

DAFTAR PUSTAKA

- Agnihotram G, Sarkar J. 2025. Evaluating Precision and Recall at Retrieval Time in Retrieval-Augmented Generation (RAG) Systems . *American Journal of Computer Science and Technology*. 8(4):174–180. <https://doi.org/10.11648/j.ajcst.20250804.11>.
- Ahadi R, Safaat Harahap N, Fikry M, Kurnia F. 2025. Retrieval-Augmented Generation in a Web-Based Question Answering System for Fiqh Books. *Journal of Artificial Intelligence and Software Engineering*. 5(2):626–635. <https://doi.org/10.30811/jaise.v5i2.7005>.
- Ahmad MA. 2025 Des 31. Islamic Chatbots in the Age of Large Language Models. <http://arxiv.org/abs/2601.06092>.
- Ammann P, Offutt J. 2017. *Introduction to Software Testing*. Revised 2nd. Cambridge University Press.
- Asl MA, Minaei-Bidgoli B. 2025. FARSIQA: Faithful & Advanced RAG System for Islamic Question Answering.
- Auffarth B. 2023. *Generative AI with LangChain: Build large language model (LLM) apps with Python*. 2024 Edition. Birmingham: Packt Publishing. [diakses 2025 Nov 25]. [https://books.google.co.id/books?id=cOrpEAAAQBAJ&pg=PA41&dq=Generative+AI+with+LangChain:+Build+large+language+model+\(LLM\)+apps+with+Python,+ChatGPT,+and+other+LLMs&hl=id&newbks=1&newbks_redir=0&sa=X&ved=2ahUKEwiAjbGkoYuRAXrS3ADHSQ5MqAQ6AF6BAgGEAM#v=onepage&q=Generative%20AI%20with%20LangChain%3A%20Build%20large%20language%20model%20\(LLM\)%20apps%20with%20Python%2C%20ChatGPT%2C%20and%20other%20LLMs&f=false](https://books.google.co.id/books?id=cOrpEAAAQBAJ&pg=PA41&dq=Generative+AI+with+LangChain:+Build+large+language+model+(LLM)+apps+with+Python,+ChatGPT,+and+other+LLMs&hl=id&newbks=1&newbks_redir=0&sa=X&ved=2ahUKEwiAjbGkoYuRAXrS3ADHSQ5MqAQ6AF6BAgGEAM#v=onepage&q=Generative%20AI%20with%20LangChain%3A%20Build%20large%20language%20model%20(LLM)%20apps%20with%20Python%2C%20ChatGPT%2C%20and%20other%20LLMs&f=false).
- Azarmi B, Vestal J, editor. 2023. *Vector Search for Practitioners with Elastic: A toolkit for building NLP solutions for search, observability, and security using vector search*. Birmingham: Packt Publishing. [diakses 2025 Nov 25]. https://books.google.co.id/books?id=ZbPIEAAAQBAJ&pg=PR14&dq=Vector+Search+for+Practitioners+with+Elastic&hl=id&newbks=1&newbks_redir=0&sa=X&ved=2ahUKEwjD87fnpIuRAXWnzzgGHTVmEhAQ6AF6BAgNEAM#v=onepage&q=Vector%20Search%20for%20Practitioners%20with%20Elastic&f=false.
- Bhojani AR, Schwarting M. 2023. Truth and Regret: Large Language Models, the Quran, and Misinformation. *Theology and Science*. 21(4):557–563. <https://doi.org/10.1080/14746700.2023.2255944>.

- Boucekif A, Gaben S, Rashwani S, Eltanbouly S, Al-Khatib M, Sbahi H, Ghaly M, Mohamed E. 2026 Jun 15. MAWARITH: A Dataset and Benchmark for Legal Inheritance Reasoning with LLMs. <http://arxiv.org/abs/2603.07539>.
- Chakraborty T. 2025. *Introduction to Large Language Models: Generative AI for Text*. New Delhi: Wiley India Pvt. Ltd.
- Es S, James J, Espinosa Anke L, Schockaert S. 2024. RAGAs: Automated Evaluation of Retrieval Augmented Generation. Di dalam: Aletras N, De Clercq O, editor. *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. St. Julians, Malta: Association for Computational Linguistics. hlm 150–158.
- Harmawati Y, Sapriya, Abdulkarim A, Bestari P, Sari BI. 2024. Data of digital literacy level measurement of Indonesian students: Based on the components of ability to use media, advanced use of digital media, managing digital learning platforms, and ethics and safety in the use of digital media. *Data Brief*. 54. <https://doi.org/10.1016/j.dib.2024.110397>.
- Hokamp C, Liu Q. 2017. Lexically Constrained Decoding for Sequence Generation Using Grid Beam Search. Di dalam: Barzilay R, Kan M-Y, editor. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Vancouver, Canada: Association for Computational Linguistics. hlm 1535–1546.
- Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, Chen Q, Peng W, Feng X, Qin B, *et al.* 2025. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions. *ACM Trans Inf Syst*. 43(2). <https://doi.org/10.1145/3703155>.
- Jeong S, Baek J, Cho S, Hwang SJ, Park J. 2024. Adaptive-RAG: Learning to Adapt Retrieval-Augmented Large Language Models through Question Complexity. Di dalam: Duh K, Gomez H, Bethard S, editor. *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. Mexico City, Mexico: Association for Computational Linguistics. hlm 7036–7050.
- Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, Ishii E, Bang YJ, Madotto A, Fung P. 2023. Survey of Hallucination in Natural Language Generation. *ACM Comput Surv*. 55(12). <https://doi.org/10.1145/3571730>.
- Jurafsky D, Martin JH. 2024. *Speech and Language Processing*. 3rd Edition draft. Stanford University / Pearson. <https://web.stanford.edu/~jurafsky/slp3/>.

- Karpukhin V, Oguz B, Min S, Lewis P, Wu L, Edunov S, Chen D, Yih W. 2020. Dense Passage Retrieval for Open-Domain Question Answering. Di dalam: Webber B, Cohn T, He Y, Liu Y, editor. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics. hlm 6769–6781.
- Khalila Z, Nasution AH, Monika W, Onan A, Murakami Y, Radi YBI, Osmani NM. 2025 Mar 20. Investigating Retrieval-Augmented Generation in Quranic Studies: A Study of 13 Open-Source Large Language Models.
- Lewis JR. 2002. Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *Int J Hum Comput Interact*. 14(3–4):463–488. <https://doi.org/10.1080/10447318.2002.9669130>.
- Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, Küttler H, Lewis M, Yih W, Rocktäschel T, *et al*. 2021 Apr 12. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. <http://arxiv.org/abs/2005.11401>.
- Li H, Liu Y, Cheng Y, Ray S, Du K, Jiang J. 2024. Eloquent: A More Robust Transmission Scheme for LLM Token Streaming. Di dalam: *Proceedings of the 2024 SIGCOMM Workshop on Networks for AI Computing*. New York, NY, USA: Association for Computing Machinery. hlm 34–40.
- Lin J. 2024 Sep 10. Operational Advice for Dense and Sparse Retrievers: HNSW, Flat, or Inverted Indexes? <http://arxiv.org/abs/2409.06464>.
- Malkov YA, Yashunin DA. 2020. Efficient and Robust Approximate Nearest Neighbor Search Using Hierarchical Navigable Small World Graphs. *IEEE Trans Pattern Anal Mach Intell*. 42(4):824–836. <https://doi.org/10.1109/TPAMI.2018.2889473>.
- Manning CD, Raghavan P, Schütze H. 2008. *Introduction to Information Retrieval*. Cambridge: Cambridge University Press. [diakses 2025 Nov 25]. <https://books.google.co.id/books?id=t1PoSh4uwVcC&printsec=frontcover&hl=id#v=onepage&q&f=false>.
- Maulana MR, Harahap NS, Okfalisa O, Yusra Y. 2025. Implementasi Chatbot Tafsir Al-Qur'an Menggunakan Chainlit dengan Pendekatan Groq. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*. 5(3):920–929. <https://doi.org/10.57152/malcom.v5i3.2082>.
- Mohammed MY, Ali SA, Ali SK, Majeed AA, Mohamed EH. 2025. Aftina: enhancing stability and preventing hallucination in AI-based Islamic fatwa generation using LLMs and RAG. *Neural Comput Appl*. 37(25):20957–20982. <https://doi.org/10.1007/s00521-025-11229-y>.

- Mubarak H, Malhas R, Mansour W, Mohamed A, Fawzi M, Hawasly M, Elsayed T, Darwish KM, Magdy W. 2025. IslamicEval 2025: The First Shared Task of Capturing LLMs Hallucination in Islamic Content.
- Pradana MG, Gusti KW, Prasetyo A, Saputro PH, Wijaya DP, Heksaputra D, Syahputra R, Kaharuddin K, Matondang N, Pramuntadi A. 2024. *Information Retrieval*. PT Penamuda Media.
- Pressman RS, Maxim BR. 2020. *Software Engineering: A Practitioner's Approach, 9/E*. Ed ke-9. New York: McGraw-Hill Education. [diakses 2025 Nov 25]. <https://openlibrary.telkomuniversity.ac.id/home/catalog/id/200715/slug/connect-access-card-for-software-engineering-a-practitioner-s-approach-9-e-.html>.
- Qdrant Team. 2023. Filtered Search Benchmark.
- Qdrant Team. 2024. Qdrant Essentials Course: What is HNSW?
- Qdrant Team. 2026. Storage Optimizers - Qdrant Documentation. <https://qdrant.tech/documentation/ops-optimization/optimizer/>.
- Ryding K, Wilmsen D. 2021. *Cambridge Handbook of Arabic Linguistics*.
- Singh IS, Aggarwal R, Allahverdiyev I, Taha M, Akalin A, Zhu K, O'Brien S. 2025. ChunkRAG: Novel LLM-Chunk Filtering Method for RAG Systems. <https://arxiv.org/abs/2410.19572>.
- Sommerville I. 2016. *Software Engineering, 10/E*. Boston: Pearson Education. [diakses 2025 Nov 25]. <https://openlibrary.telkomuniversity.ac.id/pustaka/166668/software-engineering-10-e.html>.
- StatCounter. 2026. Mobile Operating System Market Share Indonesia. *StatCounter Global Stats.*, siap terbit.
- Tunstall L, Werra L von, Wolf T. 2022. *Natural Language Processing with Transformers*. Sebastopol: O'Reilly Media. [diakses 2025 Nov 25]. https://books.google.co.id/books?id=nTxbEAAAQBAJ&pg=PT1&dq=Natural+Language+Processing+with+Transformers&hl=id&newbks=1&newbks_r edir=0&sa=X&ved=2ahUKEwjUnffinouRAxXBbmwGHfRJJAQ6AF6BA gKEAM#v=onepage&q=Natural%20Language%20Processing%20with%20 Transformers&f=false.
- Wang L, Yang N, Huang X, Jiao B, Yang L, Jiang D, Majumder R, Wei F. 2024a Feb 22. Text Embeddings by Weakly-Supervised Contrastive Pre-training. <http://arxiv.org/abs/2212.03533>.
- Wang L, Yang N, Huang X, Yang L, Majumder R, Wei F. 2024b Feb 8. Multilingual E5 Text Embeddings: A Technical Report. <http://arxiv.org/abs/2402.05672>.

WHATWG. 2026. HTML Living Standard: Server-Sent Events.

Zhao P, Zhang H, Yu Q, Wang Z, Geng Y, Fu F, Yang L, Zhang W, Jiang J, Cui B.
2024 Jun 21. Retrieval-Augmented Generation for AI-Generated Content: A
Survey. <http://arxiv.org/abs/2402.19473>.