

BAB I

PENDAHULUAN

Bab ini menyajikan dasar pemikiran dan kerangka konseptual dari penelitian yang dilakukan. Pembahasan dimulai dengan latar belakang masalah yang mengidentifikasi kesenjangan dalam akses dan pemahaman artikel ilmiah berbahasa Inggris di kalangan akademisi Indonesia, serta potensi teknologi *Large Language Model* (LLM) dalam mengatasi permasalahan tersebut. Selanjutnya, bab ini akan menguraikan rumusan masalah yang menjadi fokus penelitian, tujuan yang ingin dicapai, manfaat yang diharapkan, batasan penelitian, serta sistematika penulisan yang digunakan dalam penelitian ini

1.1 Latar Belakang

Perkembangan teknologi kecerdasan buatan / *Artificial Intelligence* (AI) telah mencapai kemajuan yang signifikan, terutama dalam bidang pemrosesan bahasa alami / *Natural Language Processing* (NLP) [1], [2]. Pencapaian terpenting dalam beberapa tahun terakhir ditandai dengan hadirnya *Large Language Model* (LLM) yang menunjukkan kemampuan luar biasa dalam memahami dan menghasilkan teks dalam berbagai bahasa [1], [3], [4]. Kemajuan ini membuka berbagai peluang untuk mengembangkan solusi inovatif di berbagai bidang, khususnya dalam konteks akademis dan penelitian ilmiah [5], [6].

Di era globalisasi ini, bahasa Inggris telah menjadi bahasa utama dalam publikasi ilmiah internasional. Namun, hal ini justru menimbulkan tantangan serius bagi banyak negara non-penutur bahasa Inggris, termasuk Indonesia. Permasalahan ini menjadi semakin nyata dalam konteks akademis di Indonesia, di mana terdapat kesenjangan yang signifikan dalam kemampuan mengakses dan memahami artikel-artikel ilmiah berbahasa Inggris. Terdapat banyak mahasiswa doktoral Indonesia yang sering menghadapi kesulitan dalam menulis dan memahami teks akademik berbahasa Inggris karena beberapa faktor, seperti pelatihan bahasa asing yang kurang memadai, kosakata yang terbatas, dan kurangnya budaya penelitian.

Kondisi ini secara langsung menghambat kemajuan akademik mereka karena kesulitan dalam mengakses dan memahami artikel ilmiah secara efektif [7]–[9].

Dalam upaya mengatasi permasalahan pemahaman dan analisis artikel ilmiah, berbagai penelitian telah dilakukan dengan pendekatan yang beragam. Telah dilakukan *literature review* komprehensif tentang sistem *Automatic Text Summarization* (ATS) untuk Bahasa Indonesia, mengungkapkan bahwa mayoritas sistem yang ada masih terbatas pada ringkasan dokumen tunggal dan menggunakan pendekatan ekstraktif berbasis statistik [10]. Selain itu, terdapat pengembangan pendekatan baru dalam otomatisasi ringkasan artikel ilmiah dengan menggunakan teknik hibrida yang menggabungkan ekstraksi dan abstraksi, namun sistem tersebut masih belum mencakup kemampuan analisis komparatif antar artikel [11].

Kemajuan dalam bidang penerjemahan otomatis juga menunjukkan perkembangan yang menjanjikan. Terdapat keberhasilan penggunaan model '*transformer*' dengan teknik *generative adversarial network* (GAN) dalam menerjemahkan teks dari bahasa Inggris ke bahasa Mandarin, dengan hasil terjemahan yang mendekati kualitas terjemahan manusia [12]. Sementara itu, dalam konteks pengembangan LLM, telah berhasil digunakan model GPT dari OpenAI untuk menghasilkan ringkasan artikel yang koheren dan mudah dipahami, membuktikan potensi besar LLM dalam memahami dan meringkas konten ilmiah dengan tingkat akurasi yang tinggi [13].

Seiring dengan perkembangan teknologi AI, beberapa platform dan *tools* berbasis AI telah dikembangkan untuk membantu peneliti dalam menganalisis dan memahami artikel ilmiah. SciSpace (sebelumnya Typeset) merupakan platform AI yang memungkinkan pengguna untuk menjelajahi dan memahami lebih dari 270 juta artikel ilmiah dengan kemampuan *chat* interaktif dengan PDF penelitian [14], [15]. Elicit adalah asisten penelitian AI yang memungkinkan pengguna untuk mencari, meringkas, mengekstrak data, dan berinteraksi dengan lebih dari 125 juta artikel ilmiah [16], [17]. Semantic Scholar menggunakan AI untuk memahami semantik literatur ilmiah dan membantu peneliti menemukan penelitian yang

relevan [18]. ResearchRabbit adalah platform penelitian komprehensif yang memanfaatkan *database* artikel akademik untuk menghasilkan rekomendasi literatur dan visualisasi yang dipersonalisasi [19].

Meskipun *tools* AI ini telah menunjukkan kemampuan yang menjanjikan dalam membantu peneliti mengakses dan memahami artikel ilmiah, sebagian besar masih berfokus pada analisis dokumen tunggal atau pencarian artikel. SciSpace menyediakan fitur ekstraksi data dari PDF dan *literature review* dengan bantuan AI, namun belum secara khusus mengintegrasikan kemampuan analisis komparatif mendalam antar artikel. Elicit mengotomatisasi sebagian *workflow* peneliti dan mempercepat proses *literature review*, tetapi masih terbatas pada ekstraksi dan sintesis bukti tanpa fokus pada analisis komparatif yang komprehensif [15], [16].

Framework LangChain telah muncul sebagai solusi menjanjikan dan memberikan peluang baru dalam pengembangan aplikasi berbasis LLM [20]–[22]. *Framework* ini tidak hanya menyediakan struktur dan komponen yang memudahkan pengembangan aplikasi AI kompleks, tetapi juga menawarkan fleksibilitas dalam mengintegrasikan berbagai sumber data dan model bahasa. Telah didemonstrasikan keunggulan *framework* ini melalui pengembangan AI *Agent* yang mampu melakukan analisis data kompleks dan visualisasi dari *input* file PDF dan CSV [23]. Bahkan telah berhasil digunakan arsitektur LangChain untuk menerjemahkan kueri bahasa manusia ke dalam SQL, membuktikan fleksibilitas *framework* ini dalam berbagai aplikasi bahasa [24].

Meskipun berbagai penelitian telah dilakukan, masih terdapat kesenjangan yang signifikan dalam pengembangan sistem yang dapat mengintegrasikan analisis komparatif artikel ilmiah dengan kemampuan penerjemahan ke bahasa Indonesia. Sistem pembuat ringkasan yang ada masih berfokus pada dokumen tunggal, sementara kebutuhan untuk membandingkan dan menganalisis lebih dari satu artikel secara bersamaan semakin meningkat. Selain itu, belum ada sistem terintegrasi yang menggabungkan kemampuan analisis komparatif dengan penyajian hasil dalam bahasa Indonesia yang mudah dipahami [11], [13], [25].

Penelitian ini mengusulkan solusi inovatif melalui pengembangan asisten virtual yang mengintegrasikan kekuatan *Large Language Model* (LLM) dan *framework* LangChain untuk melakukan analisis komparatif otomatis pada artikel jurnal ilmiah berbahasa Inggris [13], [21], [22], [26], [27]. Sistem yang diusulkan akan secara otomatis mengekstrak informasi penting dari artikel, melakukan perbandingan mendalam antar artikel untuk mengidentifikasi persamaan, perbedaan, dan hubungan antar artikel, serta menghasilkan ringkasan kontekstual hasil analisis dalam bahasa Indonesia yang komprehensif namun mudah dipahami. Pendekatan ini berbeda dari penelitian sebelumnya karena menggabungkan tiga komponen kunci dalam satu sistem terintegrasi yaitu analisis komparatif otomatis, kemampuan penerjemahan kontekstual, dan penyajian hasil yang terstruktur dalam bahasa Indonesia [10]–[13], [28].

Implementasi sistem memanfaatkan keunggulan LLM terkini yaitu model GPT dari OpenAI, dengan *framework* LangChain sebagai fondasi untuk manajemen alur kerja yang efisien pada asisten virtual yang berfungsi untuk melakukan analisis komparatif secara sistematis terhadap artikel-artikel ilmiah berbahasa Inggris serta mentransformasikan hasil analisis tersebut ke dalam bahasa Indonesia dengan pendekatan yang terstruktur dan komprehensif [13], [20], [29]–[32]. Penelitian ini diharapkan dapat memberikan solusi praktis bagi permasalahan akses dan pemahaman artikel ilmiah berbahasa Inggris di kalangan akademisi Indonesia, sekaligus membuka jalan bagi pengembangan asisten virtual akademis yang lebih canggih di masa depan [9]. Dengan mengotomatisasi proses analisis komparatif dan penyajian hasil dalam bahasa Indonesia, sistem ini akan membantu mempercepat proses penelitian dan meningkatkan kualitas kajian literatur yang dilakukan oleh peneliti Indonesia. Lebih lanjut, penelitian ini juga diharapkan dapat membuka jalan bagi pengembangan asisten virtual akademis yang lebih canggih di masa depan, mendorong inovasi dalam penggunaan AI untuk mendukung penelitian akademisi di Indonesia.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, penelitian ini difokuskan untuk menjawab beberapa pertanyaan utama terkait pengembangan sistem asisten virtual berbasis LLM dan LangChain. Adapun rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana pengembangan asisten virtual berbasis LLM (Large Language Model) dan LangChain yang dapat melakukan analisis komparatif otomatis dari dua artikel jurnal ilmiah berbahasa Inggris?
2. Bagaimana mengimplementasikan sistem yang dapat menghasilkan analisis komparatif komprehensif berbahasa Indonesia dari perbandingan dua artikel ilmiah berbahasa Inggris?
3. Seberapa efektif performa sistem dalam hal akurasi, stabilitas, dan efisiensi analisis komparatif yang dihasilkan?

1.3 Batasan Penelitian

Agar penelitian ini lebih terarah dan tidak melebar dari fokus utama, ditetapkan beberapa batasan sebagai berikut:

1. Bahasa pemrograman yang digunakan pada penelitian ini dikembangkan menggunakan bahasa pemrograman Python versi 3.11 atau yang lebih tinggi sebagai bahasa pemrograman utama. *Framework* yang digunakan adalah LangChain yang diintegrasikan dengan berbagai *library* Python untuk mengelola LLM.
2. Batasan komunikasi data sistem hanya mampu memproses artikel dalam format PDF dengan ukuran maksimal file 10 MB untuk setiap artikel. Pengoperasian sistem membutuhkan koneksi internet yang stabil untuk mengakses API LLM dan layanan terjemahan yang digunakan dalam proses analisis dan pembuatan ringkasan kontekstual.
3. Penelitian ini dibatasi pada analisis artikel jurnal berbahasa Inggris dengan maksimal dua artikel yang dapat dibandingkan secara bersamaan.

4. Fokus analisis hanya pada artikel dalam bidang sosial dan humaniora (sub-bidang sejarah & sosiologi), serta tidak mencakup analisis terhadap gambar, tabel, atau grafik dalam artikel.
5. Sistem yang dikembangkan hanya menyediakan analisis perbandingan untuk aspek persamaan, perbedaan, dan keterkaitan antar artikel. Sistem tidak mencakup analisis sitasi dan referensi, serta tidak melakukan analisis statistik mendalam terhadap data penelitian.
6. Sistem dirancang untuk tidak menyimpan artikel yang dianalisis secara permanen dalam *database*. Tidak tersedia fitur untuk membagikan hasil analisis ke platform eksternal, dan hasil analisis hanya dapat diakses oleh pengguna yang melakukan *upload* artikel ke dalam sistem.
7. Antarmuka sistem dibangun menggunakan Streamlit, dengan batasan visualisasi dan interaksi sesuai dengan fitur yang tersedia di platform tersebut. Sistem tidak mendukung antarmuka *multi-user* secara paralel maupun penyimpanan sesi pengguna.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dikemukakan, tujuan dari penelitian ini adalah sebagai berikut:

1. Mengembangkan asisten virtual berbasis LLM dan LangChain untuk melakukan analisis komparatif otomatis dari dua artikel pada jurnal ilmiah berbahasa Inggris.
2. Mengimplementasikan sistem otomatis yang dapat menghasilkan analisis komparatif komprehensif dari hasil komparasi dua artikel ilmiah berbahasa Inggris dengan penyajian dalam bahasa Indonesia.
3. Mengevaluasi performa sistem dalam hal akurasi, stabilitas, efisiensi operasional, dan penerimaan *expert* terhadap hasil analisis yang dihasilkan.

1.5 Manfaat Penelitian

Penelitian ini menawarkan sejumlah kontribusi signifikan bagi berbagai pihak. Melalui pengembangan asisten virtual berbasis kecerdasan buatan untuk analisis komparatif artikel jurnal ilmiah, penelitian ini memberikan solusi inovatif yang dapat meningkatkan efisiensi proses penelitian, meningkatkan aksesibilitas informasi ilmiah, serta mendorong kemajuan dalam bidang pemrosesan bahasa alami. Manfaat tersebut terbagi ke dalam tiga kategori utama yang dijelaskan secara rinci pada bagian berikut.

1.5.1 Manfaat bagi Peneliti dan Akademisi

Penelitian ini memberikan kontribusi nyata dalam mendukung aktivitas akademik dan proses penelitian ilmiah. Manfaat yang dapat dirasakan oleh peneliti dan kalangan akademisi meliputi:

1. Efisiensi Analisis Literatur

Asisten virtual yang dikembangkan menawarkan kemampuan untuk melaksanakan analisis perbandingan artikel jurnal ilmiah secara otomatis, sehingga memberikan penghematan waktu dan energi yang signifikan dalam proses pengkajian literatur akademis.

2. Peningkatan Aksesibilitas

Dengan menyajikan ringkasan kontekstual dalam Bahasa Indonesia, peneliti yang kurang mahir berbahasa Inggris dapat lebih mudah memahami konten artikel ilmiah berbahasa Inggris.

3. Identifikasi Gap Penelitian

Teknologi ini memungkinkan peneliti untuk lebih mudah mengenali kesenjangan penelitian dengan cara membandingkan (research gap) berbagai temuan dari beragam artikel secara sistematis.

1.5.2 Manfaat Bagi Masyarakat Umum

Selain memiliki manfaat bagi kalangan peneliti dan akademisi, sistem ini juga memberikan manfaat luas bagi masyarakat umum. Beberapa manfaat tersebut meliputi:

1. Penyebaran Pengetahuan Ilmiah

Ringkasan kontekstual dalam Bahasa Indonesia memungkinkan masyarakat awam untuk mengakses dan memahami informasi ilmiah yang sebelumnya hanya tersedia dalam bahasa asing.

2. Peningkatan Literasi Ilmiah

Dengan menyederhanakan informasi kompleks menjadi ringkasan yang mudah dipahami, sistem ini dapat meningkatkan literasi ilmiah di Indonesia.

1.5.3 Manfaat Bagi Pengetahuan dan Teknologi

Di sisi lain, penelitian ini juga berkontribusi terhadap perkembangan ilmu pengetahuan dan teknologi, khususnya di bidang kecerdasan buatan dan pemrosesan bahasa alami. Manfaat tersebut meliputi:

1. Inovasi dalam Pemrosesan Bahasa Alami

Penelitian ini berkontribusi pada pengembangan aplikasi LLM dan LangChain untuk tugas-tugas spesifik seperti analisis komparatif dan ringkasan kontekstual.

2. Referensi untuk Penelitian Selanjutnya

Temuan dan metodologi dalam penelitian ini dapat dijadikan landasan bagi pengembangan sistem serupa di berbagai bidang lainnya, seperti analisis dokumen hukum atau medis, sehingga membuka jalan bagi inovasi berkelanjutan.

1.6 Struktur Penulisan Skripsi

Penelitian ini disusun dengan mengacu pada Pedoman Penulisan Karya Ilmiah Universitas Pendidikan Indonesia (UPI) edisi 2024. Mengacu pada pedoman tersebut, struktur skripsi terdiri dari:

1. Bab 1: Pendahuluan

Bab ini menguraikan latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, serta ruang lingkup penelitian. Latar belakang menjelaskan isu atau topik yang dikaji secara sistematis dan relevan dengan permasalahan yang diteliti.

2. Bab II: Tinjauan Pustaka

Bab ini menyajikan teori-teori dan hasil penelitian terdahulu yang relevan sebagai dasar untuk mendukung penelitian. Tinjauan pustaka mencakup kerangka teori dan konsep yang menjadi landasan penelitian, serta posisi teoritis peneliti dalam konteks masalah yang dikaji.

3. Bab III: Metodologi

Bab ini menjelaskan metode yang digunakan dalam penelitian, yang mencakup jenis penelitian, populasi dan sampel, teknik pengumpulan data, serta prosedur analisis data. Metodologi disusun secara sistematis untuk menjawab rumusan masalah dan mencapai tujuan penelitian.

4. Bab IV: Hasil dan Pembahasan

Bab ini menyajikan temuan penelitian dalam bentuk teks, tabel, atau grafik, disertai dengan interpretasi dan pembahasan secara mendalam. Pembahasan mengaitkan hasil penelitian dengan teori atau temuan sebelumnya untuk menghasilkan analisis yang komprehensif.

5. Bab V: Kesimpulan dan Saran

Bab ini merangkum hasil penelitian secara ringkas serta menjawab rumusan masalah. Selain itu, bab ini juga memuat rekomendasi untuk penelitian selanjutnya dan implikasi praktis dari hasil penelitian.