

Abstrack

Penelitian ini mengembangkan Legal-Case LLM, model bahasa besar berbasis open-weight yang di-fine-tune secara khusus untuk analisis yurisprudensi Tindak Pidana Perdagangan Orang (TPPO) di Indonesia. Dataset dikurasi dari 400+ putusan Mahkamah Agung melalui scraping, validasi, ekstraksi metadata, dan konstruksi pasangan Q&A sintetis. Model Gemma 3 diadaptasi dengan teknik LoRA, dievaluasi menggunakan ROUGE, BLEU, BERTScore, dan BARTScore.

Rumusan Masalah

1. Halusinasi faktual pada model generatif skala besar terkait vonis, nomor perkara, dan pasal hukum
2. Bias linguistik akibat dominasi bahasa Inggris dalam pre-training model umum
3. Terminologi hibrida: delik, actus reus, cyber-trafficking, istilah adat & Islam
4. Format putusan formal sesuai SK KMA No. 359/KMA/SK/XII/2022 yang kaku dan spesifik

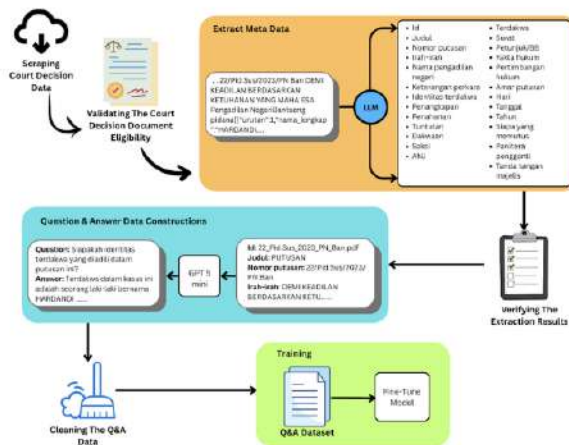
Hasil

Model	Bart	Bert	Bleu	Rouge
Gemma 3 4B	-4.37	0.889	0.374	0.555
Llama 3.2 3B	46	59	11.48	0.24
Qwen 2.5 3B	70	46	2.7	11.0
Phi-3 Mini 4K	41	44	1.84	12.0

Arsitektur & Hyperparameter

- Base Model | Gemma 3 (google/gemma-3-4b-it)
- Presisi | BF 16
- Metode | LoRA (rank=64, $\alpha=128$)
- Max Seq Len | 2048 token
- Epoch | 3 (309 steps, 822 sampel)
- Batch size | 8 (2 × grad-accum 4)
- LR | 1e-4
- Optimizer | AdamW 8-bit
- Framework | Unsloth + HuggingFace
- Hardware | GPU L4 · 24 GB VRAM

Metodologi Penelitian



Gambar 1. Diagram Alur Penelitian

Penelitian ini dilakukan dengan metode scraping dokumen putusan dari Direktori Putusan Mahkamah Agung. Dokumen yang berhasil di-scrape kemudian melalui proses validasi kelayakan. Selanjutnya, dilakukan ekstraksi metadata, yang hasilnya digunakan untuk menghasilkan dataset question-answer melalui GPT-5 mini. Dataset tersebut dibersihkan terlebih dahulu sebelum digunakan untuk fine-tuning model base Gemma 3.

Kesimpulan

Fine-tuning domain-specific pada Gemma 3 menghasilkan peningkatan signifikan: BERTScore 0.889 dan ROUGE 0.555 – melampaui semua baseline SLM open-weight terkini. Model membuktikan bahwa pendekatan parameter-efficient mampu menjembatani kesenjangan antara AI general-purpose dengan kompleksitas bahasa hukum Indonesia yang hibrida. Penerapan tetap memerlukan human-in-the-loop untuk memastikan keandalan hukum.

Tabel 1. Hasil evaluasi kuantitatif

Dataset		
408 Dokumen Awal	404 Dokumen Valid	1.019 Pasangan Q&A
260 Putusan 2023	148 Putusan 2024	80:20 Train:Test

Sumber: Direktori Putusan Mahkamah Agung RI Diterbitkan di Mendeley Data (CC BY 4.0) Anonimisasi sesuai UU No.27/2022 & SK KMA No.359/2022