

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Dalam beberapa tahun terakhir, perkembangan teknologi kecerdasan buatan (AI) telah membuka peluang baru untuk kemajuan di berbagai sektor, termasuk bidang hukum. Berbagai studi telah meneliti pengembangan kecerdasan buatan yang menargetkan berbagai kegiatan hukum (Nitta & Satoh, 2020) termasuk diantaranya penelitian hukum (Jhanvi Arora dkk., 2020), analisis perkara hukum (Drápal dkk., 2023), penyusunan (*drafting*) dan peninjauan (*review*) kontrak (Wang, 2024), *chatbot* (Shu dkk., 2024), serta otomasi (Nithya dkk., 2024). Masing-masing kegiatan hukum tersebut memiliki alur kerjanya masing-masing.

Pada penyusunan dan peninjauan kontrak, dapat terjadi proses berulang terhadap kontrak-kontrak sejenis. Penyusunan atau peninjauan kontrak sering mengandalkan *template* dari kontrak serupa sebelumnya, dengan modifikasi yang diperlukan sesuai kebutuhan (Dan Simonson dkk., 2019). Dengan sifat alamiah repetisi tersebut, proses penyusunan/peninjauan kontrak secara otomatis dapat menjadi pendorong untuk menggunakan AI untuk mencapai efisiensi waktu dan tenaga (Carneiro-Diaz dkk., 2025).

Pembuatan *template* menjadi sangat penting bagi institusi atau perusahaan yang melakukan proses kontrak yang berulang dan rutin, seperti perjanjian pengadaan, perjanjian layanan, perjanjian langganan, perjanjian jual beli, perjanjian sewa, dan sebagainya. Proses-proses yang berulang dan rutin ini sebagian besar terjadi karena kesamaan substansi dan bahkan tekstual. Hal tersebut menjadi ruang optimalisasi dalam proses penyusunan/peninjauannya, terutama untuk meningkatkan efisiensi waktu, kecepatan penyelesaian, dan konsistensi.

Salah satu tantangan utama dalam penerapan kecerdasan buatan di bidang hukum adalah kerumitan bahasa hukum itu sendiri, dimana gaya penulisan yang khas dan seringkali tidak literal, penggunaan jargon bermakna implisit, struktur norma yang berlapis, saling merujuk disertai pengecualian, serta ketergantungan pada konteks tempus dan wilayah hukum (Hannah dkk., 2025). Bahasa hukum Indonesia sendiri, terdiri dari

konstelasi hubungan yang saling mencirikan, sangat logis (bermain kata berdasarkan aturan logika yang benar dalam bahasa Indonesia), argumentatif (mengemukakan dasar, landasan, dan alasan), serta menggunakan diksi khas seperti istilah pleidoi, tersangka, terdakwa, banding, memori banding, dan lain-lain (Ola, 2009).

Penggunaan AI di bidang hukum juga memiliki beberapa keterbatasan, seperti kecenderungan untuk menyederhanakan berlebihan (*over-summarizing*), mengabaikan informasi penting, gagal menangkap hubungan sebab-akibat antara hak dan kewajiban, serta kurangnya pengetahuan domain spesifik tentang kontrak (Kim dkk., 2025). Dengan keterbatasan tersebut, AI dapat dikatakan merupakan alat bantu yang memberikan rekomendasi tambahan saat membuat keputusan (Zeng dkk., 2025).

Beberapa studi telah dilakukan terkait dengan penggunaan AI dalam proses *contract drafting*. Teknik vektorisasi dari pemrosesan bahasa alami dan algoritma klasifikasi telah digunakan untuk mengklasifikasikan jenis kalimat dan mengidentifikasi pihak terkait dalam kontrak (Dikmen dkk., 2025). Selain itu, metode berbasis pipeline yang menggabungkan pengambilan pengetahuan (text-embedding-3-small), pencocokan klausul (BM25), dan pembuatan jawaban (gpt-4o-mini) juga telah dikembangkan (Kim dkk., 2025). Model bahasa besar (*large language model*, LLM) telah menjadi bagian penting dalam proses peninjauan kontrak. Kemampuannya dalam menalar, mengikuti instruksi, dan memahami konteks, yang dapat diarahkan melalui teknik rekayasa prompt (Almeida dkk., 2024; Zhang dkk., 2023), sangat bermanfaat untuk mengoptimalkan lebih lanjut penggunaan LLM dalam proses peninjauan kontrak.

Dalam proses *review* kontrak, *lawyer* melakukan analisis pada tingkat pasal demi pasal. Hal ini meliputi *review* klausul standar pada setiap pasal, hubungan antar pasal, risiko inter dan antar pasal, serta analisis menyeluruh terhadap kontrak. Alur kerja ini pada dasarnya dapat direplikasi menggunakan *artificial intelligence* yang melibatkan LLM. Agar LLM mampu mempelajari pasal demi pasal, langkah pertama yang krusial adalah pemilihan *template* yang tepat untuk memastikan bahwa klausul standar yang telah ditetapkan sebelumnya dapat dijadikan tolok ukur perbandingan terhadap klausul dalam draft kontrak yang ditinjau.

Penelitian ini bermaksud menyelidiki pemanfaatan LLM dalam rangkaian proses automasi *review* kontrak berbasis template pada tahapan pengenalan *template* perjanjian, yang merupakan tahap krusial dalam mendapatkan template yang tepat untuk analisa

lebih lanjut. Fokus diarahkan pada bagaimana LLM dapat digunakan untuk mengidentifikasi kesesuaian struktur dan substansi antara *draft* perjanjian dan *template* terkait dengan metode *prompt engineering* yang diperkaya dengan pengetahuan tambahan (*in-context learning*) tanpa perlu menggunakan proses vektorisasi/*embedding*. Pendekatan ini dapat menjadi fondasi penting untuk membangun sistem automasi *review* draft kontrak yang lebih sederhana, akurat, efektif, dan adaptif terhadap keragaman format dan substansi kontrak, sekaligus menjaga keaslian teks dalam *template* maupun draft.

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut di atas, maka rumusan masalah yang akan dijawab dalam penelitian ini adalah sebagai berikut:

1. Bagaimana tingkat keyakinan dan akurasi metode *text similarity* berbasis leksikal/statistik dan berbasis representasi semantik dalam mengidentifikasi kesesuaian antara teks draft kontrak dengan *template* kontrak yang relevan?
2. Bagaimana pemanfaatan LLM berbasis penalaran generatif dalam proses identifikasi *template* kontrak mampu meningkatkan tingkat keyakinan dan akurasi sistem rekomendasi pemilihan *template*?

1.3. Tujuan

Penelitian ini bertujuan untuk membandingkan kinerja metode *text similarity* berbasis leksikal/statistik (Jaccard, TF-IDF), representasi semantik (BERT, LaBSE), dan penalaran berbasis model generatif dalam proses identifikasi *template* kontrak. Adapun tujuan khususnya adalah sebagai berikut:

- a. Menganalisis tingkat keyakinan dan akurasi metode *text similarity* berbasis leksikal/statistik dan berbasis representasi semantik dalam mengidentifikasi kesesuaian antara teks *draft* kontrak dengan *template* kontrak yang relevan.
- b. Mengevaluasi tingkat keyakinan dan akurasi pendekatan generatif berbasis LLM dalam proses identifikasi *template* kontrak secara lebih konsisten dan deterministik.

1.4. Manfaat

Manfaat yang didapatkan dari hasil penelitian ini adalah sebagai berikut:

- a. Memberikan kontribusi pada keilmuan bidang komputer secara umum dalam metode *text similarity* menggunakan model generatif, yang tidak hanya dapat digunakan

sebagai evaluator semantik berpenalaran yang menghasilkan hasil evaluasi yang lebih deterministik.

- b. Memberikan kontribusi pada keilmuan bidang hukum secara khusus dalam rangka memastikan akurasi pada proses automasi yang *scalable* terhadap berbagai variasi teks, sehingga dapat meningkatkan efisiensi proses kerja bidang hukum kontrak dalam melakukan proses review draft yang bersifat repetitif.
- c. Memberikan fondasi bagi penelitian dan pengembangan sistem automasi lebih lanjut terutama pada proses evaluasi substansi hukum pada pasal-pasal kontrak yang direview.
- d. Memberikan kontribusi dalam pengembangan aplikasi kecerdasan buatan untuk mendukung sistem rekomendasi yang lebih akurat dan deterministik.

1.5. Ruang Lingkup Masalah

Untuk tujuan spesifik yang hendak dicapai agar tetap pada jalur pembahasan sebagaimana rumusan masalah dimaksud, peneliti membatasi ruang lingkup permasalahan sebagai berikut:

- a. Penelitian ini menggunakan metode *text similarity* berbasis leksikal (Jaccard dan TF-IDF), metode *text similarity* berbasis embedding semantik (BERT dan LaBSE dengan *cosine similarity*), serta LLM sebagai mekanisme penalaran semantik dalam sistem rekomendasi template kontrak.
- b. Dalam penelitian ini digunakan model LLM yang tersedia secara bebas dan berukuran kecil sehingga penerapannya dapat direplikasi dengan menggunakan komputer kelas desktop. Model LLM yang digunakan adalah Qwen2.5-7B-Instruct dengan kuantisasi SmoothQuant untuk menjaga kualitas model (Guangxuan Xiao dkk., 2023) serta dapat dimuat ke dalam 1 kartu GPU Nvidia RTX 3060.
- c. Teks kontrak yang menjadi sampel untuk *draft* dan *template* adalah dokumen kontrak aktual yang diambil dari website Pejabat Pengelola Informasi dan Dokumentasi (PPID) beberapa instansi penyedia dokumen kontrak yang dapat diakses publik.
- d. Penelitian ini berfokus pada tahap automasi penentuan *template* kontrak berdasarkan *text similarity* antara *draft* dengan teks *template* dan tidak melalui tahapan analisa substansi hukum kontraktual pasal demi pasal yang memerlukan keahlian di bidang hukum.

- e. Pengujian dalam penelitian ini tidak mengukur waktu pemrosesan, mengingat urgensi mengutamakan tingkat keyakinan dan ketepatan rekomendasi template dibandingkan dengan kecepatan.
- f. Pembuatan *template* dilakukan secara otomatis menggunakan model LLM untuk membersihkan variabel-variabel yang umum terdapat dalam kontrak dan menggantinya menjadi placeholder.

1.6. Kebaruan

Penelitian ini menyajikan kebaruan dalam cara menentukan kemiripan semantik kontrak untuk pengenalan template secara deterministik, yang berbeda dari pendekatan text similarity konvensional berbasis leksikal (Jaccard, TF-IDF) maupun embedding semantik (BERT, LaBSE) yang bersifat statistik dan implisit.

Kebaruan utama penelitian ini terletak pada pemanfaatan *Large Language Model* (LLM) sebagai mesin *semantic reasoning*, bukan sebagai model generatif *end-to-end*. Melalui rekayasa prompt yang terkontrol, LLM digunakan untuk menalar kesamaan dan perbedaan substansi hukum antar teks kontrak, sehingga keputusan pemilihan template tidak hanya bergantung pada skor kemiripan numerik, tetapi pada kesesuaian makna dan fungsi hukum.

Penelitian ini juga mengusulkan pipeline komputasional bertahap yang secara eksplisit memisahkan pengenalan template kontrak pada tingkat dokumen, pemetaan pasal pada tingkat klausul, serta mengevaluasi setiap tahap secara mandiri. Pendekatan ini berbeda dari penelitian terdahulu yang umumnya mengasumsikan template telah diketahui atau langsung melakukan analisis substansi kontrak.

Kontribusi ilmiah penelitian ini yaitu:

- a. kerangka evaluasi deterministik untuk pengenalan template kontrak berbasis perbandingan semantik,
- b. analisis perilaku metode *similarity* (leksikal, embedding, dan LLM) berdasarkan ketajaman pemisahan keputusan dan stabilitas skor, serta
- c. pendekatan multi-tahap yang dapat direproduksi sebagai fondasi sistem automasi *review* kontrak yang akuntabel dalam konteks hukum.

Pendekatan ini memperluas kajian penerapan AI di bidang hukum dengan fokus pada karakteristik tahapan pengambilan keputusan yang krusial dalam sistem automasi hukum.

1.7. Hipotesis

Penelitian ini mengajukan hipotesis bahwa pemanfaatan *large language model* dengan pendekatan *in-context learning* melalui teknik *prompt engineering* mampu meningkatkan akurasi pengenalan *template* kontrak secara signifikan dalam sistem automasi *review* kontrak berbasis *template*.

Large language model diposisikan sebagai *semantic evaluator* yang tidak mengandalkan kesamaan leksikal atau perhitungan kemiripan vektor semantik, tetapi melakukan penalaran terarah berdasarkan instruksi dan konteks yang diberikan melalui *prompt*. Dengan pendekatan ini, rekomendasi *template* yang tepat berdasarkan evaluasi semantik oleh LLM pada sistem automasi *review* kontrak dapat lebih *robust* terhadap variasi struktur, redaksi, maupun penyusunan kalimat klausula.

Peningkatan ketepatan pada tahap pengenalan *template* tersebut selanjutnya diharapkan berdampak langsung pada peningkatan kualitas analisis lanjutan dalam sistem automasi *review* kontrak, mengingat tahap penentuan *template* merupakan basis bagi seluruh proses evaluasi pasal dan rekomendasi berikutnya.