

BAB II

TINJAUAN PUSTAKA

2.1. Aplikasi LLMs dalam Domain Hukum

Studi-studi terbaru menekankan ragam penerapan LLMs dalam ranah hukum. Dalam konteks bantuan hukum publik, "LegalBot-EC" memanfaatkan model BERT-like untuk memperluas akses informasi di Ekuador, dan dilaporkan mencapai akurasi 94% [21]. Pada ranah hukum pajak dan AI governance, studi tentang GPT-4 menunjukkan bahwa model tersebut berpotensi meningkatkan efisiensi dan menekan biaya, namun hingga kini masih belum memiliki autonomous reasoning yang dapat diandalkan [22]. Sejalan dengan itu, studi terkait keamanan pangan dan kepatuhan GDPR membandingkan sejumlah model (termasuk GPT-4 dan Mixtral) guna mengotomatisasi klasifikasi persyaratan; temuannya menunjukkan bahwa BERT meraih kinerja terbaik dengan F-score 87% [23]. Secara keseluruhan, temuan-temuan tersebut mengisyaratkan bahwa walaupun LLMs meningkatkan efisiensi secara signifikan, teknologi ini tetap memerlukan pengembangan lanjutan agar mencapai otonomi yang sepenuhnya reliabel.

2.2. Tantangan dan Keterbatasan Model Generatif Saat Ini

Walaupun model besar seperti GPT-4 memiliki kemampuan zero-shot yang kuat, sifat generatif mereka kerap memicu hallucinations pada vonis maupun nomor kasus, sehingga meningkatkan risiko malpraktik hukum [18]. Selain itu, pre-training yang English-heavy dapat membawa bias budaya yang kurang sesuai untuk menafsirkan hukum Indonesia [19]. Masalah ini semakin dipersulit oleh karakteristik yurisprudensi Indonesia yang bersifat hibrida antara hukum sipil Belanda, hukum Adat, dan hukum Islam, sehingga menuntut akurasi terminologi (misalnya, 'Majelis Hakim') yang kerap tidak tercakup dalam general benchmarks.

2.3. Perbandingan Arsitektur Model Hukum: Encoder vs. Decoder

Penelitian ini menonjol dibandingkan pekerjaan sebelumnya melalui pendekatan arsitektur yang berbeda. Tidak seperti model berbasis encoder terdahulu (seperti LegalBERT dan Lawformer) yang menitikberatkan tugas klasifikasi dalam Bahasa Inggris atau Mandarin [14], [20], Legal-Case LLM yang kami ajukan memakai arsitektur decoder-only (LLaMA 3.2) yang di-instruction-tuned untuk generative reasoning. Pendekatan ini dirancang untuk menjawab tantangan spesifik pada yurisprudensi Indonesia. Secara metodologis, kami menerapkan parameter-efficient synthetic data pipeline untuk mengurangi masalah factual hallucination. Hasil penelitian menunjukkan skor ROUGE meningkat menjadi 24, dibandingkan skor 11 dan 12 yang dihasilkan oleh zero-shot baselines, sekaligus

memungkinkan reproduktifitas yang transparan sebagai model open-source pertama pada domain ini.

