DAFTAR PUSTAKA

[1] K. Shuster, S. Poff, M. Chen, D. Kiela, and J. Weston, "Retrieval Augmentation Reduces Hallucination in Conversation," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, M.-F. Moens, X. Huang, L. Specia, and S. W. Yih, Eds., Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 3784–3803. doi: 10.18653/v1/2021.findings-emnlp.320.

[2] O. W. Purbo, *Internet - TCP/IP : Konsep \& Implementasi*. Penerbit Andi, 2018. [Online]. Available: https://books.google.co.id/books?id=KoHuzwEACAAJ

[3] M. Reid *et al.*, "Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context," *ArXiv*, vol. abs/2403.0, 2024, doi: https://doi.org/10.48550/arXiv.2403.05530.

[4] Y. Han, C. Liu, and P. Wang, "A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge," *ArXiv*, vol. abs/2310.1, 2023, [Online]. Available: https://api.semanticscholar.org/CorpusID:264289073

[5] Y. A. Malkov and D. A. Yashunin, "Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 824–836, 2020, doi: 10.1109/TPAMI.2018.2889473.

[6] C. V. S. Satyamurty, J. V. R. Murthy, and M. Raghava, "Metadata-Based Semantic Query in Relational Databases," in *Information Systems Design and Intelligent Applications*, V. Bhateja, B. Le Nguyen, N. G. Nguyen, S. C. Satapathy, and D.-N. Le, Eds., Singapore: Springer Singapore, 2018, pp. 183–188.

[7] C. Boettiger, "An introduction to Docker for reproducible research," *ACM SIGOPS Oper. Syst. Rev.*, vol. 49, pp. 71–79, 2014, [Online]. Available: https://api.semanticscholar.org/CorpusID:255583

[8] S. Li, L. Stenzel, C. Eickhoff, and S. A. Bahrainian, "Enhancing Retrieval-Augmented Generation: A Study of Best Practices," in *International Conference on Computational Linguistics*, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:275471724

[9] H. Yu, A. Gan, K. Zhang, S. Tong, Q. Liu, and Z. Liu, "Evaluation of Retrieval-Augmented Generation: A Survey," *ArXiv*, vol. abs/2405.0, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:269758033

[10] A. Ziletti and L. D'Ambrosi, "Retrieval augmented text-to-SQL generation for epidemiological question answering using electronic health records," in *Clinical Natural Language Processing Workshop*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:268385157

[11] P. Lewis *et al.*, "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," *arXiv e-prints*, p. arXiv:2005.11401, May 2020, doi: 10.48550/arXiv.2005.11401.

[12] S. Gupta, R. Ranjan, and S. N. Singh, "A Comprehensive Survey of Retrieval-Augmented Generation (RAG): Evolution, Current Landscape and Future Directions," *ArXiv*, vol. abs/2410.1, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:273403982

[13] A. A. Khan, M. T. Hasan, K.-K. Kemell, J. Rasku, and P. Abrahamsson, "Developing Retrieval Augmented Generation (RAG) based LLM Systems from PDFs: An Experience Report," *ArXiv*, vol. abs/2410.1, 2024, [Online]. Available: https://api.semanticscholar.org/CorpusID:273502291

[14] J. Gómez, V. H. Riaño, and G. Ramirez-Gonzalez, "Traffic Classification in IP Networks Through Machine Learning Techniques in Final Systems," *IEEE Access*, vol. 11, pp. 44932–44940, 2023, doi: 10.1109/ACCESS.2023.3272894.

[15] S. Srivastava, H. Srivastava, A. Jaymani, and P. Singh, "The Future of AI in Production : Leveraging Kubernetes for Large Language Model Deployment," *Int. J. Multidiscip. Res.*, vol. 7, no. 1, pp. 1–18, 2025, doi: https://doi.org/10.36948/ijfmr.2025.v07i01.36056.

[16] M. Openja, F. Majidi, F. Khomh, B. Chembakottu, and H. Li, "Studying the Practices of Deploying Machine Learning Projects on Docker," *Proc. 26th Int. Conf. Eval. Assess. Softw. Eng.*, 2022, doi: https://doi.org/10.1145/3530019.3530039.

[17] R. Hu, C. Peng, X. Wang, and C. Gao, "An LLM-based Agent for Reliable Docker Environment Configuration," *ArXiv*, vol. abs/2502.1, 2025, doi: https://doi.org/10.48550/arXiv.2502.13681.

[18] B. Han, T. Susnjak, and A. Mathrani, "Automating Systematic Literature Reviews with Retrieval-Augmented Generation: A Comprehensive Overview," *Appl. Sci.*, vol. 14, no. 19, 2024, doi: 10.3390/app14199103.

[19] A. Kamath, R. Jia, and P. Liang, "Selective Question Answering under Domain Shift," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, D. Jurafsky, J. Chai, N. Schluter, and J. Tetreault, Eds., Online: Association for Computational Linguistics, Jul. 2020, pp. 5684–5696. doi: 10.18653/v1/2020.acl-main.503.

[20] A. S. Tanenbaum and D. Wetherall, *Computer Networks*, 5th ed. Pearson Prentice Hall, 2011. [Online]. Available: https://books.google.co.id/books?id=I764bwAACAAJ

[21] B. A. Forouzan and S. C. Fegan, *Data Communications and Networking*, 4th ed. in Data Communications and Networking. McGraw-Hill Higher Education, 2007. [Online]. Available:

https://books.google.co.id/books?id=bwUNZvJbEeQC

[22] "Internet Protocol," Sep. 1981, *RFC Editor*. doi: 10.17487/RFC0791.

[23] R. Moskowitz, D. Karrenberg, Y. Rekhter, E. Lear, and G. J. de Groot, "Address Allocation for Private Internets," Feb. 1996, *RFC Editor*. doi: 10.17487/RFC1918.

[24] L. L. Peterson and B. S. Davie, *Computer Networks: A Systems Approach*. in The Morgan Kaufmann Series in Networking. Morgan Kaufmann, 2011. [Online]. Available: https://books.google.co.id/books?id=BvaFreun1W8C

[25] V. Fuller, T. Li, K. Varadhan, and J. Yu, "Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy," Sep. 1993, *RFC Editor*. doi: 10.17487/RFC1519.

[26] B. Manning and T. Pummill, "Variable Length Subnet Table For IPv4," Dec. 1995, *RFC Editor*. doi: 10.17487/RFC1878.

[27] A. Vaswani *et al.*, "Attention Is All You Need," *arXiv e-prints*, p. arXiv:1706.03762, Jun. 2017, doi: 10.48550/arXiv.1706.03762.

[28] OpenAI, "What is the context window of the new GPT 3.5 Turbo model (gpt-3.5-turbo-0125)?" Accessed: May 25, 2025. [Online]. Available: https://community.openai.com/t/what-is-the-context-window-of-the-the-new-gpt-3-5-turbo-model-gpt-3-5-turbo-0125/609532

[29] Anthropic, "Models overview." Accessed: May 25, 2025. [Online]. Available: https://docs.anthropic.com/en/docs/about-claude/models/overview

[30] Mistral AI team, "Mistral 7B." Accessed: May 25, 2025. [Online]. Available: https://mistral.ai/news/announcing-mistral-7b

[31] C. Coronel and S. Morris, *Database Systems: Design, Implementation, & Management*, 12th ed. Cengage Learning, 2016. [Online]. Available: https://books.google.co.id/books?id=elVZCwAAQBAJ

[32] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," *arXiv Prepr. arXiv1908.10084*, 2019.

[33] T. Taipalus, "Vector database management systems: Fundamental concepts, use-cases, and current challenges," *Cogn. Syst. Res.*, vol. 85, p. 101216, 2024, doi: https://doi.org/10.1016/j.cogsys.2024.101216.

[34] M. Douze *et al.*, "The faiss library," *arXiv Prepr. arXiv2401.08281*, 2024.

[35] D. Merkel, "Docker: lightweight Linux containers for consistent development and deployment," *Linux J.*, vol. 2014, p. 2, 2014, [Online]. Available: https://api.semanticscholar.org/CorpusID:62479797

[36] C. Boettiger, "An introduction to Docker for reproducible research,"

SIGOPS Oper. Syst. Rev., vol. 49, no. 1, pp. 71–79, Jan. 2015, doi: 10.1145/2723872.2723882.

[37] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability \& Statistics for Engineers \& Scientists*. Pearson, 2016. [Online]. Available: https://books.google.co.id/books?id=pP6SjwEACAAJ

[38] L. J. Cronbach, "Coefficient Alpha and the Internal Structure of Tests," *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951, doi: 10.1007/BF02310555.

[39] C. D. Manning, P. Raghavan, and H. Schütze, "Evaluation in information retrieval," in *Introduction to Information Retrieval*, Cambridge University Press, 2008, pp. 139–161.

[40] M. S. B. Yusoff, "ABC of Content Validation and Content Validity Index Calculation," *Educ. Med. J.*, vol. 11, pp. 49–54, 2019, doi: 10.21315/eimj2019.11.2.6.