

DAFTAR PUSTAKA

- [1] N. W. Anisa Rahmawati, Syabina Najla Amirah, “Integrasi Kecerdasan Buatan dalam Pendidikan Tinggi Indonesia: Peluang, Tantangan, dan Kerangka Implementasi,” vol. 6, no. April, pp. 114–126, 2025, doi: 10.35957/jtsi.v6i1.11329.
- [2] D. Damayanti and A. K. Nuzuli, “Studi Kasus Implementasi Teknologi Chatbot sebagai Asisten Virtual dalam Menjawab Pertanyaan Mahasiswa di Lingkungan Kampus,” *Reslaj Relig. Educ. Soc. Laa Roiba J.*, vol. 6, no. 3, pp. 1178–1192, 2024, doi: 10.47476/reslaj.v6i3.4858.
- [3] A. Amin, T. Bustomi, and K. Harianto, “Implementasi Chatbot Sebagai Asisten Virtual Untuk Membantu Mahasiswa Dalam Mendapatkan Informasi Akademik,” vol. 28, no. 2, 2024, doi: 10.46984/sebatik.v28i2.0000.
- [4] C. Sulastri, E. Hulasoh, and Widayat, “Peranan Etika Komunikasi di Grup Whatsapp Mahasiswa Ilmu Komunikasi Universitas Pamulang,” *PROMEDIA (Public Relat. Dan Media Komunikasi)*, vol. 10, no. 1, pp. 149–166, 2024.
- [5] UPN Veteran Jawa Timur, “Pejabat Pengelola Informasi dan Dokumentasi (PPID) UPN Veteran Jawa Timur.” [Online]. Available: <https://ppid.upnjatim.ac.id/>
- [6] UPN Veteran Jawa Timur, “bot telegram - UPN Veteran JATIM.” [Online]. Available: <https://t.me/upnjatimbot>
- [7] G. Nariratri, E. Rizal, and R. K. Anwar, “Bacaan digital: pengaruh aspek keunggulannya terhadap meningkatnya minat baca mahasiswa,” *J. Ilm. Multidisiplin Indones.*, vol. 2, no. 12, 2024, doi: 10.55904/nautical.v2i12.643.
- [8] UPT-TIK UPN Veteran Jawa Timur, “Sistem Informasi Aduan (SIADU) UPNVJT.” [Online]. Available: <https://siadu.upnjatim.ac.id/>
- [9] W. Assyahri and M. Mardaus, “Pengaruh Kualitas Pelayanan Akademik terhadap Kepuasan Mahasiswa Universitas Negeri Padang,” *J. Manaj. dan Ilmu Adm. Publik*, vol. 5, no. 3, pp. 239–247, 2023, doi: 10.24036/jmiap.v5i3.638.

- [10] Pangkalan Data Pendidikan Tinggi, “Jumlah Mahasiswa Aktif Universitas Pembangunan Nasional Veteran Jawa Timur.” Accessed: Nov. 29, 2025. [Online]. Available: <https://pddikti.kemdiktisaintek.go.id/detail-pt/LWDHVWsd6kLc2CGNHnEUp96puo7ZAcqcke8MpdiUfxZ9fujLi-VVripfs-2n1kI-pEaHgg==>
- [11] N. I. Cikita Berlian Hakima, Jeki Pornomo, “Evaluasi Beban Kerja Staf Administrasi Dalam Mendukung Perbaikan Kualitas Pelayanan Akademik,” *J. Teknol. dan Manaj. Ind.*, vol. 5, no. 2, p. 4351, 2024.
- [12] C. D. Manning and H. Schütze, *Foundations of Statistics Natural Language Processing*. 1999. doi: 10.4324/9781410600004-24.
- [13] W. X. Zhao *et al.*, “A Survey of Large Language Models,” pp. 1–144, 2025, doi: 10.1007/s11704-024-40663-9.
- [14] B. Min *et al.*, “Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey,” *ACM Comput. Surv.*, vol. 56, no. 2, 2023, doi: 10.1145/3605943.
- [15] Z. Ji *et al.*, “Survey of Hallucination in Natural Language Generation,” *ACM Comput. Surv.*, vol. 55, no. 12, pp. 1–38, 2023, doi: 10.1145/3571730.
- [16] G. Izacard *et al.*, “Atlas: Few-shot Learning with Retrieval Augmented Language Models,” *J. Mach. Learn. Res.*, vol. 24, no. 1, pp. 11912–11954, 2023.
- [17] I. Pujiono, I. M. Agtyaputra, and Y. Ruldeviyani, “Implementing Retrieval-Augmented Generation and Vector Databases for Chatbots in Public Services Agencies Context,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 1, pp. 216–223, 2024, doi: 10.33480/jitk.v10i1.5572.
- [18] H. W. Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Meng Wang, “Retrieval-Augmented Generation for Large Language Models: A Survey,” *Proc. - 2024 Conf. AI, Sci. Eng. Technol. AIXSET 2024*, 2024, doi: 10.48550/arXiv.2312.10997.
- [19] F. Zhang and Z. Yu, “Mind the Gap: Linguistic Divergence and Adaptation Strategies in Human-LLM Assistant vs. Human-Human Interactions,” 2025, doi: 10.48550/arXiv.2510.02645.
- [20] M. Saman, G. Ahmad, F. Alfarisy, R. Amelia, and N. R. Fadhliana, “A RAG-

- Based Academic Information Chatbot Using Lightweight LLaMA and Indocence-BERT,” *Indones. J. Artif. Intell. Data Min.*, vol. 8, no. 3, pp. 687–696, 2025, doi: 10.24014/ijaidm.v8i3.38150.
- [21] H. Tohir, N. Merlina, and M. Haris, “Utilizing Retrieval-Augmented Generation in Large Language Models to Enhance Indonesian Language NLP,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 10, no. 2, pp. 352–360, 2024, doi: 10.33480/jitk.v10i2.5916.
- [22] G. D. Albert and A. Voutama, “Pengembangan Chatbot Berbasis PDF Menggunakan Local Retrieval-Augmented Generation (RAG) dan Ollama,” *J. Inform. dan Tek. ElektroTerapan*, vol. 13, no. 2, 2025, doi: 10.23960/jitet.v13i2.6361.
- [23] K. Polin, T. Yigitcanlar, M. Limb, and T. Washington, “The Making of Smart Campus: A Review and Conceptual Framework,” *Buildings*, 2023, doi: 10.3390/buildings13040891.
- [24] S. D. Atagong, H. Tonnang, K. Senagi, M. Wamalwa, K. M. Agboka, and J. Odindi, “A review on knowledge and information extraction from PDF documents and storage approaches,” *Front. Artif. Intell.*, 2025, doi: 10.3389/frai.2025.1466092.
- [25] G. M. Csányi *et al.*, “From Fact Drafts to Operational Systems: Semantic Search in Legal Decisions Using Fact Drafts,” *Big Data Cogn. Comput.*, vol. 8, no. 12, pp. 1–22, 2024, doi: 10.3390/bdcc8120185.
- [26] N. Livathinos *et al.*, “Docling : An Efficient Open-Source Toolkit for AI-driven Document Conversion,” 2025, doi: 10.48550/arXiv.2501.17887.
- [27] Q. Zhang *et al.*, “Document Parsing Unveiled: Techniques, Challenges, and Prospects for Structured Information Extraction,” 2024, doi: 10.48550/arXiv.2410.21169.
- [28] S. J. Russell and P. Norvig, *Artificial Intelligence A Modern Approach Fourth Edition Global Edition*. 2022.
- [29] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” *SN Comput. Sci.*, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.
- [30] Y. K. Dwivedi *et al.*, “Artificial Intelligence (AI): Multidisciplinary

- perspectives on emerging challenges, opportunities, and agenda for research, practice and policy,” *Int. J. Inf. Manage.*, 2021, doi: 10.1016/j.ijinfomgt.2019.08.002.
- [31] D. Jurafsky and J. H. Martin, “Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition,” Stanford University. [Online]. Available: <https://web.stanford.edu/~jurafsky/slp3>
- [32] D. Khurana, A. Koli, K. Khatter, and S. Singh, “Natural language processing: state of the art, current trends and challenges,” *Multimed Tools Appl* 82, vol. 82, pp. 3713–3744, 2023, doi: 10.1007/s11042-022-13428-4.
- [33] Z. Nussbaum, J. X. Morris, B. Duderstadt, and A. Mulyar, “Nomic Embed : Training a Reproducible Long Context Text Embedder,” *Trans. Mach. Learn. Res.*, pp. 1–17, 2025, doi: 10.48550/arXiv.2402.01613.
- [34] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, “Deep Learning Based Text Classification: A Comprehensive Review,” *ACM Comput. Surv.*, vol. 54, no. 3, pp. 1–43, 2021, doi: 10.1145/3439726.
- [35] S. E. Prasetyo, V. A. Puteri, and Sabariman, “Implementasi Chatbot AI dan Whatsapp Untuk Mendukung Penjualan dan Konsultasi Rakitan Komputer di Toko Micro Batam,” *J. Inf. Syst. Manag.*, vol. 7, no. 1, 2025.
- [36] R. D. A. Raya and F. L. D. Cahyanti, “Perancangan Sistem Informasi Chatbot Retrieval Augmented Generation Berbasis Website Pada PT . Revolusi Cita Edukasi,” *Indones. J. Comput. Sci.*, vol. 4, no. 1, pp. 15–21, 2025, doi: 10.31294/m75d4782.
- [37] Kholisatuunisa and D. N. Misidawati, “Peran Teknologi AI dalam Pengembangan Chatbot Layanan Pelanggan,” *J. Sist. Informasi, Manaj. dan Teknol. Inf.*, vol. 2, no. 2, pp. 92–97, 2024.
- [38] R. I. Fitria, R. P. Tulodo, and A. Sofyan, “Pemanfaatan chatbot untuk informasi pendaftaran calon mahasiswa baru di perguruan tinggi,” *J. Penerapan Teknol. Inf. dan Komun.*, vol. 4, no. 2, pp. 221–232, 2025, doi: 10.24246/itexplore.v4i2.2025.
- [39] M. U. Hadi *et al.*, “Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects,” *TechRxiv*,

- 2025, doi: 10.36227/techrxiv.23589741.v8.
- [40] H. Naveed *et al.*, “A Comprehensive Overview of Large Language Models,” *ACM Trans. Intell. Syst. Technol.*, vol. 16, no. 5, pp. 1–72, 2025, doi: 10.1145/3744746.
 - [41] A. Baumann and P. Eberhard, “Experiments with Large Language Models on Retrieval- Augmented Generation for Closed-Source Simulation Software,” *Multibody Syst. Dyn.*, 2025.
 - [42] G. Pan, V. Chodnekar, A. Roy, and H. Wang, “A Cost-Benefit Analysis of On-Premise Large Language Model Deployment: Breaking Even with Commercial LLM Services,” 2025, doi: 10.48550/arXiv.2509.18101.
 - [43] Y. Chang *et al.*, “A Survey on Evaluation of Large Language Models,” *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 3, 2024, doi: 10.1145/3641289.
 - [44] P. Lewis *et al.*, “Retrieval-augmented generation for knowledge-intensive NLP tasks,” *NIPS’20 Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, pp. 9459–9474, 2021.
 - [45] A. Mishra and N. Brahmanapally, “A Comparative Performance Analysis of Locally Deployed Large Language Models Through a Retrieval-Augmented Generation Educational Assistant Application for Textual Data Extraction,” *AI*, vol. 6, no. 6, 2025, doi: 10.3390/ai6060119.
 - [46] C. Niu *et al.*, “RAGTruth: A Hallucination Corpus for Developing Trustworthy Retrieval-Augmented Language Models,” *Proc. Annu. Meet. Assoc. Comput. Linguist.*, vol. 1, 2024, doi: 10.18653/v1/2024.acl-long.585.
 - [47] C. Sharma, “Retrieval-Augmented Generation: A Comprehensive Survey of Architectures, Enhancements, and Robustness Frontiers,” *ACM TOIS*, pp. 1–32, 2025, doi: 10.48550/arXiv.2506.00054.
 - [48] J. Chen, H. Lin, X. Han, and L. Sun, “Benchmarking Large Language Models in Retrieval-Augmented Generation,” 2023, doi: 10.1609/aaai.v38i16.29728.
 - [49] M. Ghali, A. Farrag, D. Won, and Y. Jin, “Enhancing knowledge retrieval with in-context learning and semantic search through generative AI,” vol. 311, 2025, doi: 10.1016/j.knosys.2025.113047.
 - [50] G. P. Rusum and S. Anasuri, “Vector Databases in Modern Applications :

- Real-Time Search, Recommendations, and Retrieval-Augmented Generation (RAG),” *Int. J. AI, BigData, Comput. Manag. Stud.*, vol. 5, no. 4, pp. 124–136, 2024, doi: 10.63282/3050-9416.IJAIBDCMS-V5I4P113.
- [51] S. Kiran, “Hybrid Retrieval-Augmented Generation (RAG) Systems with Embedding Vector Databases,” *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol.*, vol. 11, no. 2, pp. 2694–2702, 2025, doi: 10.32628/CSEIT25112702.
- [52] Z. Jing, Y. Su, and Y. Han, “When Large Language Models Meet Vector Databases: A Survey,” *Proc. - 2025 Conf. Artif. Intell. x Multimedia, AIXMM 2025*, pp. 7–13, 2025, doi: 10.1109/AIXMM62960.2025.00008.
- [53] L. Ma, R. Zhang, Y. Han, S. Yu, and Z. Wang, “A Comprehensive Survey on Vector Database: Storage and Retrieval Technique, Challenge,” pp. 1–25, 2025, doi: 10.48550/arXiv.2310.11703.
- [54] J. J. Pan, J. Wang, and G. Li, “Survey of vector database management systems,” *VLDB J.*, vol. 33, no. 5, pp. 1591–1615, 2024, doi: 10.1007/s00778-024-00864-x.
- [55] F. Liu, Z. Kang, and X. Han, “Optimizing RAG Techniques for Automotive Industry PDF Chatbots: A Case Study with Locally Deployed Ollama Models Optimizing,” 2025, doi: 10.1145/3707292.3707358.
- [56] E. J. Husom *et al.*, “Sustainable LLM Inference for Edge AI: Evaluating Quantized LLMs for Energy Efficiency, Output Accuracy, and Inference Latency,” *ACM Trans. Internet Things*, 2025, doi: 10.1145/3767742.
- [57] V. Rajesh *et al.*, “Production-Grade Local LLM Inference on Apple Silicon: A Comparative Study of MLX, MLC-LLM, Ollama, llama.cpp, and PyTorch MPS,” 2025, doi: 10.48550/arXiv.2511.05502.
- [58] C. D. Manning, P. Raghavan, and H. Schütze, “Introduction to Modern Information Retrieval (2nd edition),” *Libr. Rev.*, vol. 53, no. 9, pp. 462–463, 2004, doi: 10.1108/00242530410565256.
- [59] W. B. Croft, D. Metzler, and T. Strohman, *Information Retrieval in Practice*. 2015. doi: 10.1016/0268-4012(89)90040-6.
- [60] J. Han, M. Kamber, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.

- [61] S. Es, J. James, L. Espinosa-anke, and S. Schockaert, “Ragas: Automated Evaluation of Retrieval Augmented Generation,” *Proc. 18th Conf. Eur. Chapter Assoc. Comput. Linguist. Syst. Demonstr.*, pp. 150–158, 2024, doi: 10.18653/v1/2024.eacl-demo.16.
- [62] M. R. Anama, A. S. Akbara, and H. Saputro, “QnA Chatbot with Mistral 7B and RAG method: Traffic Law Case Study,” *Lontar Komput.*, vol. 15, no. 3, pp. 207–218, 2024, doi: 10.24843/LKJITI.2024.v15.i03.p06.
- [63] I. Prastowo, H. Putro, J. Antoni, M. K. Adhitya, N. A. Herawati, and N. P. Utama, “Retrieval-Augmented Generation (RAG) Chatbot for Handling Customer Complaints in the Energy Sector,” *J. Infomedia Tek. Inform. Multimedia, dan Jar.*, vol. 10, no. 2, pp. 105–111, 2025, doi: 10.30811/jim.v10i2.7169.
- [64] L.-L. Ebidor and I. G. Ikhide, “Literature Review in Scientific Research: An Overview,” *East African J. Educ. Stud.*, vol. 7, no. 2, pp. 179–186, 2024, doi: 10.37284/eajes.7.2.1909.
- [65] 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.
- [66] A. Yasin, K. Kaya, and K. Kilic, “A Retrieval-Augmented Generation Framework for Academic Literature Navigation in Data Science”.