UNIVERSITAS
GADJAH MADA
Pengembangan dan Evaluasi Platform Pembelajaran Anatomi Otak Berbasis Web dengan Retrieval-Augmented Generation dan Fine-Tuning Large Language Model
Galuh Ayu Niken Adi Untari, Ir. Adhistya Erna Permanasari, S.T., M.T., Ph.D., IPM., ASEAN Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

INTISARI

Pembelajaran anatomi otak memerlukan informasi dalam bentuk teks dan visual sesuai dengan kebutuhan pengguna. Informasi yang digunakan perlu relevan dan bersumber dari pengetahuan yang telah dikurasi. Penggunaan *Large Language Model* (LLM) secara langsung memiliki keterbatasan respons yang bergantung pada pengetahuan internal model dan belum menggunakan sumber pengetahuan anatomi otak yang terkurasi. Penelitian ini mengembangkan platform pembelajaran anatomi otak berbasis web dengan *chatbot* LLM. Sistem menggunakan *rule-based routing* untuk mengklasifikasikan pertanyaan ke dalam tiga kategori, yaitu teks, gambar, atau model tiga dimensi. Respons teks dihasilkan melalui *Retrieval-Augmented Generation* (RAG) yang mengambil informasi relevan dari *knowledge base*, sedangkan gambar dan model 3D diperoleh melalui *asset retrieval*. Evaluasi kualitas jawaban dilakukan dengan membandingkan lima metode, yaitu *Prompt*, RAG, *Fine-Tuning*, kombinasi *Prompt* dan RAG, serta kombinasi RAG dan *Fine-Tuning* dengan menggunakan BERTscore dan *Semantic Answer Similarity* (SAS). Hasil evaluasi menunjukkan bahwa kombinasi RAG dan *Fine-Tuning* memperoleh nilai tertinggi, dengan BERTscore *Precision* sebesar $0,887$, *Recall* sebesar $0,881$, F1 sebesar $0,884$, dan SAS sebesar $0,706$. Validasi ahli dilakukan oleh satu dokter memperoleh nilai *mean* $4,56$ dari 5. Sementara itu, evaluasi *usability* terhadap 30 responden menunjukkan bahwa seluruh delapan subkarakteristik memperoleh indeks dalam kategori sangat baik dengan nilai tertinggi pada *Inclusivity* sebesar 88,10% dan terendah pada *User Assistance* sebesar 82,00%. Hasil penelitian menunjukkan bahwa sistem memperoleh hasil yang baik berdasarkan evaluasi kesamaan semantik jawaban, penilaian ahli, dan persepsi pengguna terhadap *usability*. Pengembangan selanjutnya dapat mempertimbangkan validasi ahli, peningkatan cakupan *knowledge base*, dan evaluasi faktualitas jawaban.

Kata kunci : Anatomi Otak, *Large Language Model*, *Retrieval-Augmented Generation*, Fine-Tuning, Usability

UNIVERSITAS
GADJAH MADA
**Pengembangan dan Evaluasi Platform Pembelajaran Anatomi Otak Berbasis Web dengan Retrieval-Augmented
Generation dan Fine-Tuning Large Language Model**
Galuh Ayu Niken Adi Untari, Ir. Adhistya Erna Permanasari, S.T., M.T., Ph.D., IPM., ASEAN Eng.
Universitas Gadjah Mada, 2026 | Diunduh dari http://etd.repository.ugm.ac.id/

ABSTRACT

Brain anatomy learning requires textual and visual information tailored to users' needs. The information provided must be relevant and derived from curated knowledge sources. Direct use of a Large Language Model (LLM) has limitations, as its responses rely on the model's internal knowledge and do not incorporate curated brain anatomy knowledge sources. This study develops a web-based brain anatomy learning platform featuring an LLM-based chatbot. The system employs rule-based routing to classify user queries into three categories: text, images, and three-dimensional (3D) models. Textual responses are generated using Retrieval-Augmented Generation (RAG), which retrieves relevant information from a knowledge base, while images and 3D models are obtained through asset retrieval. Response quality is evaluated by comparing five methods: Prompting, RAG, Fine-Tuning, a combination of Prompting and RAG, and a combination of RAG and Fine-Tuning, using BERTScore and Semantic Answer Similarity (SAS). The evaluation results show that the RAG and Fine-Tuning configuration achieves the highest scores, with a BERTscore precision of 0.887*, recall of* 0.881*, an F1 score of* 0.884*, and an SAS score of* 0.706*. Expert validation conducted by one physician yields a mean score of* 4.56 *out of* 5*. Meanwhile, a usability evaluation involving 30 respondents shows that all eight subcharacteristics fall within the very good category with the highest index observed for Inclusivity at 88.10% and the lowest for User Assistance at 82.00%. These findings indicate that the system achieves favorable results in terms of semantic response similarity, expert assessment, and user perceptions of usability. Future work may focus on expanding expert validation, increasing knowledge base coverage, and evaluating response factuality.*

Keywords : *Brain Anatomy, Large Language Model, Retrieval-Augmented Generation, Fine-Tuning, Usability*