

DAFTAR PUSTAKA

- Argene, M., Ravenscroft, C., & Kingswell, I. (2024). *Ransomware detection via cosine similarity-based machine learning on bytecode representations*.
- Asudani, D. S., Nagwani, N. K., & Singh, P. (2023). Impact of word embedding models on text analytics in deep learning environment: a review. *Artificial Intelligence Review*, 56(9). <https://doi.org/10.1007/s10462-023-10419-1>
- Auffarth, B. (2023). Generative AI with LangChain: Build large language model (LLM) apps with Python, ChatGPT, and other LLMs. In *Packt Publishing*.
- Azharyani, I., & Kusumo, D. S. (2019). Implementasi Semantic Search pada Open Library Menggunakan Metode Latent Semantic Analysis (Studi Kasus: Open Library Universitas Telkom). *Agustus*, 6(2).
- Huawei Technologies Co., Ltd. (2023). Artificial Intelligence Technology. In *Artificial Intelligence Technology*. <https://doi.org/10.1007/978-981-19-2879-6>
- Johnson, J., Douze, M., & Jegou, H. (2021). Billion-Scale Similarity Search with GPUs. *IEEE Transactions on Big Data*, 7(3). <https://doi.org/10.1109/TBDATA.2019.2921572>
- Karpukhin, V., Oğuz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. T. (2020). Dense passage retrieval for open-domain question answering. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
- Kuhail, M. A., Abu Shawar, B., & Hammad, R. (2023). Trends, applications, and challenges of chatbot technology. In *Trends, Applications, and Challenges of Chatbot Technology*. <https://doi.org/10.4018/978-1-6684-6234-8>
- Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Küttler, H., Lewis, M., Yih, W. T., Rocktäschel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 2020-December*.
- Norris, D., & Kalm, K. (2021). Chunking and data compression in verbal short-term memory. *Cognition*, 208. <https://doi.org/10.1016/j.cognition.2020.104534>
- Pichai, S., & Hassabis, D. (2023). Introducing Gemini: our largest and most capable AI model. *Google*. <https://blog.google/technology/ai/google-gemini-ai/>

- Raharjo, B. (2021). Penerapan Artificial Intelegent (AI) dalam Bisnis. *Penerbit Yayasan Prima Agus Teknik*, 1–233.
- Rick, M. (2023). *What Is Retrieval-Augmented Generation aka RAG*. NVIDIA.
- Sallam, M. (2023). ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. In *Healthcare (Switzerland)* (Vol. 11, Issue 6). <https://doi.org/10.3390/healthcare11060887>
- Shim, K., Lee, J., Chang, S., & Hwang, K. (2023). Knowledge Distillation from Non-streaming to Streaming ASR Encoder using Auxiliary Non-streaming Layer. *ArXiv Preprint ArXiv:2308.16415*.
- Shinta, A. (2022). *Apa Itu Javascript? Pengertian, Fungsi dan Contohnya*. Dewaweb.
- Tanwar, S. (2024). Machine Learning. In *Computational Science and Its Applications*. <https://doi.org/10.1201/9781003347484-2>
- Wahid, A. A. (2020). Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi. *Jurnal Ilmu-Ilmu Informatika Dan Manajemen STMIK*, November.
- Xiao, G., Tian, Y., Chen, B., Han, S., & Lewis, M. (2023). Efficient streaming language models with attention sinks. *ArXiv Preprint ArXiv:2309.17453*.