

The Capability of Logical Reasoning of LLM in Bahasa Indonesia Dataset

Course: TA

Faculty of Computer Science, UI

September 4, 2025

1 Bab 1 : Pendahuluan

- Latar Belakang
- Rumusan Masalah
- Tujuan Penelitian
- Batasan Penelitian
- Manfaat Penelitian

- Model Bahasa Besar (Large Language Models / LLM) telah menunjukkan kemampuan memadai dalam berbagai tugas bahasa alami, termasuk tanya jawab, summarization, dan penalaran.
- Salah satu area penelitian penting adalah **penalaran logis** — kemampuan model untuk menarik kesimpulan yang konsisten dari premis berdasar aturan logika.
- Kesenjangan riset: apakah LLM (termasuk versi open-source dan versi ringan) mempertahankan kemampuan penalaran ketika berhadapan dengan dataset berbahasa Indonesia yang mungkin berbeda struktur linguistik dan budaya?
- Penelitian ini menguji dan membandingkan kemampuan penalaran logis beberapa LLM pada dataset berbahasa Indonesia, khususnya LLM yang dikembangkan di Indonesia itu sendiri, serta mengevaluasi efektivitas metode pipeline (mis. kerangka Aristotle) untuk meningkatkan hasil.

Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah penelitian ini adalah:

- ➊ Sejauh mana LLM berperforma dalam tugas penalaran logis pada dataset berbahasa Indonesia?
- ➋ Bagaimana perbandingan performa antara model open-source berparameter rendah (low-resource) dengan model yang lebih besar dalam konteks penalaran logis Bahasa Indonesia?
- ➌ Seberapa efektif implementasi pipeline *logical reasoning framewrok* dalam meningkatkan kinerja penalaran pada dataset berbahasa Indonesia?
- ➍ Apa tantangan utama (mis. tokenisasi, konteks budaya, format data) yang muncul ketika memindahkan dataset/pujian penalaran dari bahasa Inggris ke bahasa Indonesia?

Tujuan Penelitian

- **Tujuan umum:** Mengevaluasi dan memahami kemampuan penalaran logis LLM pada dataset berbahasa Indonesia.
- **Tujuan khusus:**
 - ① Mengukur akurasi dan metrik terkait model-model terpilih pada tugas penalaran logis Bahasa Indonesia.
 - ② Membandingkan performa antara model low-resource dan model yang lebih besar untuk tugas penalaran, juga membandingkan performa LLM khusus bahasa Indonesia dengan LLM *general-purpose*.
 - ③ Menerapkan dan mengevaluasi pipeline penalaran (Aristotle framework) untuk melihat kontribusinya terhadap perbaikan hasil.
 - ④ Mengidentifikasi hambatan praktis dan rekomendasi pengolahan (preprocessing, tokenisasi, prompt design) untuk penelitian penalaran di Bahasa Indonesia.

- Dataset: fokus pada dataset penalaran / NLI / QA berbahasa Indonesia yang tersedia atau yang ditransformasikan dari dataset Inggris.
- Model: membandingkan beberapa LLM open-source low-resource dan beberapa model yang lebih besar bila memungkinkan (tergantung sumber daya komputasi).
- Metode: evaluasi difokuskan pada pipeline pemrosesan berbasis kerangka Aristotle (translasi, dekomposisi, pencarian komplement, penyelesaian logika).
- Evaluasi: penekanan pada metrik kuantitatif (akurasi akhir) dan analisis kasus kesalahan (qualitative error analysis).

- Menambah literatur evaluasi penalaran LLM pada bahasa selain bahasa Inggris, khususnya Bahasa Indonesia.
- Memberikan *insight* tentang kemampuan model low-resource dalam konteks penalaran logis