

DAFTAR PUSTAKA

- Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM*, 64(7), 58–65. <https://doi.org/10.1145/3448250>
- Bhat, S. R., Rudat, M., Spiekermann, J., & Flores-Herr, N. (2025). Rethinking Chunk Size For Long-Document Retrieval: A Multi-Dataset Analysis. *ArXiv Preprint ArXiv:2505.21700*. <https://doi.org/DOI:10.48550/arXiv.2505.21700>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems (NeurIPS 2020)*, 1877–1901. <https://doi.org/https://doi.org/10.48550/arXiv.2005.14165>
- Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T. S., & Li, Q. (2024). A Survey on RAG Meeting LLMs: Towards Retrieval-Augmented Large Language Models. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 6491–6501. <https://doi.org/10.1145/3637528.3671470>
- Hikmah, N., Ariyanti, D., & Pratama, F. A. (2022). Implementasi Chatbot Sebagai Virtual Assistant di Universitas Panca Marga Probolinggo menggunakan Metode TF-IDF. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 4(2), 133–148. <https://doi.org/10.35746/jtim.v4i2.225>
- Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition Third Edition draft*. <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>

- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). *Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks*. <http://arxiv.org/abs/2005.11401>
- Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025). *Large Language Models: A Survey*. <http://arxiv.org/abs/2402.06196>
- Muhajir.M.D.A, Prastiti.N, & Koeshardianto.M. (2025). Implementasi Chatbot Menggunakan Framework Langchain Berbasis Llm Gpt (Studi Kasus : Panduan Akademik Universitas Trunojoyo). In *Jurnal Mahasiswa Teknik Informatika* (Vol. 9, Issue 2). <https://ejournal.itn.ac.id/index.php/jati/article/view/13003>
- Murley, P., Ma, Z., Mason, J., Bailey, M., & Kharraz, A. (2021). Websocket adoption and the landscape of the real-time web. *The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021*, 1192–1203. <https://doi.org/10.1145/3442381.3450063>
- Nugraha, S., Irwan, M., & Nasution, P. (2024). Pengaruh Chatbot AI terhadap Beban Kerja Layanan Pelanggan Grab di Kota Medan. In *Jurnal Ekonomi Dan Bisnis Digital*(Vol.02). <https://jurnal.ittc.web.id/index.php/jebd/article/download/1929/1855/6042>
- Pressman, R. S. (2010). *Software Engineering: A Practitioner's Approach*. McGraw-Hill Higher Education. www.mhhe.com/pressman.
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *Association for Computational Linguistics*, 3982–3992. <https://doi.org/10.18653/v1/D19-1410>
- Samudra, G., Turmudi Zy, A., & Ermanto. (2025). Implementasi Retrieval Augmented Generation (RAG) Dalam. *JSAI: Journal Scientific and Applied Informatics*, 8(1). <https://doi.org/10.36085>

- Soygazi, F., & Oguz, D. (2023). An Analysis of Large Language Models and LangChain in Mathematics Education. *ACM International Conference Proceeding Series*, 92–97. <https://doi.org/10.1145/3633598.3633614>
- Tay, Y., Dehghani, M., Bahri, D., & Metzler, D. (2023). Efficient Transformers: A Survey. *ACM Computing Surveys*, 55(6). <https://doi.org/10.1145/3530811>
- Tunstall, L., Von Werra, L., & Wolf, T. (2022). *Natural Language Processing with Transformers Building Language Applications with Hugging Face* (Revised Edition). O'Reilly Media.
- Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*, 5998–6008. <https://doi.org/https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
- Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., ... Rush, A. M. (2020). Transformers: State-of-the-Art Natural Language Processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. <https://doi.org/10.18653/v1/2020.emnlp-demos.6>