-5522>
- [34] S. Larson, A. Mahendran, J. J. Peper, C. Clarke, A. Lee, P. Hill, J. K. Kummerfeld, K. Leach, M. A. Laurenzano, L. Tang, and J. Mars, “An evaluation dataset for intent classification and out-of-scope prediction,” 2019. [Online]. Available: <https://arxiv.org/abs/1909.02027>
- [35] J. Howard and S. Ruder, “Universal language model fine-tuning for text classification,” 2018.
- [36] P. He, J. Gao, and W. Chen, “Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing,” 2023.
- [37] I. Casanueva, T. Temčinas, D. Gerz, M. Henderson, and I. Vulić, “Efficient intent detection with dual sentence encoders,” in *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*, T.-H. Wen, A. Celikyilmaz, Z. Yu, A. Papangelis, M. Eric, A. Kumar, I. Casanueva, and R. Shah, Eds. Online: Association for Computational Linguistics, Jul. 2020, pp. 38–45. [Online]. Available: <https://aclanthology.org/2020.nlp4convai-1.5>

- [38] X. Liu, A. Eshghi, P. Swietojanski, and V. Rieser, “Benchmarking natural language understanding services for building conversational agents,” 2019. [Online]. Available: <https://arxiv.org/abs/1903.05566>
- [39] J. Zhang, K. Hashimoto, W. Liu, C.-S. Wu, Y. Wan, P. Yu, R. Socher, and C. Xiong, “Discriminative nearest neighbor few-shot intent detection by transferring natural language inference,” 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. 5064–5082. [Online]. Available: <https://aclanthology.org/2020.emnlp-main.411>
- [40] S. Mehri, M. Eric, and D. Hakkani-Tur, “Dialogue: A natural language understanding benchmark for task-oriented dialogue,” 2020. [Online]. Available: <https://arxiv.org/abs/2009.13570>
- [41] S. Mehri and M. Eric, “Example-driven intent prediction with observers,” *arXiv preprint arXiv:2010.08684*, 2020.
- [42] S. Kumar, S. Deep, and P. Kalra, “Enhancing customer service in banking with ai: Intent classification using distilbert,” 2024.
- [43] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter,” 2020. [Online]. Available: <https://arxiv.org/abs/1910.01108>
- [44] L. Loukas, I. Stogiannidis, O. Diamantopoulos, P. Malakasiotis, and S. Vassos, “Making llms worth every penny: Resource-limited text classification in banking,” in *4th ACM International Conference on AI in Finance*, ser. ICAIF 23. ACM, Nov. 2023. [Online]. Available: <http://dx.doi.org/10.1145/3604237.3626891>
- [45] K. Chowdhary and K. Chowdhary, “Natural language processing,” *Fundamentals of artificial intelligence*, pp. 603–649, 2020.
- [46] 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.
- [47] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz *et al.*, “Transformers: State-of-the-art natural language processing,” in *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, 2020, pp. 38–45.
- [48] V. N. Gudivada, D. L. Rao, and A. R. Gudivada, *Information Retrieval: Concepts, Models, and Systems*. Elsevier, 2018, p. 331401. [Online]. Available: <http://dx.doi.org/10.1016/bs.host.2018.07.009>
- [49] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, D. Bikel, L. Blecher, C. C. Ferrer, M. Chen, G. Cucurull, D. Esiobu, J. Fernandes, J. Fu, W. Fu, B. Fuller, C. Gao, V. Goswami, N. Goyal, A. Hartshorn, S. Hosseini, R. Hou, H. Inan, M. Kardas, V. Kerkez, M. Khabsa, I. Kloumann, A. Korenev, P. S. Koura, M.-A. Lachaux, T. Lavril, J. Lee, D. Liskovich, Y. Lu, Y. Mao, X. Martinet, T. Mihaylov,

- P. Mishra, I. Molybog, Y. Nie, A. Poulton, J. Reizenstein, R. Rungta, K. Saladi, A. Schelten, R. Silva, E. M. Smith, R. Subramanian, X. E. Tan, B. Tang, R. Taylor, A. Williams, J. X. Kuan, P. Xu, Z. Yan, I. Zarov, Y. Zhang, A. Fan, M. Kambadur, S. Narang, A. Rodriguez, R. Stojnic, S. Edunov, and T. Scialom, "Llama 2: Open foundation and fine-tuned chat models," 2023. [Online]. Available: <https://arxiv.org/abs/2307.09288>
- [50] B. Rokh, A. Azarpeyvand, and A. Khanteymoori, "A comprehensive survey on model quantization for deep neural networks in image classification," *ACM Transactions on Intelligent Systems and Technology*, vol. 14, no. 6, p. 150, Nov. 2023. [Online]. Available: <http://dx.doi.org/10.1145/3623402>
- [51] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei, "Language models are few-shot learners," 2020. [Online]. Available: <https://arxiv.org/abs/2005.14165>
- [52] B. Lester, R. Al-Rfou, and N. Constant, "The power of scale for parameter-efficient prompt tuning," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 3045–3059. [Online]. Available: <https://aclanthology.org/2021.emnlp-main.243>
- [53] N. Houlsby, A. Giurgiu, S. Jastrzebski, B. Morrone, Q. De Laroussilhe, A. Gesmundo, M. Attariyan, and S. Gelly, "Parameter-efficient transfer learning for NLP," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 2790–2799. [Online]. Available: <https://proceedings.mlr.press/v97/houlsby19a.html>
- [54] X. L. Li and P. Liang, "Prefix-tuning: Optimizing continuous prompts for generation," 2021. [Online]. Available: <https://arxiv.org/abs/2101.00190>
- [55] M. M. Noel, A. Banerjee, G. B. A. D, and V. Muthiah-Nakarajan, "Alternate loss functions for classification and robust regression can improve the accuracy of artificial neural networks," 2023. [Online]. Available: <https://arxiv.org/abs/2303.09935>
- [56] D. P. Amur, "QLoRA: Fine-Tuning Large Language Models (LLMs) — dilliprasad60," <https://medium.com/@dilliprasad60/qlora-explained-a-deep-dive-into-parametric-efficient-fine-tuning-in-large-language-models-llms-2023>, [Accessed 25-08-2024].
- [57] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019. [Online]. Available: <https://arxiv.org/abs/1810.04805>

- [58] A. Benayas, M. A. Sicilia, and M. Mora-Cantallops, “A comparative analysis of encoder only and decoder only models in intent classification and sentiment analysis: Navigating the trade-offs in model size and performance,” Jan. 2024. [Online]. Available: <http://dx.doi.org/10.21203/rs.3.rs-3865391/v1>
- [59] OpenAI, “Gpt-3.5 turbo,” <https://platform.openai.com/docs/models/gpt-3-5-turbo>, 2024, [Accessed 18-08-2024].
- [60] W.-L. Chiang, L. Zheng, Y. Sheng, A. N. Angelopoulos, T. Li, D. Li, H. Zhang, B. Zhu, M. Jordan, J. E. Gonzalez, and I. Stoica, “Chatbot arena: An open platform for evaluating llms by human preference,” 2024.
- [61] S. Cahyawijaya, H. Lovenia, F. Koto, R. A. Putri, E. Dave, J. Lee, N. Shadieq, W. Cenggoro, S. M. Akbar, M. I. Mahendra, D. A. Putri, B. Wilie, G. I. Winata, A. F. Aji, A. Purwarianti, and P. Fung, “Cendol: Open instruction-tuned generative large language models for indonesian languages,” 2024. [Online]. Available: <https://arxiv.org/abs/2404.06138>
- [62] A. Kabra, E. Liu, S. Khanuja, A. F. Aji, G. Winata, S. Cahyawijaya, A. Aremu, P. Ogayo, and G. Neubig, “Multi-lingual and multi-cultural figurative language understanding,” in *Findings of the Association for Computational Linguistics: ACL 2023*, A. Rogers, J. Boyd-Graber, and N. Okazaki, Eds. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 8269–8284. [Online]. Available: <https://aclanthology.org/2023.findings-acl.525>
- [63] X.-P. Nguyen, W. Zhang, X. Li, M. Aljunied, Z. Hu, C. Shen, Y. K. Chia, X. Li, J. Wang, Q. Tan, L. Cheng, G. Chen, Y. Deng, S. Yang, C. Liu, H. Zhang, and L. Bing, “Seallms – large language models for southeast asia,” 2023.
- [64] L. Owen, V. Tripathi, A. Kumar, and B. Ahmed, “Komodo: A linguistic expedition into indonesia’s regional languages,” 2024. [Online]. Available: <https://arxiv.org/abs/2403.09362>
- [65] M. Ichsan, “Merak-7b: The llm for bahasa indonesia,” *Hugging Face Repository*, 2023.
- [66] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi, “The curious case of neural text degeneration,” 2020. [Online]. Available: <https://arxiv.org/abs/1904.09751>