

PENGEMBANGAN CHATBOT DENGAN PENDEKATAN RAG UNTUK MENDUKUNG PENGELOLAAN LIMBAH RADIOAKTIF

Ihsan Aulia Rahman¹, Tati Mardiana², Riki Supriyadi³

Program Studi Sains Data, Fakultas Teknologi Informasi, Universitas Nusa Mandiri

15210014@nusamandiri.ac.id,^{2,3},
Universitas Nusa Mandiri

Abstract

Limited access to comprehensive technical information on Disused Sealed Radioactive Sources (DSRS) poses a challenge in radioactive waste management in Indonesia. The need for a digital system capable of interactively presenting the classification, handling procedures, and reuse of DSRS has driven the development of artificial intelligence-based solutions. This research aims to develop a chatbot system based on a Large Language Model (LLM) equipped with a Retrieval-Augmented Generation (RAG) approach to enhance the chatbot's contextual understanding of technical documents related to DSRS. The system integrates two types of LLM models: a local model via Ollama and an external model based on the OpenAI API, accessed through the OpenWebUI user interface. The RAG architecture is used to extract context from technical documents using semantic vectors before the generative process. Evaluation using four metrics—Cosine Similarity, ROUGE-L, BERTScore F1, and Fuzzy Ratio—shows that the GPT-4 model with Deep Thinking and RAG generates the most accurate and relevant responses to the ground truth, with the highest Cosine Similarity score of 0.9026 and BERTScore F1 of 0.8632. These results indicate that the integration of LLM and RAG significantly improves the quality of chatbot responses in the domain of radioactive waste management. This research contributes in the form of an adaptive information system prototype that can support decision-making processes and knowledge dissemination more effectively in the nuclear field.

Keywords: Radioactive Waste; DSRS; Chatbot; Large Language Model; RAG

Abstrak

Keterbatasan akses terhadap informasi teknis yang komprehensif mengenai Disused Sealed Radioactive Sources (DSRS) menjadi tantangan dalam pengelolaan limbah radioaktif di Indonesia. Kebutuhan akan sistem digital yang mampu menyajikan klasifikasi, prosedur penanganan, dan pemanfaatan ulang DSRS secara interaktif mendorong pengembangan solusi berbasis kecerdasan buatan. Penelitian ini bertujuan untuk mengembangkan sistem chatbot berbasis Large Language Model (LLM) yang dilengkapi dengan pendekatan Retrieval-Augmented Generation (RAG) guna meningkatkan pemahaman kontekstual chatbot terhadap dokumen teknis terkait DSRS. Sistem mengintegrasikan dua jenis model LLM, yaitu model lokal melalui Ollama dan model eksternal berbasis API OpenAI, yang diakses melalui antarmuka pengguna Open WebUI. Arsitektur RAG digunakan untuk mengekstraksi konteks dari dokumen teknis menggunakan vektor semantik sebelum dilakukan proses generatif. Evaluasi menggunakan empat metrik, yaitu Cosine Similarity, ROUGE-L, BERTScore F1, dan Fuzzy Ratio menunjukkan bahwa model GPT-4 - Deep Thinking - With RAG menghasilkan respons paling akurat dan relevan terhadap ground truth, dengan skor Cosine Similarity tertinggi sebesar 0.9026 dan BERTScore F1 sebesar 0.8632. Hasil ini mengindikasikan bahwa integrasi LLM dan RAG secara signifikan meningkatkan kualitas jawaban chatbot dalam domain pengelolaan limbah radioaktif. Penelitian ini memberikan kontribusi dalam bentuk prototipe sistem informasi yang adaptif dan dapat menunjang proses pengambilan keputusan serta diseminasi pengetahuan secara lebih efektif di bidang nuklir.

Kata kunci: Limbah Radioaktif; DSRS; Chatbot; Large Language Model; RAG

PENDAHULUAN

Pengelolaan *Disused Sealed Radioactive Sources* (DSRS) merupakan isu strategis dalam keselamatan nuklir di Indonesia. Dengan lebih dari 10.000 sumber radioaktif tersegregasi terdaftar hingga

2022, banyak di antaranya telah habis masa pakainya namun masih memiliki aktivitas radioaktif signifikan. Keterbatasan akses terhadap informasi teknis yang cepat dan akurat menjadi hambatan utama dalam pengelolaan yang aman.

Perkembangan *Large Language Models* (LLM) menawarkan peluang baru untuk memfasilitasi akses informasi teknis melalui *chatbot*. Namun, LLM standar sering mengalami *hallucination* atau memberikan jawaban yang tidak akurat ketika dihadapkan pada domain teknis spesifik seperti nuklir. Untuk mengatasi ini, pendekatan *Retrieval-Augmented Generation* (RAG) diperkenalkan. RAG menggabungkan proses pencarian dokumen (retrieval) dengan pembuatan jawaban (generation), sehingga LLM dapat mengakses dokumen eksternal sebagai landasan jawaban yang lebih akurat dan kontekstual.

Hingga saat ini, belum ada penelitian yang khusus mengembangkan *chatbot* berbasis RAG untuk pengelolaan DSRS di Indonesia. Selain itu, implementasi yang menggabungkan model lokal (via Ollama) dan model *cloud* (via OpenAI) di sektor strategis ini juga belum banyak diteliti. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi *chatbot* cerdas berbasis LLM dengan pendekatan RAG yang mampu memberikan jawaban teknis akurat terkait pengelolaan DSRS, serta membandingkan performanya dengan LLM standar.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini menggunakan pendekatan *Research and Development* (R&D) dengan model pengembangan sistem *Rapid Prototyping*. Pendekatan ini bersifat kuantitatif, melalui pengukuran performa model menggunakan metrik evaluasi, dan kualitatif, melalui analisis komparatif terhadap relevansi dan kualitas jawaban yang dihasilkan *chatbot*.

Waktu dan Tempat Penelitian

Objek penelitian adalah sistem pengelolaan DSRS di Badan Riset dan Inovasi Nasional (BRIN). Pengumpulan data dan observasi dilakukan di lingkungan Pusat Riset Teknik Bahan Nuklir dan Limbah Radioaktif (PRTBNLR) BRIN.

Prosedur Penelitian

Tahapan penelitian meliputi:

1. Studi Literatur dan Analisis Kebutuhan: Melakukan telaah dokumen teknis (IAEA, BAPETEN) serta observasi dan wawancara semi-terstruktur dengan peneliti dan praktisi di PRTBNLR BRIN untuk memahami alur kerja dan kebutuhan informasi.

2. Perancangan Arsitektur Sistem: Merancang arsitektur teknologi yang mencakup pemilihan LLM (lokal dan *cloud*), *pipeline* RAG, *vector store*, dan antarmuka pengguna.
3. Pembuatan dan Integrasi Sistem: Dokumen teknis (PDF, Word, dll.) diolah menjadi *embedding* menggunakan model sentence-transformers/all-MiniLM-L6-v2 dan *nommic-embed*. Hasil *embedding* disimpan dalam *vector store* ChromaDB.
4. Integrasi Antarmuka: Sistem *chatbot* diintegrasikan dengan antarmuka OpenWebUI, yang memungkinkan pengguna berinteraksi dan memilih model LLM.
5. Pengujian dan Evaluasi: Melakukan pengujian fungsional dan evaluasi kualitas respons menggunakan skenario pertanyaan spesifik DSRS.

Data, Instrumen, dan Teknik Analisis Data

Sumber data utama adalah dokumen teknis publik seperti pedoman IAEA, regulasi BAPETEN, dan laporan riset internal BRIN.

Instrumen yang digunakan adalah sistem *chatbot* DSRSafe-Bot yang mengintegrasikan empat model LLM:

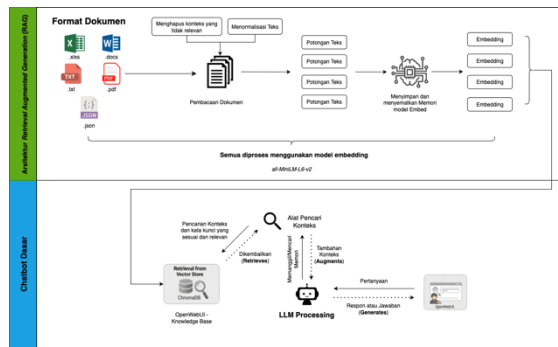
- **Cloud (OpenAI):** GPT-4 ("01" - *Deep Thinking*) dan GPT-3.5 Turbo ("03-mini" - *Fast Thinking*).
- **Lokal (Ollama):** Deepseek R1 14B (*Deep Thinking*) dan LLaMA 3.1 8B (*Fast Thinking*).

Teknik analisis data berfokus pada perbandingan performa model dalam dua kondisi: tanpa RAG (*baseline*) dan dengan RAG. Kualitas jawaban dievaluasi secara kuantitatif menggunakan empat metrik: *Cosine Similarity*, *ROUGE-L*, *BERTScore F1*, dan *Fuzzy Ratio* terhadap *ground truth*.

HASIL DAN PEMBAHASAN

Hasil Sistem Chatbot DSRSafe-Bot, sebuah *chatbot* cerdas yang dirancang untuk domain DSRS. Arsitektur sistem (terlihat pada Gambar 1) memproses masukan pengguna, mengubahnya menjadi *embedding*, lalu melakukan pencarian semantik di *vector store* (ChromaDB) untuk menemukan konteks dokumen yang relevan. Konteks ini kemudian digabungkan (*augmented*)

dengan pertanyaan asli dan dikirim ke model LLM (lokal atau *cloud*) untuk menghasilkan jawaban.



Gambar 1. Arsitektur Chatbot dengan RAG

Analisis Kualitas Jawaban (RAG vs Non-RAG) Pengujian dilakukan dengan memberikan pertanyaan "apa itu DSRS?".

1. **Model Tanpa RAG (Baseline):** Model lokal (LLaMA3.1 8B dan DeepSeek R1 14B) mengalami *hallucination*. Keduanya salah menafsirkan DSRS sebagai "*Daily Step Rate Scale*", yaitu metrik kebugaran untuk menghitung langkah harian. Jawaban ini, meskipun terstruktur, sepenuhnya tidak relevan dengan konteks nuklir. Model *cloud* (GPT) memberikan jawaban umum tanpa referensi teknis spesifik.
2. **Model Dengan RAG:** Setelah RAG diaktifkan, semua model (lokal dan *cloud*) mampu memberikan jawaban yang akurat. Sistem dengan benar mengidentifikasi DSRS sebagai "*Disused Sealed Radioactive Sources*" (Sumber Radioaktif Tersegel Tidak Digunakan Lagi). Model juga mampu menjelaskan konteksnya terkait bahan bakar nuklir, radioisotop, dan perlunya penanganan khusus oleh IAEA.

Evaluasi Kinerja Kuantitatif metrik menunjukkan peningkatan performa yang konsisten dan signifikan pada semua model yang menggunakan RAG.

Tabel 1. Perbandingan Relevansi Model Tanpa Dan Dengan RAG

Model	Cosine Similarity	ROUGE-L	BERTScore F1	Fuzzy Ratio
LLaMA 3.1	0.2981	0.1176	0.8218	0.39

LLaMA 3.1 + RAG	0.8853	0.2752	0.8598	0.52
-----------------	--------	--------	--------	------

DeepSeek	0.3766	0.1207	0.8275	0.41
----------	--------	--------	--------	------

DeepSeek + RAG	0.8604	0.2222	0.8515	0.46
----------------	--------	--------	--------	------

GPT o3 mini	0.3150	0.1379	0.8276	0.46
-------------	--------	--------	--------	------

GPT o3 mini + RAG	0.8575	0.2857	0.8586	0.49
-------------------	--------	--------	--------	------

GPT o1	0.3639	0.0882	0.8290	0.37
--------	--------	--------	--------	------

GPT o1 + RAG	0.9026	0.2796	0.8632	0.53
--------------	--------	--------	--------	------

Analisis Performa Teknis Model Lokal Performa teknis model lokal juga dievaluasi untuk membandingkan efisiensi pemrosesan.

Tabel 2. Perbandingan Kecepatan Respon Model Tanpa Dan Dengan RAG

Model	RAG ?	Total Duration	Response Tokens/s	Load Duration	Completion Tokens	Total Tokens
LLaMA3.1-8B	Tidak	7.30s	117.92 token/de tik	5.18s	219	234
LLaMA3.1-8B	Ya	3.73s	97.16 token/s	0.79s	134	4230

Dee pSe ek R1- 14B	Tida k	12.5 8 s	64.7 7 toke n/s	7.07 s	345	579
Dee pSe ek R1- 14B	Ya	11.9 6 s	58.5 0 toke n/s	0.73 s	362	4458

Hasil pada Tabel 2 menunjukkan bahwa aktivasi RAG meningkatkan Total Tokens secara signifikan karena penambahan konteks dari dokumen. Menariknya, Total Duration (waktu respons total) untuk model LLaMA3.1 8B justru menjadi lebih cepat saat menggunakan RAG (dari 7.30s menjadi 3.73s), dan Load Duration (waktu muat model) turun drastis untuk kedua model. Hal ini mengindikasikan bahwa meskipun beban token lebih berat, *pipeline* RAG yang terintegrasi dapat berjalan sangat efisien.

Keterbatasan Penelitian ini memiliki beberapa keterbatasan. Pengujian dilakukan pada set pertanyaan spesifik dan belum mencakup ratusan variasi. Selain itu, *chatbot* terkadang masih menghasilkan jawaban di luar domain DSRS, menunjukkan perlunya pembatasan konteks yang lebih ketat.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian, dapat ditarik kesimpulan sebagai berikut:

1. Sistem *chatbot* DSRSafe-Bot telah berhasil dirancang dengan mengintegrasikan LLM dan RAG, yang mampu memahami konteks teknis pengelolaan DSRS, termasuk data dari IAEA dan simulasi peluruhan.
2. Penerapan RAG secara signifikan meningkatkan kualitas, akurasi, dan relevansi jawaban. Model *GPT 01 (GPT-4) Deep Thinking With RAG* menunjukkan performa terbaik dengan skor *Cosine Similarity* 0.9026 dan *BERTScore F1* 0.8632.
3. Integrasi model lokal (Ollama) dan *cloud* (OpenAI) dalam satu antarmuka OpenWebUI berhasil dilakukan,

memberikan fleksibilitas penggunaan sistem.

4. Meskipun berhasil, masih ditemukan indikasi *hallucination* di luar konteks DSRS, yang mengindikasikan perlunya *fine-tuning* lebih lanjut.

Saran

Untuk pengembangan sistem di masa mendatang, diajukan beberapa saran:

1. Pelibatan Ahli Nuklir: Melibatkan ahli nuklir secara formal dalam proses validasi teknis untuk memverifikasi keakuratan jawaban *chatbot*.
2. *Fine-Tuning* Model: Melakukan *fine-tuning* LLM menggunakan korpus dokumen DSRS yang spesifik (regulasi IAEA, laporan teknis BRIN) untuk mengurangi potensi *hallucination*.
3. Integrasi RLHF: Menerapkan *Reinforcement Learning from Human Feedback* (RLHF) agar *chatbot* dapat belajar dari umpan balik pengguna (pakar) secara langsung.
4. Pelatihan Berkala: Melatih ulang model secara periodik untuk menjaga stabilitas dan adaptasi terhadap data atau regulasi baru.
5. Pengembangan Fitur: Menambahkan fitur evaluasi otomatis dan memungkinkan pengguna memperkaya *knowledge base* dengan dokumen baru.

DAFTAR PUSTAKA

- BAPETEN. (2022). *Data jumlah sumber radioaktif tersegel di Indonesia*.
- Berger, P., & von Garrel, J. (2023). How to design a value-based Chatbot for the manufacturing industry: An empirical study of an internal assistance for employees. *KI - Kunstliche Intelligenz*, 37(2-4), 203-211. <https://doi.org/10.1007/s13218-023-00817-6>
- Bommasani, R., et al. (2022). *On the Opportunities and Risks of Foundation Models*. arXiv. <http://arxiv.org/abs/2108.07258>
- Brown, T. B., et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 2020-Decem*.
- Casadei, A., Schlogl, S., & Bergmann, M. (2022). Chatbots for Robotic Process Automation: Investigating Perceived Trust and User

- Satisfaction. *Proceedings of the 2022 IEEE International Conference on Human-Machine Systems, ICHMS 2022*. <https://doi.org/10.1109/ICHMS56717.2022.9980826>
- Chau, H. K. L., Ngo, T. T. A., Bui, C. T., & Tran, N. P. N. (2025). Human-AI interaction in E-Commerce: The impact of AI-powered customer service on user experience and decision-making. *Computers in Human Behavior Reports*, 19, 100725. <https://doi.org/10.1016/j.chbr.2025.100725>
- Cheng, X., Bao, Y., Zarifis, A., Gong, W., & Mou, J. (2022). Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure. *Internet Research*, 32(2), 496–517. <https://doi.org/10.1108/INTR-08-2020-0460>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171–4186.
- Ding, Y., & Najaf, M. (2024). Interactivity, humanness, and trust: a psychological approach to AI chatbot adoption in e-commerce. *BMC Psychology*, 12(1). <https://doi.org/10.1186/s40359-024-02083-z>
- Gao, K. Y., Xiong, Y., Gao, L., & Liu, X. (2024). *Retrieval-Augmented Generation for Large Language Models*.
- Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing*.
- Grattafiori, A., et al. (2024). *The Llama 3 Herd of Models*. arXiv. <http://arxiv.org/abs/2407.21783>
- Ha, J., Kambe, M., & Pe, J. (2011). *Data Mining: Concepts and Techniques*. Elsevier. <https://doi.org/10.1016/C2009-0-61819-5>
- Iaea. (2006). *GC(51)/5 - IAEA Annual Report 2006*.
- IAEA. (2018). IAEA ANNUAL REPORT 2018. *International Atomic Energy Agency*, 63(5), 27–198.
- International Atomic Energy Agency. (2022). *Artificial Intelligence for Accelerating Nuclear Applications, Science and Technology*. <https://www.iaea.org/publications/15198/artificial-intelligence-for-accelerating-nuclear-applications-science-and-technology>
- International Atomic Energy Agency. (2022). *Management of Naturally Occurring Radioactive Material (NORM) in Industry*. https://www-pub.iaea.org/MTCD/Publications/PDF/PU_B1998_web.pdf
- Izcard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open domain question answering. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2021)*, 874–880. <https://doi.org/10.18653/v1/2021.eacl-main.74>
- Jurafaky. (n.d.). *Chatbots and Dialogue Systems Introduction to Chatbots and Dialogue Systems Conversational Agents*.
- Jurafsky, J. H., & Martin, D. (2023). *Speech and Language Processing*.
- Karkosz, S., Szymański, R., Sanna, K., & Michałowski, J. (2024). Effectiveness of a Web-based and Mobile Therapy Chatbot on Anxiety and Depressive Symptoms in Subclinical Young Adults: Randomized Controlled Trial. *JMIR Formative Research*, 8. <https://doi.org/10.2196/47960>
- Karpukhin, V., et al. (2020). Dense passage retrieval for open-domain question answering. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6769–6781. <https://doi.org/10.18653/v1/2020.emnlp-main.550>
- Kelly, A., Noctor, E., Ryan, L., & van de Ven, P. (2025). The Effectiveness of a Custom AI Chatbot for Type 2 Diabetes Mellitus Health Literacy: Development and Evaluation Study. *Journal of Medical Internet Research*, 27, 1–20. <https://doi.org/10.2196/70131>
- Khan, A. A., Hasan, M. T., Kemell, K. K., Rasku, J., & Abrahamsson, P. (2024). *Developing Retrieval Augmented Generation (RAG) based LLM Systems from PDFs: An Experience Report*. arXiv. <http://arxiv.org/abs/2410.15944>
- Laloy, E., Rogiers, B., Bielen, A., & Boden, S. (2021). Bayesian inference of 1D activity profiles from segmented gamma scanning of a heterogeneous radioactive waste drum. *Applied Radiation and Isotopes*, 175,

- 109803.<https://doi.org/10.1016/j.apradiso.2021.109803>
- Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems, 2020-Decem.*
- Lin, C. Y. (2004, July). Rouge: A package for automatic evaluation of summaries. *Proceedings of the workshop on text summarization branches out (WAS 2004)* (pp. 25–26). Association for Computational Linguistics.
- Liu, F., Kang, Z., & Han, X. (2025). Optimizing RAG Techniques for Automotive Industry PDF Chatbots: A Case Study with Locally Deployed Ollama Models. *Proceedings of 2024 3rd International Conference on Artificial Intelligence and Intelligent Information Processing, AIIP 2024*, 152–159.
<https://doi.org/10.1145/3707292.3707358>
- Ltifi, M. (2023). Trust in the chatbot: a semi-human relationship. *Future Business Journal*, 9(1).
<https://doi.org/10.1186/s43093-023-00288-z>
- Mohiuddin, K., et al. (2023). Attention Is All You Need. *International Conference on Information and Knowledge Management, Proceedings*, 4752–4758.
<https://doi.org/10.1145/3583780.3615497>
- Nilsson, R., & Rosman, O. (2025). *Design and Evaluation of a RAG Chatbot in an Industrial Setting*.
- Ojovan, M. I., & Lee, W. E. (2014). *An Introduction to Nuclear Waste Immobilisation*.
- OpenAI, et al. (2024). *GPT-4 Technical Report*. arXiv.
<http://arxiv.org/abs/2303.08774>
- Openwebui. (n.d.). *OpenWebUI Docs*.
- Orden-Mejía, M., Carvache-Franco, M., Huertas, A., Carvache-Franco, O., & Carvache-Franco, W. (2025). Analysing how AI-powered chatbots influence destination decisions. *PLoS One*, 20(3), 1–20.
<https://doi.org/10.1371/journal.pone.0319463>
- Panigrahi, R. R., et al. (2023). AI Chatbot Adoption in SMEs for Sustainable Manufacturing Supply Chain Performance. *Sustainability (Switzerland)*, 15(18).
- Potts, C., et al. (2023). A Multilingual Digital Mental Health and Well-Being Chatbot (ChatPal): Pre-Post Multicenter Intervention Study. *Journal of Medical Internet Research*, 25.
<https://doi.org/10.2196/43051>
- Rahman, I. A., et al. (2025). Integration of machine learning models for enhancing radioactive waste management of disused sealed radioactive sources. *Nuclear Engineering and Design*, 442, 114272.
<https://doi.org/10.1016/j.nucengdes.2025.114272>
- Rao, P. J., Rao, K. N., & Gokuruboyina, S. (2022). An Experimental Study with Fuzzy-Wuzzy (Partial Ratio) for Identifying the Similarity between English and French Languages for Plagiarism Detection. *International Journal of Advanced Computer Science and Applications*, 13(10), 393–401.
<https://doi.org/10.14569/IJACSA.2022.0131047>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using siamese BERT-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3982–3992.
<https://doi.org/10.18653/v1/d19-1410>
- Shiri, A. (2004). Introduction to Modern Information Retrieval (2nd edition). *Library Review*, 53(9), 462–463.
<https://doi.org/10.1108/00242530410565256>
- Syaiful Bachri Mustamin, S., & Atnang, M. (2024). *Data Science untuk Pemula: Menaklukkan Tantangan dan Peluang di Era Big Data*. CV. Science Tech Group.
- Thirunavukarasu, A. J., Jassar, S., Fote, G., & Ghassemi, M. (2024). *Large Language Models*.
- Wijaya, D. R. (2024). *Dasar Data Science: Dari Teori Ke Praktik*. Deepublish.
- Wisnubroto, D. S., Zamroni, H., Sumarbagiono, R., & Nurliati, G. (2021). Challenges of implementing the policy and strategy for management of radioactive waste and nuclear spent fuel in Indonesia. *Nuclear Engineering and Technology*, 53(2), 549–561.
<https://doi.org/10.1016/j.net.2020.07.005>
- Zhang, X., Chen, A. L., Piao, X., Yu, M., Zhang, Y., & Zhang, L. (2024). Is AI chatbot recommendation convincing customer? An analytical response based on the elaboration likelihood model. *Acta Psychologica*, 250.
<https://doi.org/10.1016/j.actpsy.2024.104501>

Zhang, Y., et al. (2023). *Siren's Song in the AI Ocean: A Survey on Hallucination in Large Language Models*. arXiv.
<http://arxiv.org/abs/2309.01219>