

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Untuk memposisikan penelitian ini agar sejalan dengan konteks perkembangan ilmu pengetahuan terkini, dilakukan kajian terhadap beberapa penelitian sebelumnya yang berfokus pada aplikasi AI dalam bidang hukum.

1. Yao et al. (2025), "Intelligent Legal Assistant: An Interactive Clarification System for Legal Question Answering"

Penelitian ini mengembangkan asisten hukum cerdas yang menggunakan pendekatan klarifikasi interaktif. Sebelum memberikan jawaban, sistem akan mengajukan pertanyaan lanjutan untuk memastikan pemahaman yang mendalam terhadap kebutuhan pengguna. Hasilnya menunjukkan peningkatan signifikan pada akurasi dan kepuasan pengguna (skor 4.8/5.0), bahkan melampaui GPT-4o dalam preferensi pengguna untuk konsultasi hukum. Relevansinya dengan penelitian ini terletak pada pentingnya pemahaman konteks pengguna untuk meningkatkan kualitas jawaban AI.

2. Panchal et al. (2025), "LawPal: A Retrieval-Augmented Generation Based System for Enhanced Legal Accessibility in India"

LawPal adalah sistem tanya jawab hukum yang menggunakan arsitektur Retrieval-Augmented Generation (RAG) untuk menyediakan jawaban yang relevan secara kontekstual dengan hukum di India. Sistem ini berhasil mencapai akurasi lebih dari 90% dalam penarikan informasi dan generasi jawaban. Penelitian ini menjadi acuan penting bagi OSOORA dalam hal penerapan teknik RAG untuk memastikan jawaban yang dihasilkan tidak hanya generatif tetapi juga berlandaskan pada basis pengetahuan hukum yang valid.

3. Graham et al. (2025), "Chatbots in Customer Service within Banking and Finance"

Meskipun berfokus pada sektor perbankan, penelitian ini memberikan wawasan tentang efektivitas *chatbot* dalam meningkatkan efisiensi layanan pelanggan. Namun, studi ini juga menekankan bahwa intervensi manusia masih krusial untuk menangani kasus-kasus kompleks. Hal ini menjadi pengingat bahwa OSOORA, sebagai *chatbot*, berfungsi sebagai alat bantu informasi awal dan bukan pengganti sepenuhnya dari konsultasi hukum profesional.

Tabel 2.1 Perbandingan Penelitian Terdahulu

Judul Penelitian	Fokus Utama	Kekuatan	Relevansi dengan OSOORA
Intelligent Legal Assistant (Yao et al., 2025)	Klarifikasi interaktif untuk meningkatkan relevansi jawaban.	Tingkat kepuasan dan akurasi pengguna sangat tinggi.	Menginspirasi desain interaksi yang berpusat pada pengguna untuk memahami kebutuhan secara spesifik.
LawPal (Panchal et al., 2025)	Penerapan RAG untuk aksesibilitas hukum di India.	Akurasi tinggi (>90%) dalam memberikan jawaban berbasis data hukum.	Menjadi referensi utama dalam implementasi arsitektur RAG untuk OSOORA.
Chatbots in Customer Service	Efektivitas <i>chatbot</i> dalam	Menunjukkan efisiensi <i>chatbot</i> , namun	Memperkuat posisi OSOORA sebagai alat bantu informasi, bukan pengganti advokat.

Judul Penelitian	Fokus Utama	Kekuatan	Relevansi dengan OSOORA
(Graham et al., 2025)	layanan pelanggan.	menyoroti perlunya intervensi manusia.	

Penelitian ini dibangun berdasarkan kualitas berbagai studi sebelumnya sambil mempertahankan penekanan dan kontribusi yang unik. Berbeda dengan “LawPal,” yang berfokus pada sistem hukum India, “OSOORA” secara eksplisit dirancang untuk mengatasi tantangan hukum perdata dalam kerangka sistem hukum dan birokrasi Indonesia yang unik, dengan memasukkan studi kasus mengenai administrasi kartu identitas dan pengelolaan tanah yang erat kaitannya dengan masyarakat. Penelitian ini memprioritaskan penggunaan arsitektur RAG secara komprehensif untuk memberikan respons yang faktual dan berbasis data, berbeda dengan “*Intelligent Legal Assistant*” yang berfokus pada teknik klarifikasi interaktif. Sesuai dengan kesimpulan Graham dkk., pengembangan OSOORA dilakukan dengan pemahaman yang jelas bahwa sistem ini berfungsi sebagai sumber informasi awal (pertolongan pertama hukum), bukan pengganti pengacara, untuk memberdayakan non-profesional sebelum mereka terlibat lebih lanjut dalam sistem hukum formal.

2.2 Teori Terkait

Bagian ini menjelaskan konsep-konsep teknis dan metodologis yang menjadi pilar dalam perancangan dan implementasi OSOORA.

2.2.1 Teknologi Pengembangan Web

HTML, CSS, dan JavaScript: Merupakan tiga pilar utama pengembangan antarmuka web. HTML (*HyperText Markup Language*) digunakan untuk menstrukturkan konten, CSS (*Cascading Style Sheets*) untuk mendesain

tampilan visual, dan JavaScript untuk menciptakan fungsionalitas interaktif di sisi klien (Flanagan, 2020).

Node.js dan Express.js: Node.js adalah sebuah *runtime environment* yang memungkinkan eksekusi JavaScript di sisi server. Dipadukan dengan Express.js, sebuah kerangka kerja minimalis, pengembang dapat membangun API (*Application Programming Interface*) yang tangguh dan efisien untuk menangani logika bisnis aplikasi (Tilkov & Vinoski, 2010).

MongoDB Atlas: Sebuah layanan basis data NoSQL berbasis *cloud* yang menyimpan data dalam format dokumen mirip JSON (BSON). Fleksibilitas skemanya sangat cocok untuk menyimpan data percakapan yang dinamis dan data pengguna (Chodorow, 2013).

2.2.2 Generative Artificial Intelligence (GenAI)

GenAI merepresentasikan cabang dari kecerdasan buatan yang berfokus pada pembuatan konten baru, seperti teks, gambar, atau kode, berdasarkan data yang telah dipelajarinya. Large Language Models (LLM) adalah salah satu bentuk GenAI yang dilatih pada *dataset* teks masif untuk memahami dan menghasilkan bahasa manusia. Dalam penelitian ini, model Qwen 2.5 72B Instruct dari Alibaba Cloud Model Studio digunakan sebagai mesin utama untuk memproses pertanyaan dan menghasilkan jawaban hukum.

2.2.3 Arsitektur Sistem

Diagram Konteks: Menggambarkan sistem secara keseluruhan dan interaksinya dengan entitas eksternal (pengguna, sistem AI) tanpa memerinci komponen internal (Yourdon, 1989).

Use Case Diagram: Memodelkan fungsionalitas sistem dari perspektif pengguna, menunjukkan interaksi antara aktor (pengguna) dan kasus penggunaan (misalnya, mengirim pesan, melihat riwayat) (Booch et al., 2005).

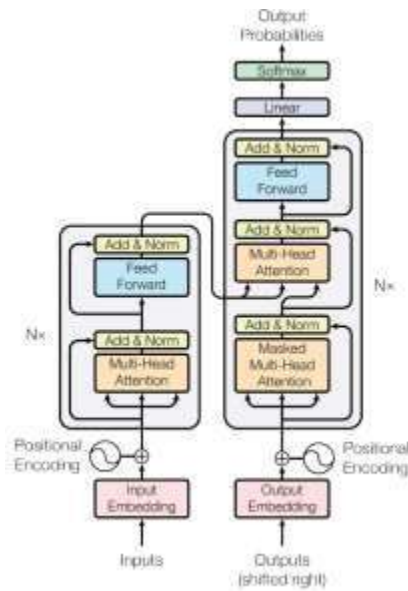
Deployment Diagram: Memvisualisasikan arsitektur fisik sistem, menunjukkan bagaimana komponen perangkat lunak (misalnya, *frontend*, *backend*, basis data) didistribusikan ke perangkat keras (misalnya, server client, server aplikasi, server basis data) (Fowler, 2004).

2.2.4 Metodologi *Adaptive Waterfall*

Model *Adaptive Waterfall* adalah modifikasi dari model *Waterfall* klasik. Model ini mempertahankan alur kerja sekuensial (analisis → desain → implementasi → pengujian), namun memperbolehkan adanya *feedback loop* antar tahapan. Fleksibilitas ini memungkinkan adanya penyesuaian terhadap kebutuhan atau temuan baru selama proses pengembangan sehingga lebih cocok untuk proyek-proyek inovatif seperti pengembangan aplikasi AI (Sommerville, 2011).

2.2.5 Arsitektur *Transformer*

Arsitektur *transformer* merupakan kerangka kerja revolusioner dalam pemrosesan bahasa alami (NLP) yang pertama kali diperkenalkan dalam artikel pionir “Attention Is All You Need” (Vaswani et al, 2017). Berbeda dari model pendahulu seperti Jaringan Neural Berulang (RNN) atau Memori Jangka Pendek Panjang (LSTM), yang memproses data secara berurutan, Transformer memungkinkan pemrosesan simultan dari seluruh data input. Terobosan fundamental ini secara signifikan meningkatkan efisiensi pelatihan dan memungkinkan model untuk memahami hubungan kontekstual jarak jauh dalam teks. Arsitektur ini menjadi dasar hampir semua model bahasa skala besar (LLM) kontemporer, terutama model Qwen 2.5 72B Instruct yang menjadi inti sistem OSOORA.



Gambar 2.1 Model Arsitektur Transformer

Efektivitas *Transformer* secara fundamental bergantung pada sejumlah komponen esensial. Pertama, Embedding dan Positional Encoding mengubah teks input menjadi vektor numerik sambil menyertakan informasi posisi, memungkinkan model memahami urutan kata. Mekanisme self-attention adalah elemen paling kritis, memungkinkan model menilai pentingnya setiap kata dalam kalimat saat memproses kata tertentu. Pada dasarnya, perhatian memungkinkan model memahami konteks secara dinamis dengan bertanya, “Untuk memahami kata ini, kata mana dalam kalimat yang memerlukan fokus terbesar?” Hasil lapisan perhatian ini diproses lebih lanjut oleh lapisan Jaringan Feed-Forward untuk meningkatkan kedalaman komputasi sebelum dikirim ke lapisan Transformer berikutnya.

Arsitektur ini sangat relevan dengan studi ini. Kemampuan OSOORA untuk memahami pertanyaan pengguna yang kompleks, mengenali nuansa, dan

memberikan respons yang relevan serta secara gramatikal benar, menunjukkan keefektifan desain Transformer. Tanpa kemampuan self-attention untuk memahami konteks secara mendalam, sebuah chatbot tidak akan mampu memberikan nasihat hukum yang relevan dan kredibel.

2.2.6 Alibaba Cloud AI Platform: Model Studio dan DashScope

Dalam ekosistem Alibaba Cloud, Model Studio berfungsi sebagai platform pengembangan atau "bengkel" AI, tempat pengembang dapat memilih, mengonfigurasi, menguji, dan men-deploy berbagai model AI generatif. Di sinilah System Prompt untuk OSOORA didefinisikan untuk membentuk perilaku model. Sementara itu, DashScope adalah nama layanan API (Model-as-a-Service) yang menyediakan akses terkelola ke model-model yang telah siap di Model Studio. Aplikasi OSOORA berinteraksi dengan API DashScope untuk mengirim kueri pengguna dan menerima respons yang dihasilkan oleh model yang telah dikonfigurasi sebelumnya (Alibaba Cloud, 2023).

2.2.7 Kerangka Teori

Kerangka teori penelitian ini dibangun atas sinergi antara teknologi pengembangan web modern, kekuatan pemrosesan bahasa dari GenAI, dan metodologi pengembangan yang terstruktur namun fleksibel.

Fondasi Aplikasi: Teknologi frontend (HTML, CSS, JS) dan backend (Node.js, Express) membentuk platform web yang interaktif dan responsif.

Manajemen Data: MongoDB Atlas berfungsi sebagai repositori untuk data pengguna dan riwayat percakapan, memastikan persistensi dan keamanan data.

Mesin Kecerdasan: Model LLM dari Alibaba Cloud Model Studio bertindak sebagai "otak" dari sistem, yang bertanggung jawab untuk memahami dan menjawab pertanyaan pengguna.

Struktur dan Alur: Arsitektur sistem (Diagram Konteks, Use Case, Deployment) dan metodologi Adaptive Waterfall menjadi panduan untuk memastikan pengembangan berjalan secara sistematis, terarah, dan dapat beradaptasi.

Secara keseluruhan, kerangka ini mengintegrasikan berbagai komponen untuk menciptakan sebuah sistem chatbot yang tidak hanya fungsional secara teknis tetapi juga relevan dalam menjawab kebutuhan pengguna akan informasi hukum.

2.2.8 Teknologi Pengumpulan Data Semi-Otomatis

Bagian ini menguraikan teknologi yang digunakan dalam proses pengumpulan data kualitatif, yang memungkinkan eksplorasi web secara semi-otomatis untuk mengidentifikasi sentimen dan keluhan publik.

LangChain: Merupakan sebuah kerangka kerja (framework) sumber terbuka yang dirancang untuk menyederhanakan pengembangan aplikasi yang ditenagai oleh Large Language Models (LLM). LangChain menyediakan komponen modular dan agent yang dapat digabungkan untuk menciptakan alur kerja yang kompleks. Dalam penelitian ini, LangChain digunakan untuk membangun AI Agent yang mampu menerima instruksi dalam bahasa alami dan berinteraksi dengan alat lain, seperti Playwright, untuk menjalankan tugas-tugas di web (Chase, 2023).

Playwright: Adalah sebuah pustaka (library) dari Microsoft untuk otomatisasi peramban web (browser automation). Playwright memungkinkan skrip untuk mengontrol peramban seperti Chromium, Firefox, dan WebKit. Fungsinya adalah untuk menavigasi halaman web, berinteraksi dengan elemen (seperti mengklik tombol atau mengisi formulir), dan mengekstrak konten. Dalam konteks ini, Playwright bertindak sebagai "tangan" dari AI Agent yang dibangun dengan LangChain, yang secara fisik menjalankan perintah di peramban.

Browser Use: Merupakan sebuah alat bantu (toolkit) yang dirancang khusus untuk membuat situs web lebih mudah diakses oleh AI Agent. Alat ini menyediakan serangkaian perintah sederhana yang dapat dipahami oleh LLM untuk melakukan tugas-tugas navigasi web yang kompleks. Dengan

mengintegrasikan Browser Use, AI Agent yang dibangun dengan LangChain dan Playwright dapat menjalankan tugas eksplorasi web secara lebih efisien dan andal, bahkan pada situs web dinamis yang sulit ditangani oleh metode scraping tradisional (Browser Use contributors, 2025).

Secara keseluruhan, kerangka ini mengintegrasikan berbagai komponen untuk menciptakan sebuah sistem chatbot yang tidak hanya fungsional secara teknis tetapi juga relevan dalam menjawab kebutuhan pengguna akan informasi hukum.