

Pengembangan Question Answering System Menggunakan Large Language Model UU TNI 2025

Yola Permata Bunda^{*1}, Elyani², Muhammad Fadlan Siregar³

¹Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, Universitas Tjut Nyak Dhien
Indonesia

²Fakultas Sosial dan Humaniora, Program Studi Ilmu Hukum, Universitas Tjut Nyak Dhien, Indonesia

³Postgraduate Student University Malaysia Perlis, Pauh Street Perlis, Malaysia

Email: ¹yolapermata@utnd.ac.id, ²elyani@utnd.ac.id, ³fadlansiregar@studentmail.unimap.edu.my

Abstrak

Undang-Undang Tentara Nasional Indonesia (UU TNI) merupakan landasan hukum yang mengatur peran, tugas, dan fungsi TNI dalam menjaga kedaulatan negara. Seiring dengan perkembangan zaman, RUU TNI 2025 sedang dibahas untuk memperbarui undang-undang tersebut agar sesuai dengan tantangan zaman, khususnya dalam menghadapi ancaman siber dan perubahan global lainnya. Meskipun demikian, revisi RUU ini menuai kontroversi, terutama terkait dengan tiga pasal yang dianggap kontroversial dan berpotensi memperlebar ruang lingkup tugas TNI hingga melibatkan sektor sipil. Penelitian ini bertujuan untuk mengembangkan sebuah sistem tanya jawab berbasis Large Language Model (LLM) untuk mempermudah masyarakat memperoleh informasi secara cepat, akurat, dan mudah dipahami mengenai pasal-pasal kontroversial dalam RUU TNI 2025. Dengan menggunakan model BERT, sistem ini dapat memberikan jawaban otomatis yang relevan dan memproses pertanyaan yang diajukan oleh pengguna. Metode penelitian melibatkan pengumpulan data, pemrosesan teks, pelatihan model LLM, dan evaluasi kinerja menggunakan metrik precision, recall, dan F1 score. Hasil evaluasi menunjukkan sistem mencapai precision 0,7240, recall 0,7850, dan F1 score 0,7529, yang menandakan performa tinggi dalam memberikan jawaban relevan dan akurat. Hasil penelitian menunjukkan bahwa sistem yang dikembangkan mampu memberikan jawaban yang akurat dan relevan dengan tingkat keakuratan yang tinggi. Kesimpulannya, sistem tanya jawab berbasis LLM ini dapat menjadi solusi yang efektif dalam meningkatkan pemahaman masyarakat terhadap RUU TNI 2025, memperkenalkan teknologi kecerdasan buatan dalam mendukung transparansi informasi publik, serta mengurangi kebingungan terkait dengan peraturan yang kompleks.

Kata kunci: BERT, Large Language Model, Question Answering System, RUU TNI 2025.

Development Question Answering System Using Large Language Model TNI Law 2025

Abstract

The Indonesian National Armed Forces Act (TNI ACT) is the legal basis that regulates the role, duties, and functions of the TNI in maintaining the sovereignty of the state. In line with the times, the 2025 TNI Bill is being discussed to update the law to meet the challenges of the times, particularly in facing cyber threats and other global changes. However, the revision of this bill has sparked controversy, especially regarding three articles that are considered controversial and have the potential to broaden the scope of the TNI's duties to involve the civilian sector. This study aims to develop a question-and-answer system based on a Large Language Model (LLM) to make it easier for the public to obtain information quickly, accurately, and in an easy-to-understand manner regarding the controversial articles in the 2025 TNI Bill. Using the BERT model, this system can provide relevant automatic answers and process questions submitted by users. The research method involved data collection, text processing, LLM model training, and performance evaluation using precision, recall, and F1 score metrics. The evaluation results showed that the system achieved a precision of 0.7240, a recall of 0.7850, and an F1 score of 0.7529, indicating high performance in providing relevant and accurate answers. The research results show that the developed system is capable of providing accurate and relevant answers with a high level of accuracy. In conclusion, this LLM-based question and answer system can be an effective solution in improving public understanding of the 2025 TNI Bill, introducing artificial intelligence technology to support public information transparency, and reducing confusion related to complex regulations.

Keywords: BERT, Large Language Model, Question Answering System, RUU TNI 2025.

1. PENDAHULUAN

Undang-Undang Tentara Nasional Indonesia (UU TNI) merupakan dasar hukum yang mengatur mengenai keberadaan, peran, tugas, serta fungsi Tentara Nasional Indonesia (TNI) dalam menjaga kedaulatan negara, mempertahankan keutuhan wilayah, dan melindungi segenap bangsa Indonesia. Undang-Undang TNI yang telah disahkan pada tahun 2004, dengan nomor UU No. 34 Tahun 2004, menjadi acuan penting dalam kebijakan pertahanan negara [1].

Pada tahun 2025, RUU TNI 2025 sedang dibahas untuk memperbarui dan menyempurnakan UU TNI, dengan tujuan untuk meningkatkan efektivitas TNI dalam menjalankan tugas dan fungsinya. RUU ini bertujuan untuk menyelaraskan aturan mengenai organisasi, peran, serta kebijakan TNI dengan perkembangan teknologi modern, sistem pertahanan yang lebih terintegrasi, serta dinamika ancaman yang semakin beragam.

Revisi RUU TNI 2025 menuai kontroversi, terutama terkait dengan tiga pasal yang dianggap bermasalah. Pertama, penambahan tiga tugas baru bagi TNI, yaitu penanganan ancaman siber, penyelamatan WNI di luar negeri, dan penanggulangan narkoba, yang dinilai dapat mengganggu profesionalisme TNI dan memperluas cakupan operasi militer ke ranah sipil. Kedua, perwira TNI aktif yang kini dapat menjabat di 16 kementerian/lembaga, yang sebelumnya hanya dapat dilakukan oleh perwira yang telah pensiun, menimbulkan potensi konflik kepentingan. Ketiga, perpanjangan usia pensiun TNI, yang kini mencakup perwira hingga usia 62 tahun, berisiko memperpanjang pengaruh militer dalam struktur pemerintahan.

Kekhawatiran ini semakin kuat karena revisi ini dianggap berpotensi mengembalikan dwifungsi militer, yang sebelumnya menyebabkan keterlibatan TNI dalam bidang sipil. Hal ini bisa mempersempit ruang bagi lembaga sipil dan meningkatkan potensi pelanggaran HAM, terutama terkait dengan penanganan ancaman siber dan proyek strategis nasional yang melibatkan militer. Pendekatan yang lebih militeristik ini dikhawatirkan akan melemahkan prinsip demokrasi dan supremasi sipil.

Untuk mengatasi permasalahan tersebut, dibutuhkan sebuah sistem yang dapat mempermudah masyarakat dalam memperoleh informasi secara cepat, akurat, dan mudah dipahami mengenai isi RUU TNI 2025, khususnya yang berkaitan dengan pasal-pasal kontroversial. Salah satu solusi yang dapat diterapkan adalah pengembangan *Question Answering System* (Sistem Tanya Jawab) menggunakan *Large Language Model* (LLM), seperti GPT (*Generative Pre-trained Transformer*) [2]. Sistem ini dapat memberikan jawaban otomatis berdasarkan teks yang ada, dengan kemampuan memproses bahasa alami dan memahami konteks [3].

LLM memiliki kemampuan untuk memproses bahasa alami dan memahami konteks dari berbagai pertanyaan, serta memberikan jawaban yang relevan berdasarkan teks yang ada [4]. Dengan mengintegrasikan LLM dalam sistem tanya jawab, masyarakat dapat dengan mudah mengajukan pertanyaan mengenai RUU TNI 2025 dan memperoleh jawaban yang jelas tanpa harus membaca seluruh teks perundang-undangan. Teknologi ini telah terbukti efektif dalam berbagai aplikasi lain, seperti di bidang medis, hukum, dan layanan pelanggan, sehingga dapat menjadi solusi yang tepat untuk masalah pemahaman peraturan yang kompleks [5].

penelitian chatbot yang memanfaatkan LLM dan Retrieval Augmented Generation (RAG) untuk menjawab pertanyaan terkait artikel ilmiah di platform GARUDA. Evaluasi menggunakan ROUGE score menunjukkan rentang nilai antara 42,68% hingga 68,03%, menandakan kinerja sistem yang cukup baik dalam memberikan jawaban yang relevan [6].

sistem tanya jawab untuk Tafsir Al-Azhar dengan menggunakan Langchain dan LLM dalam bentuk chatbot Telegram. Pengujian menunjukkan akurasi sebesar 83,71% dan skor relevansi kontekstual sebesar 79%, menunjukkan efektivitas sistem dalam memberikan jawaban yang sesuai dan mudah dipahami oleh pengguna [7].

Potensi dan Tantangan Penggunaan ChatGPT dalam Pemasaran Penelitian mengenai pemanfaatan ChatGPT dalam bidang pemasaran menunjukkan bahwa teknologi ini memiliki kemampuan signifikan untuk mengotomatisasi pembuatan konten, mempercepat riset pasar, dan meningkatkan personalisasi layanan pelanggan. ChatGPT mampu menyajikan konten dengan cepat dan memberikan wawasan tentang perilaku konsumen yang membantu tim pemasaran dalam pengambilan keputusan. Namun, studi ini juga menyoroti risiko etis seperti kemungkinan pelanggaran privasi, bias data yang bisa memperkuat stereotip sosial, potensi penggantian tenaga kerja manusia, serta ketergantungan berlebihan terhadap teknologi AI. Oleh karena itu, diperlukan penerapan prinsip transparansi, mitigasi bias, perlindungan data pribadi, serta pengawasan manusia guna memastikan penggunaan ChatGPT yang bertanggung jawab dan beretika dalam pemasaran [8]. Dampak ChatGPT dan GPT-4 pada Dunia Akademik Hukum Penelitian ini mengkaji dampak penggunaan ChatGPT dan versi lebih lanjut GPT-4 dalam penulisan karya akademik dan ujian di bidang hukum. ChatGPT dapat membantu mahasiswa dalam memperbaiki kualitas bahasa serta mempercepat proses penulisan akademik. Namun, terdapat kekhawatiran terkait plagiarisme yang semakin sulit terdeteksi dan potensi

melemahkan kemampuan berpikir kritis mahasiswa. Ketergantungan berlebihan terhadap AI dapat mengurangi keterlibatan mahasiswa dalam riset tradisional dan diskusi mendalam. Oleh sebab itu, institusi pendidikan disarankan untuk mengatur penggunaan AI melalui kebijakan yang mengedepankan transparansi, mengajarkan keterampilan analisis kritis, dan memodifikasi metode evaluasi agar selaras dengan perkembangan teknologi baru ini [9].

Sistem OntoChatGPT: Integrasi Ontologi dengan ChatGPT untuk Peningkatan Kinerja Penelitian ini menghadirkan sistem OntoChatGPT yang menggabungkan ontologi dengan prompt terstruktur untuk meningkatkan performa ChatGPT, khususnya dalam konteks domain spesifik seperti rehabilitasi. Sistem ini menggunakan meta-ontologi sebagai fondasi pembuatan prompt yang dinamis dan terstruktur sehingga memungkinkan AI menghasilkan respon yang lebih relevan dan terarah. Implementasi OntoChatGPT di ranah bahasa Ukraina menunjukkan hasil baik dalam hal akurasi dan kemampuan ekstraksi entitas serta analisis kontekstual. Meski demikian, tantangan seperti menurunkan jumlah respon positif palsu masih menjadi fokus pengembangan lebih lanjut.[10]

Urgensi Pengawasan Regulasi untuk Model Bahasa Besar di Bidang Kesehatan Artikel ini menyoroti pentingnya regulasi khusus untuk model bahasa besar (Large Language Models/LLMs) seperti GPT-4 dalam bidang kesehatan yang memiliki potensi besar namun juga risiko serius. LLM memiliki kompleksitas dan kemampuan adaptasi tinggi yang berbeda dengan teknologi medis berbasis AI sebelumnya. Regulasi diperlukan untuk menjamin keamanan data pasien, mengatasi bias yang mungkin muncul, menetapkan tanggung jawab jika terjadi kesalahan medis akibat rekomendasi AI, serta memastikan transparansi dan keadilan dalam pemakaiannya. Penulis mengusulkan pembentukan kategori regulasi baru yang spesifik, pengawasan berkelanjutan, serta keterlibatan pasien dalam proses desain dan pengawasan guna menjaga kepercayaan dan integritas pelayanan kesehatan [11].

Pendekatan Audit Tiga Lapis untuk Model Bahasa Besar Penelitian ini memperkenalkan kerangka audit tiga lapis untuk memastikan model bahasa besar seperti ChatGPT dirancang dan digunakan dengan cara yang legal, etis, dan teknis andal. Audit dilakukan pada tiga tingkat yaitu audit tata kelola penyedia teknologi, audit kualitas model sebelum peluncuran, dan audit aplikasi yang menggunakan model tersebut. Dengan menggabungkan audit di ketiga level ini, risiko sosial dan etis dapat teridentifikasi serta dikelola secara lebih efektif. Artikel juga membahas batasan serta tantangan dalam proses audit dan menekankan perlunya kolaborasi antara penyedia teknologi, pembuat kebijakan, dan auditor independen agar tata kelola AI dapat berjalan optimal dan transparan [12]. sistem tanya jawab berbasis chatbot Telegram yang mengintegrasikan Langchain dan LLM untuk menjawab pertanyaan terkait UU Kesehatan. Hasil pengujian menggunakan BERTScore menunjukkan nilai precision 76%, recall 80%, dan F1-score 78%, menandakan sistem mampu memberikan jawaban yang akurat dan relevan [13]. mengembangkan QAS berbasis chatbot yang terintegrasi dengan Telegram, memanfaatkan Langchain dan LLM. Sistem ini menggunakan UU No. 17 Tahun 2023 tentang Kesehatan sebagai basis pengetahuan, dengan evaluasi menggunakan metrik BERTScore yang menunjukkan nilai precision 76%, recall 80%, dan F1-score 78% [14].

Pemecahan masalah yang ditawarkan dalam penelitian ini memiliki keunggulan dari penelitian ini yang terletak pada kemampuannya untuk menyediakan solusi yang efektif dalam memecahkan masalah terkait pemahaman regulasi yang kompleks dan memberikan akses yang mudah serta interaktif melalui penggunaan *Large Language Model* (LLM) dalam Sistem Tanya Jawab. kebaruan dari penelitian ini terletak pada pengembangan Sistem Tanya Jawab berbasis LLM yang disesuaikan dengan RUU TNI 2025 untuk meningkatkan transparansi, pemahaman, dan aksesibilitas informasi hukum yang spesifik dan kompleks di bidang pertahanan negara.

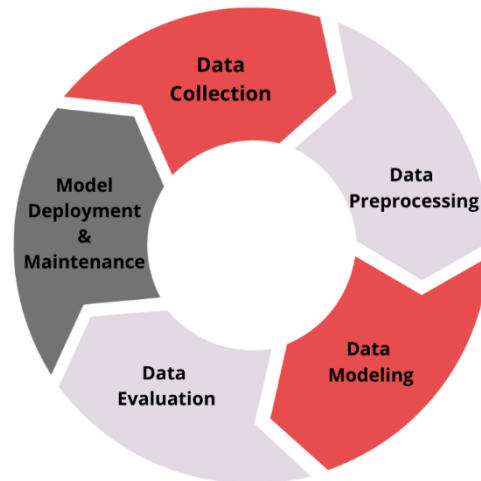
Melalui penggunaan *Question Answering System* berbasis LLM, diharapkan dapat meningkatkan aksesibilitas masyarakat terhadap informasi terkait RUU TNI 2025, mempermudah pemahaman terhadap regulasi yang berlaku, serta memperkenalkan teknologi kecerdasan buatan sebagai solusi untuk mendukung transparansi dan pemahaman terhadap peraturan hukum yang vital bagi pertahanan negara [15].

Rumusan Masalah dalam penelitian ini adalah bagaimana membuat sebuah sistem yang dapat mempermudah masyarakat dalam memperoleh informasi secara cepat, akurat, dan mudah dipahami mengenai isi RUU TNI 2025, khususnya yang berkaitan dengan pasal-pasal kontroversial?

2. METODE PENELITIAN

Bagian ini akan menjelaskan tahapan-tahapan yang dilakukan dalam penelitian ini. Proses penelitian ini terdiri dari serangkaian langkah yang disusun secara terstruktur untuk mencapai tujuan yang telah ditetapkan. Setiap tahapan dalam proses ini dirancang untuk memastikan bahwa sistem yang dibangun memiliki tingkat keakuratan yang tinggi. Selain itu, langkah-langkah tersebut saling terkait dan mendukung untuk memastikan setiap tahap memberikan kontribusi yang maksimal terhadap hasil akhir penelitian. Dengan pendekatan ini,

diharapkan penelitian ini dapat menghasilkan sistem yang kuat dan dapat diandalkan dalam menjawab pertanyaan terkait UU TNI 2025. Alur metode penelitian dimulai dari pengumpulan data, pembangunan model, hingga pengujian dengan data terbaru, seperti yang dapat dilihat pada Gambar 1.



Gambar 1. Alur Metode Penelitian

2.1. Data Collection

Pada tahap ini, data yang relevan tentang UU TNI 2025 dikumpulkan dari berbagai sumber resmi seperti dokumen pemerintah, peraturan resmi, artikel ilmiah, jurnal hukum, dan laporan analisis terkait. Data disimpan dalam format CSV yang memudahkan proses selanjutnya. Dataset terdiri dari 1 dokumen utama (RUU TNI 2025), 100 pertanyaan sampel yang dibagi menjadi 70% data latih, 10% data validasi, dan 20% data uji.

2.2. Data Preprocessing

Pada tahap ini mempersiapkan dan membersihkan data agar dapat diproses lebih lanjut oleh model. Proses dimulai dengan pembersihan teks (text cleaning), yang melibatkan penghapusan karakter-karakter yang tidak perlu, seperti tanda baca yang tidak relevan, spasi ekstra, atau karakter khusus yang tidak memberikan nilai tambah untuk analisis teks [16]. Selanjutnya, dilakukan case folding, yaitu mengubah semua teks menjadi huruf kecil (lowercase) agar tidak terjadi duplikasi yang tidak perlu akibat perbedaan antara huruf kapital dan kecil. Proses berikutnya adalah stopword removal, yang bertujuan untuk menghapus kata-kata umum dan tidak signifikan dalam konteks analisis, seperti "dan", "atau", "yang", dan "adalah". Selain itu, dilakukan juga stemming dan lemmatization, yang berfungsi untuk mengubah kata-kata dengan bentuk yang berbeda menjadi bentuk dasar, misalnya mengubah "berjalan" menjadi "jalan", guna mengurangi kompleksitas data dan memperjelas makna [17]. Tokenisasi juga menjadi langkah penting, di mana teks dipecah menjadi kata-kata atau token yang lebih kecil, yang kemudian dapat dianalisis secara individu. Terakhir, normalisasi dilakukan untuk menstandarkan format data, seperti angka atau tanggal, agar sesuai dengan standar yang diinginkan, sehingga data dapat lebih mudah diproses dan dianalisis dalam tahap berikutnya. Model yang dipilih adalah BERT, sebuah Large Language Model (LLM) yang sudah umum dan kuat dalam pemahaman bahasa alami. Model ini di-fine-tune menggunakan dataset spesifik RUU TNI 2025 agar dapat memahami konteks hukum dengan lebih tepat. Fine-tuning memakai parameter teknis: learning rate $2e-5$, batch size 16, epoch 5, optimizer AdamW, dan menggunakan scheduler linear warmup. Validasi model dilakukan menggunakan data validasi untuk mencegah overfitting serta memastikan performa yang optimal.

2.3. Data Modelling

Tahapan ketiga dalam pengembangan *Question Answering System* adalah pemodelan data, yang merupakan inti dari proses. Pada tahap ini, model *Large Language Model* (LLM) dibangun untuk memproses teks dan memberikan jawaban berdasarkan data yang telah diproses sebelumnya [18]. Langkah pertama adalah pemilihan model, di mana jenis LLM yang akan digunakan dipilih sesuai dengan kebutuhan dan tujuan proyek. Pada penelitian ini model yang digunakan adalah *BERT*. Setelah model dipilih, langkah selanjutnya adalah pre-training dan fine-tuning. Model LLM biasanya telah dipelajari dengan dataset besar secara umum (pre-trained), namun untuk kebutuhan yang lebih spesifik seperti UU TNI 2025, fine-tuning perlu dilakukan agar model lebih

memahami konteks hukum dan pertanyaan yang berkaitan dengan topik ini. Fine-tuning melibatkan penyesuaian model menggunakan dataset yang lebih relevan, seperti teks UU TNI, sehingga model dapat memberikan jawaban yang lebih akurat dan sesuai dengan konteks hukum yang dimaksud. Selain itu, pengaturan hyperparameter juga dilakukan untuk memastikan model belajar dengan optimal. *Hyperparameter* seperti *learning rate*, *batch size*, dan jumlah *epoch* perlu diatur dengan cermat agar proses pelatihan dapat berjalan dengan baik. Setelah model dilatih, validasi model dilakukan untuk memeriksa kinerjanya menggunakan data validasi [19]. Validasi ini bertujuan untuk menghindari *overfitting*, yaitu kondisi di mana model terlalu terlatih pada data pelatihan dan kurang mampu menggeneralisasi pada data baru. Dengan validasi yang tepat, performa model dapat dipastikan sesuai dengan harapan dan siap untuk diimplementasikan pada tahap berikutnya.

2.4. Data Evaluation

Tahapan keempat dalam pengembangan *Question Answering System* adalah evaluasi data, yang bertujuan untuk menilai performa model dalam memberikan jawaban yang akurat dan relevan. Langkah pertama adalah pengujian dengan data uji, di mana dataset yang berbeda dari data pelatihan digunakan untuk menguji kemampuan model dalam menjawab pertanyaan. Data uji ini dirancang untuk mencakup berbagai jenis pertanyaan yang relevan dengan UU TNI 2025, sehingga model dapat diuji dalam konteks yang lebih beragam dan nyata. Setelah itu, dilakukan penggunaan metrik evaluasi, yang merupakan langkah penting dalam mengukur kualitas model. Metrik yang umum digunakan dalam *Question Answering* meliputi *Accuracy*, *F1-score*, *Precision*, *Recall*, dan *ROUGE*. Metrik ini membantu mengukur sejauh mana model dapat memberikan jawaban yang benar, seberapa baik ia menangani jawaban yang tidak relevan, dan sejauh mana model dapat merangkum informasi dengan baik jika diperlukan. Terakhir, feedback dari pengguna sangat penting untuk mengetahui kualitas sistem dalam situasi dunia nyata. Dengan melakukan uji coba sistem oleh pengguna akhir, kita dapat mendapatkan umpan balik langsung mengenai kualitas jawaban yang diberikan oleh model dan seberapa akurat model tersebut dalam menjawab pertanyaan yang berkaitan dengan UU TNI 2025. Umpan balik ini memungkinkan perbaikan dan penyempurnaan lebih lanjut pada model, memastikan bahwa sistem berfungsi dengan baik dalam konteks penggunaannya.

2.4. Model Deployment & Maintenance

Tahapan terakhir dalam pengembangan *Question Answering System* adalah deployment dan pemeliharaan model, yang bertujuan untuk mengimplementasikan model dalam aplikasi nyata yang dapat digunakan oleh pengguna untuk mendapatkan jawaban terkait UU TNI 2025. Langkah pertama adalah deploy model, yaitu meng-host model pada platform hosting. Platform ini memungkinkan model untuk diakses oleh pengguna melalui aplikasi berbasis *web* atau *web mobile*, sehingga dapat digunakan secara praktis dalam kehidupan sehari-hari. Model ini dapat diintegrasikan dalam sistem yang mendukung tanya jawab otomatis, memungkinkan pengguna untuk mengajukan pertanyaan dan menerima jawaban secara langsung.

3. HASIL DAN PEMBAHASAN

3.1. Data RUU TNI 2025

Data dikumpulkan didapat dari website database peraturan pemerintah dari link <https://peraturan.bpk.go.id/Details/319166/uu-no-3-tahun-2025> [20] dalam bentuk PDF. Dari data pdf tersebut diolah menjadi sample question, answer dan context untuk di ujicoba. Data sample dapat dilihat pada tabel 1.

Tabel 1. Tabel Sampel Question, Answer Dan Context

Question	Answer	Context
Apa tujuan utama dari UU TNI yang baru?	Tujuan utama UU TNI adalah memperkuat pertahanan negara dan menjaga kedaulatan NKRI dari segala ancaman.	UU TNI 2025 mengatur tentang penguatan sistem pertahanan negara Indonesia.
Tujuan utama UU TNI adalah?	TNI dipimpin oleh Panglima TNI yang BERTanggung jawab langsung kepada Presiden sebagai Panglima Tertinggi.	Kepemimpinan TNI diatur dalam struktur komando yang jelas dan hierarkis.
Bagaimana struktur organisasi TNI dalam UU baru?	TNI terdiri dari TNI Angkatan Darat, TNI Angkatan Laut, dan TNI Angkatan Udara di bawah komando Panglima TNI.	Organisasi TNI terdiri dari tiga matra yang bekerja secara sinergis.
Apa peran TNI dalam menjaga kedaulatan	TNI berperan menjaga kedaulatan negara melalui operasi pertahanan, pengamanan wilayah, dan	Kedaulatan negara menjadi prioritas utama dalam tugas TNI.

negara? Bagaimana mekanisme pengangkatan perwira TNI?	penangkalan ancaman. Pengangkatan perwira TNI dilakukan melalui sistem pendidikan militer dan promosi berdasarkan prestasi dan kompetensi.	Sistem karier TNI dibangun berdasarkan merit dan profesionalisme.
---	---	---

Pejelasan dari tabel 1 diatas kolom satu question merupakan Kolom yang berisi pertanyaan yang diajukan terkait dengan konteks atau topik tertentu (misalnya terkait TNI dalam konteks ini). Sedangkan kolom answer berisi jawaban atau informasi yang bisa menjadi referensi untuk menjawab pertanyaan yang diajukan di kolom dan kolom context berisi informasi tambahan atau konteks yang memberikan latar belakang lebih mendalam terkait dengan pertanyaan dan jawaban.

3.2. Data Preprocessing dan Data Modelling

Langkah pertama adalah pembersihan teks (*Text Cleaning*), di mana teks yang akan dianalisis dibersihkan dari karakter-karakter yang tidak diinginkan, seperti tanda baca, karakter kontrol, atau spasi ekstra. Setelah teks dibersihkan, teks tersebut dibagi menjadi paragraf-paragraf berdasarkan pemisah baris baru. Setiap paragraf yang telah dibersihkan ini kemudian diubah menjadi vektor menggunakan *SentenceTransformer*, yang akan menghasilkan embedding untuk setiap paragraf.

Selanjutnya, proses pencarian paragraf yang paling relevan dengan pertanyaan dimulai dengan mengonversi pertanyaan itu sendiri menjadi vektor embedding. Dengan menggunakan cosine similarity, sistem membandingkan vektor embedding dari pertanyaan dengan vektor-vektor paragraf yang telah dibuat sebelumnya. Dengan cara ini, sistem dapat menemukan paragraf yang memiliki kesamaan tertinggi dengan pertanyaan, yang dianggap sebagai paragraf yang paling relevan untuk menjawab pertanyaan tersebut.

Setelah menemukan paragraf yang paling relevan, model extractive QA (dalam hal ini, model *xlm-roBERTa-base-squad2*) digunakan untuk mengekstrak jawaban dari paragraf tersebut. Model ini bekerja dengan memilih jawaban langsung dari konteks yang diberikan, yaitu paragraf yang ditemukan sebagai yang paling relevan. Jawaban yang diprediksi kemudian disimpan dalam kolom candidate di dataframe untuk setiap pertanyaan yang diajukan yang tampak pada tabel 2 berikut.

Tabel 2. Tabel Sampel Question, Answer Dan Context

Question	Answer	Context	Candidate
Apa tujuan utama dari UU TNI yang baru?	Tujuan utama UU TNI adalah memperkuat pertahanan negara dan menjaga kedaulatan NKRI dari segala ancaman.	UU TNI 2025 mengatur tentang penguatan sistem pertahanan negara Indonesia.	menegakkan kedaulatan
Tujuan utama UU TNI adalah?	TNI dipimpin oleh Panglima TNI yang BERTanggung jawab langsung kepada Presiden sebagai Panglima Tertinggi.	Kepemimpinan TNI diatur dalam struktur komando yang jelas dan hierarkis.	berhubungan dengan penegakan hukum dilaut
Bagaimana struktur organisasi TNI dalam UU baru?	TNI terdiri dari TNI Angkatan Darat, TNI Angkatan Laut, dan TNI Angkatan Udara di bawah komando Panglima TNI.	Organisasi TNI terdiri dari tiga matra yang bekerja secara sinergis.	berhubungan dengan penegakan hukum dilaut
Apa peran TNI dalam menjaga kedaulatan negara?	TNI berperan menjaga kedaulatan negara melalui operasi pertahanan, pengamanan wilayah, dan penangkalan ancaman.	Kedaulatan negara menjadi prioritas utama dalam tugas TNI.	Tugas pokok
Bagaimana mekanisme pengangkatan perwira TNI?	Pengangkatan perwira TNI dilakukan melalui sistem pendidikan militer dan promosi berdasarkan prestasi dan kompetensi.	Sistem karier TNI dibangun berdasarkan merit dan profesionalisme.	ketentuan peraturan

Dalam tabel 1, setiap baris menunjukkan pertanyaan yang diajukan, jawaban yang diambil dari teks yang relevan, konteks yang digunakan untuk memberikan jawaban, dan jawaban prediksi yang dihasilkan oleh model QA. Hasil yang terlihat di gambar ini menunjukkan bagaimana sistem dapat memberikan jawaban yang relevan berdasarkan paragraf yang ditemukan, yang sesuai dengan pertanyaan yang diberikan.

- Langkah selanjutnya evaluasi model menggunakan *BERTScore* dari tabel sebelumnya dengan hasil evaluasi
1. *Precision*
Mengukur seberapa banyak kata yang dihasilkan oleh model yang relevan dengan referensi. Semakin tinggi *precision*, semakin banyak kata yang dihasilkan oleh model yang sesuai dengan yang ada dalam referensi. Pada Simulasi Chatbot, *precision* lebih tinggi (0.7240) dibandingkan dengan *Baseline Retrieval-QA* (0.7034). Ini menunjukkan bahwa model chatbot menghasilkan kata-kata yang lebih relevan dibandingkan model retrieval-QA.
 2. *Recall*:
Recall mengukur seberapa banyak kata dari referensi yang tercapture oleh model. Semakin tinggi *recall*, semakin banyak kata dalam referensi yang berhasil diambil oleh model. Simulasi Chatbot memiliki *recall* yang lebih rendah (0.7850) dibandingkan *Baseline Retrieval-QA* (0.7764), menunjukkan bahwa model retrieval-QA sedikit lebih baik dalam mencakup kata-kata dari referensi.
 3. *F1 score*:
F1 score adalah rata-rata harmonik antara *precision* dan *recall*, yang memberikan gambaran umum tentang kinerja model. *F1 score* yang lebih tinggi menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Simulasi Chatbot memiliki *F1 score* yang sedikit lebih tinggi (0.7529) dibandingkan *Baseline Retrieval-QA* (0.7389), yang menunjukkan bahwa secara keseluruhan, model chatbot lebih baik dalam menyeimbangkan *precision* dan *recall*.

Tabel 3. Tabel Perbandingan Hasil Evaluasi

Question	Answer	Context	Candidate
Simulasi Chatbot	0.7240	0.7850	0.7529
Baseline Retrieval-QA	0.7034	0.7764	0.7389

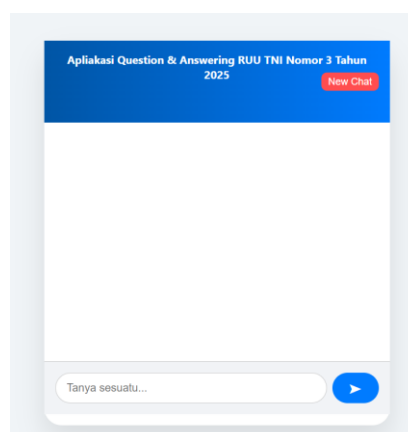
Dari tabel 3 diatas dapat disimpulkan Simulasi Chatbot memiliki *precision* yang lebih tinggi, artinya lebih sedikit kata yang tidak relevan dihasilkan oleh model ini. *Baseline Retrieval-QA* memiliki *recall* yang sedikit lebih tinggi, yang berarti model ini lebih baik dalam menangkap kata-kata dari referensi. Secara keseluruhan, Simulasi Chatbot memiliki *F1 score* yang sedikit lebih tinggi, menunjukkan bahwa model ini lebih seimbang dalam hal *precision* dan *recall* dibandingkan *Baseline Retrieval-QA*. Dengan menggunakan *BERTScore*, kita dapat memahami dengan lebih jelas kekuatan dan kelemahan masing-masing model dalam menangani tugas-tugas seperti *question-answering* dan *information retrieval*.

3.3. Model Deployment & Maintenance

Pada Deployment aplikasi disini akan menampilkan menu menu atau design interface dari sistem yang telah dibuat.

- a. Halaman Utama Aplikasi Aplikasi Question & Answering RUU TNI 2025

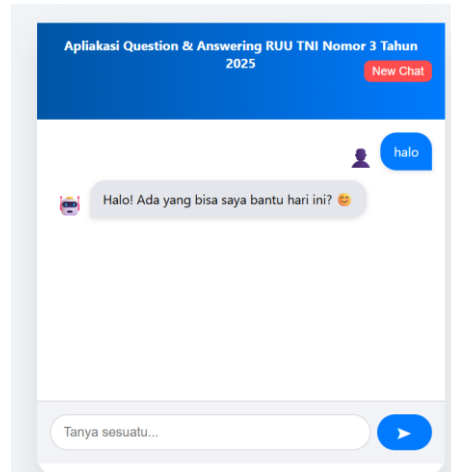
Pada halaman ini menampilkan halama kosong untuk melakukan pertanyaan dan beberapa tombol yang mana setiap tombol berfungsi sesuai kegunaannya



Gambar 2. Halaman Utama Aplikasi Aplikasi Question & Answering RUU TNI 2025

b. Halaman Tampilan Pertanyaan Dan Jawaban

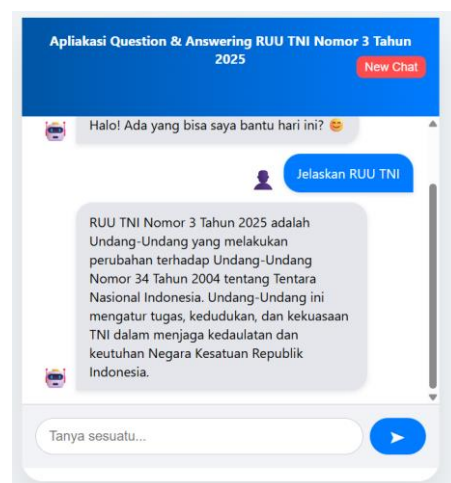
Pada halaman ini menampilkan user melakukan pertanyaan dan sistem memberikan jawaban, dimana ditampilkan ini user melakukan uji coba pertanyaan halo dan sistem menjawab sesuai dengan pertanyaan



Gambar 3. Halaman Utama Halaman tampilan pertanyaan dan jawaban

c. Halaman Tampilan Pertanyaan Dan Jawaban Kedua

Pada halaman ini menampilkan user melakukan pertanyaan dan sistem memberikan jawaban, dimana ditampilkan ini user melakukan uji coba pertanyaan tentang RUU TNI dan sistem menjawab dengan detail tentang RUU TNI 2025.



Gambar 4. Halaman Utama Halaman tampilan pertanyaan dan jawaban Kedua

Selanjutnya Setelah aplikasi di-deployment, pemeliharaan yang berkelanjutan sangat penting untuk menjaga agar aplikasi tetap berfungsi dengan baik dan relevan dengan kebutuhan pengguna. Pemeliharaan aplikasi meliputi beberapa langkah sebagai berikut:

a. Pemantauan Kinerja Aplikasi

Melakukan pemantauan terus-menerus terhadap performa aplikasi untuk memastikan bahwa aplikasi berjalan lancar. Hal ini meliputi uji coba pertanyaan, pengelolaan penggunaan CPU/GPU, dan pengukuran waktu respons untuk memastikan aplikasi dapat memberikan jawaban yang cepat dan efisien.

b. Pembaruan Model dan Databas

Secara berkala, memperbarui dataset dan melakukan fine-tuning model untuk memastikan aplikasi dapat memberikan jawaban yang akurat dan relevan. Penambahan fitur baru, seperti pencarian atau analisis statistik, juga dapat meningkatkan fungsionalitas aplikasi.

c. Pemeliharaan Sistem Backend

Pembersihan data yang tidak diperlukan, pembaruan API, dan peningkatan keamanan dilakukan untuk menjaga kinerja sistem dan melindungi aplikasi dari risiko keamanan.

- d. Mengatasi Bug dan Masalah Pengguna
Tim pemeliharaan harus siap menangani masalah yang dilaporkan pengguna dan melakukan update bugfix untuk memperbaiki kesalahan teknis agar aplikasi tetap optimal.
- e. Penyempurnaan Antarmuka Pengguna (UI)
Mengumpulkan feedback pengguna untuk meningkatkan pengalaman mereka, serta memastikan desain UI responsif untuk berbagai perangkat seperti desktop, mobile, dan tablet.
- f. Backup dan Pemulihan Data
Melakukan backup data secara berkala dan memastikan aplikasi memiliki prosedur pemulihan data untuk menghindari kehilangan data yang tidak diinginkan.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan sistem Question Answering berbasis model bahasa besar BERT yang dikustomisasi khusus untuk konteks RUU TNI 2025. Sistem yang dihasilkan menunjukkan performa evaluasi yang lebih baik dibanding baseline, dengan nilai precision 0,7240, recall 0,7850, dan F1 score 0,7529, menandakan kemampuan tinggi dalam memberikan jawaban yang akurat dan relevan.

Hasil ini mengindikasikan potensi besar teknologi Large Language Model (LLM) dalam meningkatkan pemahaman masyarakat terhadap regulasi hukum yang kompleks. Sistem ini dapat diintegrasikan ke dalam platform e-Government untuk mendukung transparansi informasi dan literasi hukum publik secara lebih luas dan praktis.

Untuk pengembangan selanjutnya, penelitian disarankan memperluas dataset dengan memasukkan berbagai dokumen hukum tambahan dan referendum pertanyaan yang lebih beragam. Selain itu, integrasi metode Retrieval Augmented Generation (RAG) dan penggunaan model generatif terbaru seperti GPT-4 dapat diterapkan untuk meningkatkan kemampuan sistem dalam menjawab pertanyaan yang sangat kompleks dan kontekstual.

5. UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada DRTPM Kemditik saintek (Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat Kementerian Pendidikan Tinggi, Sains, dan Teknologi) yang telah mendanai penelitian ini. Dukungan yang diberikan sangat berarti bagi kelancaran dan keberhasilan penelitian ini, serta bagi pengembangan teknologi yang dapat bermanfaat bagi masyarakat.

DAFTAR PUSTAKA

- [1] R. Gupta, "UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 34 TAHUN 2004," *CWL Publ. Enterp. Inc., Madison*, vol. 2004, no. May, p. 352, 2004, [Online]. Available: <http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract>
- [2] J. Robinson and D. Wingate, "Leveraging Large Language Models for Multiple Choice Question Answering," *11th Int. Conf. Learn. Represent. ICLR 2023*, pp. 1–28, 2023.
- [3] Y. Gao, D. T. Baptista-Hon, and K. Zhang, "The inevitable transformation of medicine and research by large language models: The possibilities and pitfalls," *MedComm - Futur. Med.*, vol. 2, no. 2, pp. 2–3, 2023, doi: <https://doi.org/10.1002/mef2.49>.
- [4] H. Xiong, J. Bian, S. Yang, X. Zhang, L. Kong, and D. Zhang, "Natural Language based Context Modeling and Reasoning for Ubiquitous Computing with Large Language Models: A Tutorial," pp. 1–28, 2023, doi: <https://doi.org/10.48550/arXiv.2309.15074>.
- [5] J. Cui *et al.*, "Chatlaw: A Multi-Agent Collaborative Legal Assistant with Knowledge Graph Enhanced Mixture-of-Experts Large Language Model," pp. 1–30, 2023, doi: <https://doi.org/10.48550/arXiv.2309.15074>.
- [6] D. Milasanti, "Sistem Chatbot Berbasis Large Language Model (LLM) dan Retrieval Augmented Generation (RAG) Pada Artikel Ilmiah Garuda Kemendikbud," *Ayan*, vol. 15, no. 1, pp. 37–48, 2024.
- [7] Aji Bayu Permadi, Nazruddin Safaat H, Lestari Handayani, and Yusra, "Implementasi Question Answering System Tafsir Al-Azhar Menggunakan Langchain Dan Large Language Model Berbasis Chatbot Telegram," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 12, no. 1, pp. 62–69, 2024, doi: <https://doi.org/10.21063/jtif.2024.V12.1.62-69>.
- [8] P. Rivas and L. Zhao, "Marketing with ChatGPT: Navigating the Ethical Terrain of GPT-Based Chatbot Technology," *AI*, vol. 4, no. 2, pp. 375–384, 2023, doi: <https://doi.org/10.3390/ai4020019>.

-
- [9] S. Koos and S. Wachsmann, "Navigating the Impact of ChatGPT/GPT4 on Legal Academic Examinations: Challenges, Opportunities and Recommendations," *Media Iuris*, vol. 6, no. 2, pp. 255–270, 2023, doi: <https://doi.org/10.20473/mi.v6i2.45270>.
- [10] O. Palagin, V. Kaverinsky, A. Litvin, and K. Malakhov, "OntoChatGPT Information System: Ontology-Driven Structured Prompts for ChatGPT Meta-Learning," *Int. J. Comput.*, vol. 22, no. 2, pp. 170–183, 2023, doi: <https://doi.org/10.47839/ijc.22.2.3086>.
- [11] B. Meskó and E. J. Topol, "The imperative for regulatory oversight of large language models (or generative AI) in healthcare," *npj Digit. Med.*, vol. 6, no. 1, pp. 1–6, 2023, doi: <https://doi.org/10.1038/s41746-023-00873-0>.
- [12] J. Mökander, J. Schuett, H. R. Kirk, and L. Floridi, *Auditing large language models: a three-layered approach*, vol. 4, no. 4. Springer International Publishing, 2024. doi: 10.1007/s43681-023-00289-2.
- [13] A. T. U. B. Lubis, N. S. Harahap, S. Agustian, M. Irsyad, and I. Afrianty, "Question Answering System pada Chatbot Telegram Menggunakan Large Language Models (LLM) dan Langchain (Studi Kasus UU Kesehatan)," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 955–964, 2024, doi: <https://doi.org/10.57152/malcom.v4i3.1378>.
- [14] L. R. M. Gede Putra Nugraha, Lya Hulliyyatus Suadaa*, Nori Wilantika, "Pengembangan Aplikasi Chatbot dengan Large Language Model untuk Text-to-SQL Generation," *Semin. Nas. Off. Stat. 2024*, vol. 1, no. 1, pp. 831–840, 2024, doi: <https://doi.org/10.34123/semnasoffstat.v2024i1.2252>.
- [15] J. Nay, "Large Language Models as Fiduciaries: A Case Study Toward Robustly Communicating With Artificial Intelligence Through Legal Standards," *SSRN Electron. J.*, pp. 1–27, 2023, doi: <https://dx.doi.org/10.2139/ssrn.4335945>.
- [16] D. C. K. Gomathy, "Data Mining Preparation: Process, Techniques and Major Issues in Data Analysis," *Interantional J. Sci. Res. Eng. Manag.*, vol. 06, no. 11, pp. 1–6, 2022, doi: <https://doi.org/10.55041/ijrsrem16833>.
- [17] A. L. Lezama-Sánchez, M. Tovar Vidal, and J. A. Reyes-Ortiz, "Integrating Text Classification into Topic Discovery Using Semantic Embedding Models," *Appl. Sci.*, vol. 13, no. 17, pp. 1–15, 2023, doi: <https://doi.org/10.3390/app13179857>.
- [18] Q. Huang *et al.*, "Lawyer LLaMA Technical Report," pp. 1–19, 2023, doi: <https://doi.org/10.48550/arXiv.2305.15062>.
- [19] F. Yang *et al.*, "Empower Large Language Model to Perform Better on Industrial Domain-Specific Question Answering," *EMNLP 2023 - 2023 Conf. Empir. Methods Nat. Lang. Process. Proc. Ind. Track*, pp. 294–312, 2023, doi: <https://doi.org/10.18653/v1/2023.emnlp-industry.29>.
- [20] Prabowo Subiyanto, "UU No. 3 Tahun 2025," *Undang. Nomor 3 Tahun 2025 Tentang TNI*, no. 255658, 2025.