

**PENGEMBANGAN SISTEM CHATBOT FIKIH BERBASIS NLP
DENGAN PENDEKATAN RAG MENGGUNAKAN KORPUS KITAB
SAFINATUN NAJAH**

SKRIPSI

Diajukan Untuk Memenuhi Salah Satu Syarat Guna Memperoleh Gelar Sarjana
Komputer (S.Kom) Pada Program Studi Teknik Informatika Fakultas Teknik dan
Ilmu Komputer Universitas Nusantara PGRI Kediri



Oleh :

Muhammad Ichsan

NPM : 2213020050

**FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS NUSANTARA PGRI KEDIRI**

2026

Skripsi oleh:

Muhammad Ichsan
NPM : 2213020050

Judul :

**PENGEMBANGAN SISTEM CHATBOT FIKIH BERBASIS NLP
DENGAN PENDEKATAN RAG MENGGUNAKAN KORPUS KITAB
SAFINATUN NAJAH**

Telah Disetujui Untuk Diajukan Kepada Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer
Universitas Nusantara PGRI Kediri

Tanggal : 14 Juni 2026

Pembimbing I



Ardi Sanjaya, S.Kom, M.Kom.
NIDN. 0706118101

Pembimbing II



Ahmad Bagus Setiawan, S.T, M.M, M.Kom.
NIDN. 0703018704

Skripsi oleh:

Muhammad Ichsan
NPM : 2213020050

Judul :

**PENGEMBANGAN SISTEM CHATBOT FIKIH BERBASIS NLP
DENGAN PENDEKATAN RAG MENGGUNAKAN KORPUS KITAB
SAFINATUN NAJAH**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi
Program Studi Teknik Informatika Fakultas Teknik dan Ilmu Komputer

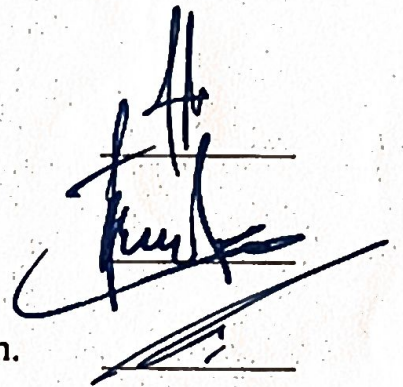
Universitas Nusantara PGRI Kediri

Pada tanggal : 14 Juli 2026

Dan Dinyatakan telah Memenuhi Syarat

Panitia Penguji :

1. Ketua : Ardi Sanjaya, S.Kom, M.Kom.
2. Penguji I : Dr. Risky Aswi Ramadhani, M. Kom.
3. Penguji II : Ahmad Bagus Setiawan, S.T, M.M, M.Kom.



Mengetahui,
Dekan FTIK

Dr. Sulistiono, M.si
NIDN. 0007076801

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Muhammad Ichsan
Jenis Kelamin : Laki-Laki
Tempat/Tgl Lahir : Palembang, 31 Desember 2003
NPM : 2213020050
Fakultas/Prodi : Teknik dan Ilmu Komputer/ Teknik Informatika

Menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak dapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 14 Juli 2026
Yang Menyatakan



Muhammad Ichsan

NPM : 2213020050

HALAMAN PERSEMBAHAN

Penulisan skripsi ini dengan tulus saya dedikasikan kepada:

1. Kedua orang tua saya, yang dengan penuh kesabaran senantiasa mendoakan, memberikan dukungan terbaik, serta menjadi sumber motivasi tak henti-hentinya dalam perjalanan menyelesaikan skripsi ini.
2. Adik-adik saya, yang selalu memberikan semangat dan dukungan moral, sehingga saya dapat menyelesaikan proses ini dengan baik.
3. Seluruh dosen Universitas Nusantara PGRI Kediri, yang telah memberikan bimbingan, ilmu, dan pelajaran berharga, baik dalam ranah akademik maupun kehidupan sehari-hari.
4. Teman-teman seperjuangan di kampus, yang menjadi tempat berbagi suka dan duka, serta saling menyemangati selama menjalani masa perkuliahan hingga tahap akhir ini.
5. Almamater tercinta, Universitas Nusantara PGRI Kediri, yang telah menjadi tempat saya tumbuh, belajar, dan mengembangkan diri selama masa perkuliahan.
6. Semua pihak yang tidak dapat saya sebutkan satu per satu, yang telah memberikan kontribusi dalam berbagai bentuk dukungan selama proses penyusunan skripsi ini.

Semoga dedikasi ini dapat menjadi penghormatan atas semua dukungan, doa, dan kebersamaan yang telah diberikan.

HALAMAN MOTTO

“Barang siapa menempuh suatu jalan untuk mencari ilmu, niscaya Allah akan memudahkan baginya jalan menuju surga” — (HR. Muslim No. 2699)

RINGKASAN

Muhammad Ichsan, *Pengembangan Chatbot Fiqih Berbasis Retrieval-Augmented Generation (RAG) Menggunakan Kitab Safinatun Najah sebagai Media Pembelajaran Santri*, Skripsi, Program Studi Teknik Informatika, Fakultas Teknik, Universitas Nusantara PGRI Kediri, 2026.

Kata Kunci: Chatbot, Fiqih, Large Language Model, Retrieval-Augmented Generation, Safinatun Najah.

Penelitian ini bertujuan mengembangkan chatbot fikih berbasis *Retrieval-Augmented Generation* (RAG) menggunakan kitab *Safinatun Najah* sebagai media pendukung pembelajaran santri. Sistem dikembangkan menggunakan bahasa pemrograman Python dengan framework Streamlit dan LangChain, model embedding *intfloat/multilingual-e5-large*, ChromaDB sebagai *vector database*, serta Google Gemini 2.5 Flash sebagai *Large Language Model*. Proses *retrieval* dilakukan menggunakan *cosine similarity* untuk memperoleh dokumen yang paling relevan, kemudian digunakan sebagai konteks dalam menghasilkan jawaban berdasarkan isi kitab. Hasil pengujian terhadap 250 data uji menunjukkan nilai *Accuracy*, *Precision*, *Recall*, dan *F1-Score* sebesar 76,8% dengan rata-rata *Cosine Similarity* sebesar 0,7644. Sistem mampu memberikan jawaban yang relevan disertai teks Arab, terjemahan, dan sumber rujukan kitab, sehingga dapat membantu santri memperoleh informasi fikih secara lebih cepat, akurat, dan tetap sesuai dengan kitab *Safinatun Najah* sebagai sumber utama.

PRAKATA

Puji Syukur dipanjatkan kepada Tuhan Yang Maha Esa, karena atas ridha dan karunianya peneliti dapat menyelesaikan penyusunan laporan penelitian ini. Penulisan ini juga tak lepas dari dukungan pihak yang selalu membantu dalam penulisan penelitian ini. Oleh karenanya peneliti ingin mengucapkan terima kasih kepada:

1. Dr. Zainal Afandi, M.Pd. selaku Rektor Universitas Nusantara PGRI Kediri.
2. Dr. Sulistiono, M.Si. selaku Dekan Fakultas Teknik dan Ilmu Komputer Universitas Nusantara PGRI Kediri.
3. Risa Helilintar, M.Kom selaku Ketua Program Studi Teknik Informatika Universitas Nusantara PGRI Kediri.
4. Ardi Sanjaya, M.Kom. dan Intan Nur Farida, M.Kom selaku Dosen Pembimbing Skripsi yang telah dan mengarahkan kami selama mengerjakan skripsi.
5. Kedua Orang Tua saya dan Keluarga atas doa dan dukungannya.
6. Ucapan terimakasih juga disampaikan kepada pihak-pihak lain yang tidak dapat disebutkan satu persatu, yang telah banyak membantu menyelesaikan penulisan penelitian ini.

Disadari penelitian ini masih banyak kekurangan, maka diharapkan kritik dan saran dari berbagai pihak sangat diharapkan. Semoga penelitian ini bermanfaat bagi semua pihak.

Kediri, 14 Juli 2026

Muhammad Ichsan

NPM 2213020050

DAFTAR ISI

HALAMAN PERNYATAAN	iv
HALAMAN PERSEMBAHAN.....	v
HALAMAN MOTTO	vi
RINGKASAN	vii
PRAKATA.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
A. Latar Belakang	1
B. Identifikasi Masalah.....	3
C. Rumusan Masalah.....	3
D. Batasan Masalah	4
E. Tujuan Penelitian	4
F. Manfaat Penelitian	5
BAB II LANDASAN TEORI	6
A. Teori dan Penelitian Terdahulu.....	6
B. Kerangka Berfikir	18
BAB III METODE PENELITIAN.....	21
A. Desain Penelitian	21
B. Instrumen Penelitian	24
C. Tempat dan Jadwal Penelitian	27
D. Objek Penelitian/Subjek Penelitan.....	29
E. Prosedur Penelitian	35
F. Teknik Analisis Data.....	37
BAB IV Hasil dan Pembahasan.....	54
A. Hasil Penelitian	54
B. Pembahasan.....	63
BAB V PENUTUP.....	65
A. Kesimpulan	65

B. Saran	66
DAFTAR PUSTAKA	67
Lampiran	69

DAFTAR TABEL

Tabel 3. 1 Jadwal Penelitian.....	28
Tabel 3. 2 Dataset kitab Safinatun Najah.....	32
Tabel 3. 3 Data <i>Input</i>	46
Tabel 4.1 Halaman Beranda.....	58
Tabel 4.2 Pengujian Konsultasi Fikih.....	59
Tabel 4.3 Pengujian Informasi	59
Tabel 4.4 Pengujian Halaman Tentang Sistem	60
Tabel 4.5 Skenario Pengujian Retrieval.....	61
Tabel 4.6 Hasil Pengujian Retrieval.....	62
Tabel 4.7 Hasil Evaluasi Retrieval.....	62

DAFTAR GAMBAR

Gambar 2. 1 Kerangka Berfikir.....	18
Gambar 3, 1 Prosedur Penelitian.....	35
Gambar 3, 2 <i>Use Case Diagram</i>	37
Gambar 3, 3 <i>Activity Diagram</i>	38
Gambar 3, 4 <i>Sequence Diagram</i>	39
Gambar 3, 5 <i>Class Diagram</i>	41
Gambar 3, 6 Desain Halaman Utama <i>Chatbot</i>	44
Gambar 3, 7 Desain Input Pertanyaan.....	44
Gambar 3, 8 Desain <i>Output</i>	45
Gambar 4. 1 Halaman Beranda Sistem	54
Gambar 4. 2 Halaman Konsultasi Fikih.....	55
Gambar 4. 3 Halaman Tentang Kitab	56
Gambar 4. 4 Halaman Tentang Sistem	56

BAB I

PENDAHULUAN

A. Latar Belakang

Perkembangan teknologi *Artificial Intelligence (AI)* telah memberikan dampak signifikan terhadap berbagai aspek kehidupan, termasuk dalam bidang pendidikan agama Islam. Salah satu cabang *AI* yang banyak dimanfaatkan adalah *Natural Language Processing (NLP)*, yakni teknologi yang memungkinkan komputer untuk memahami dan memproses bahasa manusia secara alami. Penerapan *NLP* semakin berkembang dalam sistem pembelajaran berbasis digital, termasuk dalam konteks pesantren yang menekankan kajian kitab kuning.

Pondok Pesantren Al-Ma'ruf Kediri merupakan institusi pendidikan Islam yang menyelenggarakan pembelajaran kitab klasik berbahasa Arab, khususnya kitab fikih dasar pada jenjang kelas 1 *Asasiyah*. Pembelajaran fikih di pesantren ini masih dilakukan secara konvensional, dengan mengandalkan penjelasan langsung dari ustadz sebagai sumber utama pemahaman materi. Pola pembelajaran tersebut telah menjadi karakteristik khas pesantren dalam menjaga transmisi keilmuan Islam secara tradisional.

Berdasarkan hasil observasi yang telah dilakukan, santri kelas 1 *Asasiyah* masih menghadapi berbagai kendala dalam proses pembelajaran fikih. Kesulitan utama yang dialami santri adalah memahami teks Arab dalam kitab fikih serta mencari penjelasan hukum secara mandiri. Proses pencarian materi yang masih dilakukan secara manual menyebabkan santri sangat bergantung pada ustadz, sehingga interaksi mandiri santri dengan sumber pengetahuan menjadi terbatas dan efektivitas pembelajaran kurang optimal.

Di sisi lain, perkembangan teknologi kecerdasan buatan (*Artificial Intelligence/AI*), khususnya pada bidang *Natural Language Processing (NLP)*, telah membuka peluang penerapan teknologi dalam sistem pembelajaran berbasis digital. *NLP* memungkinkan sistem komputer untuk memahami dan memproses bahasa manusia secara alami, sehingga berpotensi digunakan

sebagai media pendukung pembelajaran yang bersifat interaktif. Pemanfaatan teknologi ini mulai banyak diterapkan dalam berbagai sistem pembelajaran digital, meskipun penerapannya dalam konteks pesantren masih relatif terbatas.

Beberapa penelitian terdahulu menunjukkan bahwa penerapan *NLP* dalam sistem *chatbot* mampu meningkatkan kualitas layanan informasi berbasis tanya jawab. Penelitian oleh membuktikan bahwa *chatbot* berbasis *NLP* dapat memberikan respons yang lebih akurat dan efisien dalam layanan informasi (Muhammad Hamzani Wahid et al., 2025). Selain itu, pendekatan *Retrieval-Augmented Generation (RAG)* juga dikembangkan untuk menghasilkan jawaban yang relevan dengan mengintegrasikan proses pencarian informasi dan pembuatan respons generatif berbasis sumber data tertentu (Procko & Ochoa, 2024).

Retrieval-Augmented Generation (RAG) merupakan pendekatan yang menggabungkan proses pencarian informasi (*retrieval*) dengan kemampuan pembangkitan jawaban (*generation*) pada *Large Language Model (LLM)*. Proses *retrieval* merupakan tahapan dalam *Retrieval-Augmented Generation (RAG)* yang bertujuan mencari dokumen paling relevan terhadap pertanyaan pengguna. Agar proses *retrieval* dapat menemukan dokumen yang memiliki makna paling relevan dengan pertanyaan pengguna, penelitian ini memanfaatkan teknik *embedding*. *Embedding* merupakan proses mengubah teks menjadi representasi dalam bentuk vektor numerik sehingga hubungan makna antar kata atau kalimat dapat dipahami oleh sistem.

Pendekatan *Retrieval-Augmented Generation (RAG)* dipilih dalam penelitian ini karena mampu mengintegrasikan proses pencarian informasi (*retrieval*) dengan kemampuan generatif *Large Language Model (LLM)* dalam menghasilkan respons. Berbeda dengan *chatbot* generatif yang hanya mengandalkan pengetahuan yang telah dipelajari selama proses pelatihan model, pendekatan *RAG* terlebih dahulu melakukan pencarian terhadap dokumen yang memiliki tingkat relevansi tertinggi dari basis pengetahuan berdasarkan pertanyaan pengguna. Dokumen yang diperoleh kemudian digunakan sebagai konteks utama dalam proses pembangkitan jawaban oleh

LLM. Melalui mekanisme tersebut, sistem diharapkan mampu menghasilkan jawaban yang lebih relevan dengan kebutuhan pengguna serta tetap mengacu pada sumber informasi yang telah ditetapkan. Selain itu, penerapan *RAG* berperan dalam meminimalkan terjadinya *hallucination*, yaitu kondisi ketika *LLM* menghasilkan informasi yang tidak didukung oleh sumber referensi yang valid, sehingga konsistensi dan keandalan jawaban yang dihasilkan dapat lebih terjaga.

Melalui pengembangan sistem ini, diharapkan proses pembelajaran fikih di Pondok Pesantren Al-Ma'ruf Kediri dapat didukung oleh media pembelajaran digital yang membantu santri memahami materi fikih secara mandiri tanpa mengesampingkan otoritas kitab dan pembelajaran pesantren.

B. Identifikasi Masalah

Berdasarkan deskripsi dalam latar belakang, proses pembelajaran fikih di Pondok Pesantren Al-Ma'ruf Kediri masih dihadapkan pada beberapa permasalahan. Santri kelas 1 *Asasiyah* masih mengalami kesulitan dalam memahami teks Arab pada kitab fikih klasik akibat keterbatasan penguasaan bahasa dan pemahaman konteks hukum. Proses pencarian hukum fikih dalam kitab masih dilakukan secara manual, sehingga memerlukan waktu yang relatif lama dan kurang efisien. Metode pembelajaran yang bersifat tradisional menyebabkan santri cenderung pasif dan memiliki ketergantungan tinggi terhadap penjelasan ustadz. Selain itu, hingga saat ini belum tersedia aplikasi berbasis teknologi yang dapat memfasilitasi santri untuk melakukan tanya jawab fikih secara mandiri dengan merujuk pada kitab yang autentik.

C. Rumusan Masalah

Berdasarkan identifikasi masalah tersebut, rumusan masalah dalam penelitian ini adalah bagaimana merancang dan mengimplementasikan *chatbot* fikih berbasis *NLP* yang mampu memahami pertanyaan santri dalam bahasa Indonesia dan memberikan jawaban sesuai konteks pembelajaran fikih dasar. Selain itu, penelitian ini juga merumuskan permasalahan mengenai bagaimana penerapan pendekatan *Retrieval-Augmented Generation (RAG)* dalam menghasilkan jawaban yang relevan berdasarkan dataset kitab *Safinaturun*

Najah, serta sejauh mana tingkat akurasi dan relevansi jawaban *chatbot* dalam merespons pertanyaan santri kelas 1 *Asasiyah* di Pondok Pesantren Al-Ma'ruf Kediri.

D. Batasan Masalah

Agar penelitian ini tetap terfokus dan tidak melampaui batas, penulis menetapkan beberapa pembatasan sebagai ruang lingkup penelitian. Batasan masalah ditetapkan sebagai berikut:

1. Dataset yang digunakan dalam pengembangan sistem bersumber secara eksklusif dari kitab *Safinatun Najah* dalam format digital *csv*.

Teknologi yang digunakan:

- a. Bahasa pemrograman: *Python 3.11*
 - b. *Framework*: *Streamlit* (untuk antarmuka), *LangChain* (untuk *RAG*)
 - c. *Vector database*: *ChromaDB* (lokal)
 - d. *Model embedding*: *intfloat/multilingual-e5-large*
 - e. *Large Language Model*: *Google Gemini 2.5 Flash* (via *API*)
 - f. *Deployment*: *Streamlit*
2. Fitur yang dikembangkan mencakup:
 - a. Pencarian dalil berbasis *RAG*
 - b. Penampilan teks Arab asli kitab
 - c. Antarmuka *web* responsif (bisa dibuka di HP dan laptop)
 3. Penggunaan dan bahasa:
 - a. Sistem hanya menerima pertanyaan menggunakan bahasa Indonesia.
 - b. Pengguna yang dituju adalah santri, pelajar, dan masyarakat umum yang menggunakan bahasa Indonesia sehari-hari.

E. Tujuan Penelitian

Penelitian ini bertujuan untuk membangun dan menganalisis sistem *chatbot* fikih berbasis *NLP* yang dapat mendukung proses pembelajaran santri di Pondok Pesantren Al-Ma'ruf Kediri. Secara spesifik, penelitian ini bertujuan untuk merancang dan membangun aplikasi *chatbot* yang mampu memahami pertanyaan santri dalam bahasa alami serta memberikan jawaban berdasarkan

rujukan kitab *Safinatun Najah*. Selain itu, penelitian ini juga bertujuan menerapkan metode *RAG* untuk menghasilkan jawaban yang lebih akurat, kontekstual, dan selaras dengan isi kitab fikih. Hasil akhir penelitian ini diharapkan dapat dimanfaatkan untuk mengevaluasi tingkat akurasi dan efektivitas *chatbot* sebagai media pembelajaran digital yang mendorong kemandirian belajar santri di lingkungan pesantren.

F. Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat baik secara teoretis maupun praktis. Secara teoretis, penelitian ini memberikan kontribusi ilmiah terhadap pengembangan teknologi *NLP* dan pendekatan *RAG*, khususnya dalam aplikasinya pada bidang pendidikan Islam. Penelitian ini juga dapat menjadi referensi bagi pengembangan aplikasi serupa yang berfokus pada pengolahan teks kitab klasik berbahasa Arab. Secara praktis, penelitian ini diharapkan dapat memfasilitasi santri dalam memahami ilmu fikih dasar melalui pendekatan yang lebih cepat dan interaktif, serta membantu ustadz dalam menyediakan media pembelajaran digital yang efisien. Selain itu, sistem ini diharapkan dapat menjadi langkah awal dalam mewujudkan digitalisasi pendidikan Islam yang adaptif terhadap perkembangan teknologi.

DAFTAR PUSTAKA

- Adamopoulou, E., & Moussiades, L. (2020). An Overview of Chatbot Technologybased chatbots. In *55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)* (Vol. 1). Springer International Publishing. <https://doi.org/10.1007/978-3-030-49186-4>
- Alayba, A. M. (2025). Arabic Natural Language Processing (NLP): A Comprehensive Review of Challenges, Techniques, and Emerging Trends. *Computers*, 14(11). <https://doi.org/10.3390/computers14110497>
- Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 2020-Decem*.
- Dahlan, H. R., Windy, R., Saputra, & Sahid, H. N. (2025). A LITERATURE REVIEW ON USING AI CHATBOT APPLICATION TO ENHANCE STUDENTS PERFORMANCE ON STEAM. *Jurnal Review Pendidikan Dan Pengajaran (Atau Nama Lengkap Sesuai Sumber)*, 8(1), 2950–2956. <https://doi.org/https://doi.org/10.31004/jrpp.v8i1.43173>
- Dam, S. K., Hong, C. S., Qiao, Y., & Zhang, C. (2024). *A Complete Survey on LLM-based AI Chatbots*. 1–23. <http://arxiv.org/abs/2406.16937>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 1(Mlm)*, 4171–4186.
- Elysia, S., & Herianto. (2024). Chatbot Berbasis Retrieval Augmented Generation (RAG) untuk Peningkatan Layanan Informasi Sekolah. *Journal TIFDA (Technology Information and Data Analytic)*, 1(2), 52–58. <https://doi.org/10.70491/tifda.v1i2.52>
- Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Guo, Q., Wang, M., & Wang, H. (2023). *Retrieval-Augmented Generation for Large Language Models: A Survey*. *arXiv:2312.10997*. <https://arxiv.org/pdf/2312.10997>
- 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-Decem*.
- Muhammad Hamzani Wahid, Danang Tejo Kumoro, & Zaenal Abidin. (2025). Implementasi Chatbot Berbasis Natural Language Processing Pada Web Lppmb Di Uniqhba. *SainsTech Innovation Journal*, 7(2), 489–496.

<https://doi.org/10.37824/sij.v7i2.2024.758>

- Procko, T. T., & Ochoa, O. (2024). Graph Retrieval-Augmented Generation for Large Language Models: A Survey. *Proceedings - 2024 Conference on AI, Science, Engineering, and Technology, AIXSET 2024*, 166–169. <https://doi.org/10.1109/AIXSET62544.2024.00030>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*.
- Wang, L., Yang, N., Huang, X., Yang, L., Majumder, R., & Wei, F. (2024). *Multilingual E5 Text Embeddings: A Technical Report*. <http://arxiv.org/abs/2402.05672>
- Weizenbaum, J. (1966). ELIZA-A computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36–45. <https://doi.org/10.1145/365153.365168>
- Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, Z., Yang, R., & Wen, J.-R. (2023). *A Survey of Large Language Models*. <https://doi.org/10.48550/arXiv.2303.18223>