UNIVERSITAS
GADJAH MADA
**Pengembangan model chatbot hukum berbasis hybrid retrieval augmented generation (RAG) dengan metode
hybrid retrieval dan MT5 terkait Undang-Undang ITE di Indonesia**
Robertus Anom Wibisono, Dr. Sri Mulyana M. Kom
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

DAFTAR PUSTAKA

Amato, F., Fonisto, M., Giacalone, M. dan Sansone, C., 2023, An intelligent conversational agent for the legal domain, *Information*, 14(6), 307, https://doi.org/10.3390/info14060307

Ashman, R., 2024, The aRt of RAG Part 4: Retrieval evaluation, Medium, https://medium.com/@rossashman/the-art-of-rag-part-4-retrieval-evaluation-427bb5db0475, 30 Mei 2024, diakses 12 April 2026.

Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R.C., Mellor, S., Schwaber, K., Sutherland, J. dan Thomas, D., 2001, *Manifesto for Agile Software Development*, http://agilemanifesto.org/

Bergstra, J., & Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research*, 13(10), 281–305

Cahn, J., 2017, Chatbot: Architecture, design, & development, Skripsi Senior, University of Pennsylvania, School of Engineering and Applied Science, Philadelphia.

Cahyani, L.D. dan Gischa, S., 2025, Perkembangan kecerdasan buatan (AI): Inovasi tanpa batas, *Kompas.com*, https://www.kompas.com/stori/read/2025/09/17/230000679/perkembangan-kecerdasan-buatan-ai---inovasi-tanpa-batas, 17 September 2025, diakses 25 Maret 2026.

Devlin, J., Chang, M.W., Lee, K. dan Toutanova, K., 2019, BERT: Pre-training of deep bidirectional transformers for language understanding, dalam *Proceedings of NAACL-HLT 2019*, hlm. 4171–4186, Association for Computational Linguistics, https://doi.org/10.18653/v1/N19-1423

Di Iorio, A., 2007, Segmentation patterns pentaformat IML digital documents, Disertasi Doktoral, Università di Bologna, Bologna, https://doi.org/10.6092/unibo/amsdottorato/370

Douze, M., Guzhva, A., Deng, C., Johnson, J., Szilvasy, G., Mazaré, P.-E., Lomeli, M., Hosseini, L. dan Jégou, H., 2025, The FAISS library, *IEEE Transactions on Big Data*, advance online publication, https://doi.org/10.1109/TBDATA.2025.3618474

Dybå, T. dan Dingsøyr, T., 2008, Empirical studies of agile software development: A systematic review, *Information and Software Technology*, 50(9–10), 833–859.

Feng, F., Yang, Y., Cer, D., Arivazhagan, N. dan Wang, W., 2022, Language-agnostic BERT sentence embedding, dalam *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, hlm. 878–891, https://arxiv.org/abs/2007.01852

UNIVERSITAS
GADJAH MADA
**Pengembangan model chatbot hukum berbasis hybrid retrieval augmented generation (RAG) dengan metode
hybrid retrieval dan MT5 terkait Undang-Undang ITE di Indonesia**
Robertus Anom Wibisono, Dr. Sri Mulyana M. Kom
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Freitag, M. dan Al-Onaizan, Y., 2017, Beam search strategies for neural machine translation, dalam *Proceedings of the First Workshop on Neural Machine Translation*, Vancouver, Canada, hlm. 56–60, https://arxiv.org/abs/1702.01806, 7 Februari 2017 diakses 21 April 2026.

Harris, Z.S., 1954, Distributional structure, *WORD*, 10(2–3), 146–162, https://doi.org/10.1080/00437956.1954.11659520

Järvelin, K. dan Kekäläinen, J., 2002, Cumulated gain-based evaluation of IR techniques, *ACM Transactions on Information Systems*, 20(4), 422–446, https://doi.org/10.1145/582415.582418

Johnson, J., Douze, M. dan Jégou, H., 2019, Billion-scale similarity search with GPUs, *IEEE Transactions on Big Data*, 7(3), 535–547, https://arxiv.org/abs/1702.08734, 28 February 2019, diakses 21 April 2026.

Jurafsky, D. dan Martin, J.H., 2024, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, edisi ke-3 draf, https://web.stanford.edu/~jurafsky/slp3, diakses 2 November 2025.

Katz, D.M., Hartung, D., Gerlach, L., Jana, A. dan Bommarito, M.J., 2023, Natural language processing in the legal domain, *arXiv*:2302.12039, https://doi.org/10.48550/arXiv.2302.12039

Lewis, P., Perez, E., Piktus, A., Petroni, F., Rocktäschel, T., Shuster, K., Wu, Y., Miller, A., Weston, J. dan Kiela, D., 2020, Retrieval-augmented generation for knowledge-intensive NLP tasks, dalam *Advances in Neural Information Processing Systems (NeurIPS 2020)*, 33, hlm. 9459–9474, https://arxiv.org/abs/2005.11401, 22 May 2020, diakses 21 April 2026.

Lin, C.-Y., 2004, ROUGE: A package for automatic evaluation of summaries, dalam *Proceedings of the ACL Workshop on Text Summarization Branches Out*, Barcelona, Spanyol, hlm. 74–81, https://aclanthology.org/W04-1013/, Juli 2004, diakses 21 April 2026.

Loshchilov, I. dan Hutter, F., 2019, Decoupled weight decay regularization, dalam *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019)*, New Orleans, USA, https://arxiv.org/abs/1711.05101, 14 November 2019, diakses 21 April 2026.

Manning, C.D., Raghavan, P. dan Schütze, H., 2008, *Introduction to Information Retrieval*, Cambridge University Press, Cambridge, https://nlp.stanford.edu/IR-book/

Mohammed, M. dan Aref, M.M., 2020, Chatbot system architecture, *EasyChair Preprints*, 2961, Ain Shams University.

UNIVERSITAS
GADJAH MADA
**Pengembangan model chatbot hukum berbasis hybrid retrieval augmented generation (RAG) dengan metode
hybrid retrieval dan MT5 terkait Undang-Undang ITE di Indonesia**
Robertus Anom Wibisono, Dr. Sri Mulyana M. Kom
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Ott, M., Edunov, S., Grangier, D. dan Auli, M., 2018, Scaling neural machine translation, dalam *Proceedings of the Third Conference on Machine Translation (WMT 2018)*, Brussels, Belgia, hlm. 1–9, https://arxiv.org/abs/1806.00187, 1 Juni 2018, diakses 21 April 2026.

Palmer, M.C., 2022, Ethical considerations of legal chatbots, *Attorney at Work*, https://www.attorneyatwork.com/legal-chatbots-ethical-considerations-for-law-firms/

Paulus, R., Xiong, C. dan Socher, R., 2018, A deep reinforced model for abstractive summarization, dalam *Proceedings of the 6th International Conference on Learning Representations (ICLR 2018)*, Vancouver, Kanada, https://arxiv.org/abs/1705.04304, 11 Mei 2018, diakses 21 April 2026.

Perry, D.E. dan Wolf, A.L., 1992, Foundations for the study of software architecture, *ACM SIGSOFT Software Engineering Notes*, 17(4), 40–52, https://doi.org/10.1145/141874.141884

Pinecone, 2023, Evaluation measures in information retrieval, https://www.pinecone.io/learn/offline-evaluation/, 30 Juni 2023, diakses 20 April 2026.

Pressman, R.S., 2010, *Software Engineering: A Practitioner's Approach*, edisi ke-7, McGraw-Hill, New York.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. dan Liu, P.J., 2020, Exploring the limits of transfer learning with a unified text-to-text transformer, *Journal of Machine Learning Research*, 21(140), 1–67, https://arxiv.org/abs/1910.10683, 23 October 2020, diakses 23 Maret 2026.

Ramesh, S., 2022, Evaluating your NLP systems, *Towards Data Science*, https://towardsdatascience.com/evaluating-your-nlp-systems-2ba0d3f09f04/, 21 September 2022, diakses 12 Mei 2026

Reimers, N. dan Gurevych, I., 2019, Sentence-BERT: Sentence embeddings using siamese BERT-networks, dalam *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, Hong Kong, hlm. 3982–3992, https://arxiv.org/abs/1908.10084 , 27 August 2019, diakses 21 April 2026.

Robertson, S.E. dan Walker, S., 1994, Some simple effective approximations to the 2-Poisson model for probabilistic weighted retrieval, dalam *SIGIR '94: Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, hlm. 232–241, Springer, London, https://doi.org/10.1007/978-1-4471-2099-5_24

UNIVERSITAS
GADJAH MADA
**Pengembangan model chatbot hukum berbasis hybrid retrieval augmented generation (RAG) dengan metode
hybrid retrieval dan MT5 terkait Undang-Undang ITE di Indonesia**
Robertus Anom Wibisono, Dr. Sri Mulyana M. Kom
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Robertson, S.E. dan Zaragoza, H., 2009, The probabilistic relevance framework: BM25 and beyond, Foundations and Trends in Information Retrieval, 3(4), 333–389

Robertson, S.E., 1977, The probability ranking principle in IR, *Journal of Documentation*, 33(4), 294–304, https://doi.org/10.1108/eb026647

Rosa, G.M., Rodrigues, R.C., Lotufo, R.A. dan Nogueira, R., 2021, Yes, BM25 is a strong baseline for legal case retrieval, *arXiv*:2105.05686, https://doi.org/10.48550/arXiv.2105.05686

Rozaq, A., Fitri, R., Akbar, M., Sumari, M., Khafiyya, N., Wardiah, I. dan Sholihin, F., 2025, Legal literacy in Indonesia: Leveraging semantic-based AI and NLP for enhanced civil law access, *E3S Web of Conferences*, 622, 03002, https://doi.org/10.1051/e3sconf/202562203002

Salton, G., Wong, A. dan Yang, C.S., 1975, A vector space model for automatic indexing, *Communications of the ACM*, 18(11), 613–620, https://doi.org/10.1145/361219.361220

Sansonnet, J.-P., Leray, D. dan Martin, J.-C., 2006, Architecture of a framework for generic assisting conversational agents, dalam J. Gratch dkk. (Ed.), *Intelligent Virtual Agents (IVA 2006)*, Lecture Notes in Artificial Intelligence, Vol. 4133, hlm. 145–156, Springer, https://doi.org/10.1007/11821830_13

Sawarkar, K., Mangal, A. dan Rane, S., 2024, Blended RAG: Improving RAG accuracy with semantic search and hybrid query-based retrievers, *arXiv*:2404.07220, https://arxiv.org/abs/2404.07220

Sommerville, I., 2011, *Software Engineering*, edisi ke-9, Pearson, Boston.

Sultania, D., Lu, Z., Naik, T., Dernoncourt, F., Yoon, S., Sharma, S. dan Friedrich, M., 2024, Domain-specific question answering with hybrid search, *arXiv*:2412.03736, https://doi.org/10.48550/arXiv.2412.03736

Sutskever, I., Vinyals, O. dan Le, Q.V., 2014, Sequence to sequence learning with neural networks, dalam Advances in Neural Information Processing Systems (NeurIPS 2014), 27, hlm. 3104–3112, https://arxiv.org/abs/1409.3215, 10 September 2014, diakses 21 April 2026.

Towards Data Science, 2025, How to evaluate retrieval quality in RAG pipelines: Precision@k, Recall@k, and F1@k, https://towardsdatascience.com/how-to-evaluate-retrieval-quality-in-rag-pipelines-precisionk-recallk-and-f1k/

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. dan Polosukhin, I., 2017, Attention is all you need, dalam *Advances in Neural Information Processing Systems (NeurIPS 2017)*, 30, https://doi.org/10.48550/arXiv.1706.03762

UNIVERSITAS
GADJAH MADA
**Pengembangan model chatbot hukum berbasis hybrid retrieval augmented generation (RAG) dengan metode
hybrid retrieval dan MT5 terkait Undang-Undang ITE di Indonesia**
Robertus Anom Wibisono, Dr. Sri Mulyana M. Kom
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

Voorhees, E.M., 1999, The TREC-8 question answering track report, dalam *Proceedings of the 8th Text Retrieval Conference (TREC-8)*, hlm. 77–82, NIST, https://trec.nist.gov/pubs/trec8/papers/overview.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., Drame, M., Lhoest, Q. dan Rush, A.M., 2020, Transformers: State-of-the-art natural language processing, dalam Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020): System Demonstrations, Online, hlm. 38–45, https://arxiv.org/abs/1910.03771, 9 Oktober 2020, diakses 21 April 2026.

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A. dan Raffel, C., 2021, mT5: A massively multilingual pre-trained text-to-text transformer, dalam Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT 2021), Online, hlm. 483–498, https://arxiv.org/abs/2010.11934, 22 Oktober 2021, diakses 16 April 2026.

Yulianti, E., Bhary, N., Abdurrohman, J., Dwitilas, F.W., Nuranti, E.Q. dan Husin, H.S., 2024, Named entity recognition on Indonesian legal documents: A dataset and study using transformer-based models, *International Journal of Electrical and Computer Engineering*, 14(5), 5489–5501, https://doi.org/10.11591/ijece.v14i5.pp5489-5501