UNIVERSITAS INDONESIA

PENALARAN LOGIKA BERBASIS FRAMEWORK TRANSLATION-DECOMPOSITION-SEARCH-RESOLVE PADA OPEN-WEIGHT LARGE LANGUAGE MODEL DENGAN DATASET BERBAHASA INDONESIA

SKRIPSI

MIKHAEL DEO BARLI
1906350572

FAKULTAS ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER
DEPOK
BULAN TAHUN

UNIVERSITAS INDONESIA

PENALARAN LOGIKA BERBASIS FRAMEWORK TRANSLATION-DECOMPOSITION-SEARCH-RESOLVE PADA OPEN-WEIGHT LARGE LANGUAGE MODEL DENGAN DATASET BERBAHASA INDONESIA

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar
Gelar Jurusan Anda

MIKHAEL DEO BARLI
1906350572

FAKULTAS ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER
DEPOK
BULAN TAHUN

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri,
dan semua sumber baik yang dikutip maupun dirujuk
telah saya nyatakan dengan benar.

Nama	:	**Mikhael Deo Barli**
NPM	:	**1906350572**
Tanda Tangan	:	
Tanggal	:	**Tanggal Bulan Tahun**

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh	:	
Nama	:	Mikhael Deo Barli
NPM	:	1906350572
Program Studi	:	Ilmu Komputer
Judul Skripsi	:	Penalaran Logika Berbasis Framework Translation-Decomposition-Search-Resolve pada Open-Weight Large Language Model dengan Dataset Berbahasa Indonesia

Telah berhasil dipertahankan di hadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Indonesia.

DEWAN PENGUJI

Pembimbing 1	:	Ari Saptawijaya, S.Kom., M.Sc., Ph.D	( )
Penguji 1	:	Penguji Pertama Anda	( )
Penguji 2	:	Penguji Kedua Anda	( )

Ditetapkan di	:	Depok
Tanggal	:	Tanggal Bulan Tahun

KATA PENGANTAR

Template ini disediakan untuk orang-orang yang berencana menggunakan LaTeX untuk membuat dokumen tugas akhir.

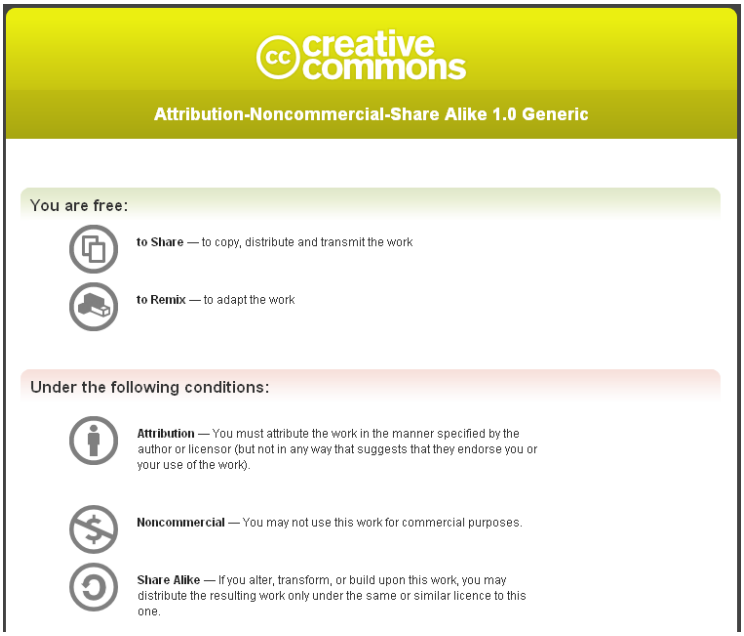
@todo

Silakan ganti pesan ini dengan pendahuluan kata pengantar Anda.

Ucapan Terima Kasih:

1. Pembimbing.
2. Dosen.
3. Instansi.
4. Orang tua.
5. Sahabat.
6. Teman.

Penulis menyadari bahwa laporan Skripsi ini masih jauh dari sempurna. Oleh karena itu, apabila terdapat kesalahan atau kekurangan dalam laporan ini, Penulis memohon agar kritik dan saran bisa disampaikan langsung melalui *e-mail* `emailanda@mail.id`.



Creative Common License 1.0 Generic

Terkait template ini, gambar lisensi di atas diambil dari `http://creativecommons.org/licenses/by-nc-sa/1.0/deed.en_CA`. Jika ingin mengentahui lebih lengkap mengenai *Creative Common License 1.0 Generic*, silahkan buka `http://creativecommons.org/licenses/by-nc-sa/1.0/legalcode`. Seluruh dokumen yang dibuat dengan menggunakan template ini sepenuhnya menjadi hak milik pembuat dokumen dan bebas didistribusikan sesuai dengan keperluan masing-masing. Lisensi hanya berlaku jika ada orang yang membuat template baru dengan menggunakan template ini sebagai dasarnya.

Penyusun template ingin berterima kasih kepada Andreas Febrian, Lia Sadita, Fahrurrozi Rahman, Andre Tampubolon, dan Erik Dominikus atas kontribusinya dalam template yang menjadi pendahulu template ini. Penyusun template juga ingin mengucapkan terima kasih kepada Azhar Kurnia atas kontribusinya dalam template yang menjadi pendahulu template ini.

Semoga template ini dapat membantu orang-orang yang ingin mencoba menggunakan LaTeX. Semoga template ini juga tidak berhenti disini dengan ada kontribusi dari para penggunanya. Jika Anda memiliki perubahan yang dirasa penting untuk disertakan dalam template, silakan lakukan *fork* repositori Git template ini di `https://gitlab.com/ichlaffterlalu/latex-skripsi-ui-2017`, lalu lakukan *merge request* perubahan Anda terhadap *branch* `master`. Kami berharap agar *template* ini dapat terus diperbarui mengikuti perubahan ketentuan dari pihak Rektorat Universitas Indonesia, dan hal itu tidak mungkin terjadi tanpa kontribusi dari teman-teman sekalian.

Depok, Tanggal Bulan Tahun

Mikhael Deo Barli

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Indonesia, saya yang bertanda tangan di bawah ini:

Nama	:	Mikhael Deo Barli
NPM	:	1906350572
Program Studi	:	Ilmu Komputer
Jenis Karya	:	Skripsi

demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Indonesia **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty Free Right*)** atas karya ilmiah saya yang berjudul:

Penalaran Logika Berbasis Framework Translation-Decomposition-Search-Resolve pada Open-Weight Large Language Model dengan Dataset Berbahasa Indonesia

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Indonesia berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok
Pada tanggal : Tanggal Bulan Tahun
Yang menyatakan

(Mikhael Deo Barli)

ABSTRAK

Nama	:	Mikhael Deo Barli
Program Studi	:	Ilmu Komputer
Judul	:	Penalaran Logika Berbasis Framework Translation-Decomposition-Search-Resolve pada Open-Weight Large Language Model dengan Dataset Berbahasa Indonesia
Pembimbing	:	Ari Saptawijaya, S.Kom., M.Sc., Ph.D

Isi abstrak.

Kata kunci:
Open-weight LLM, Penalaran logika dataset berbahasa Indonesia, *translation-decompose-search-resolve framework*

ABSTRACT

Name	:	Mikhael Deo Barli
Study Program	:	Computer Science
Title	:	Logical Inference Framework Translation-Decomposition-Search-Resolve on Open-Weight Large Language Model with Indonesian Language Dataset
Counselor	:	Ari Saptawijaya, S.Kom., M.Sc., Ph.D

Abstract content.

Key words:
Open-weight LLM, Bahasa Indonesia logical inference dataset, translation-decompose-search-resolve framework

DAFTAR ISI

HALAMAN JUDUL . . . . . **i**
HALAMAN ORISINALITAS . . . . . **ii**
LEMBAR PENGESAHAN . . . . . **iii**
KATA PENGANTAR . . . . . **iv**
LEMBAR PERSETUJUAN KARYA ILMIAH . . . . . **vi**
ABSTRAK . . . . . **vii**
DAFTAR ISI . . . . . **ix**
DAFTAR GAMBAR . . . . . **xi**
DAFTAR TABEL . . . . . **xii**
DAFTAR KODE PROGRAM . . . . . **xii**
DAFTAR LAMPIRAN . . . . . **xiii**

1. Pendahuluan **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 3
1.3 Tujuan Penelitian 3
1.4 Batasan Penelitian 3
1.5 Manfaat Penelitian 4
1.6 Posisi Penelitian 4
1.7 Sistematika Penulisan 6

2. Landasan Teori **7**
2.1 *First-Order Logic* (FOL) 7
2.1.1 Sintaks: Alfabet dan Aturan Pembentukan 7
2.1.2 Semantik: Interpretasi dan Kebenaran 8
2.2 (Konsekuensi Logis *(Entailment)*) 8
2.3 Resolusi dan Unifikasi 8
2.3.1 *Proof by Refutation* (Pembuktian Kontradiksi) 8
2.3.2 Konversi ke *Conjunctive Normal Form* (CNF) 9
2.3.3 Aturan Resolusi Robinson dan Unifikasi 9

3. Cara Kerja Framework **10**
3.1 Desain Eksperimen 10
3.2 Instrumen Data dan Adaptasi Linguistik 11
3.2.1 Proses Adaptasi ke Bahasa Indonesia 11
3.3 Konfigurasi Model 12
3.3.1 Pemilihan Model 12
3.3.2 Parameterisasi Inferensi 12
3.4 Kerangka Kerja Aristotle 14
3.4.1 Logika Translasi (*Logical Translation*) 14
3.4.2 Dekomposer (*Decomposer*) 14
3.4.3 Penyelesai Pencarian (*Search Router*) 15
3.4.4 Resolusi (*Resolver*) 15
3.5 Prosedur Eksekusi: Pipeline Aristotle 15
3.5.1 Tahap 1: Translasi Logika (Logical Translator) 15

3.5.2 Tahap 2: Dekomposisi dan Normalisasi 16
3.5.3 Tahap 3: Pencarian dan Resolusi (Search & Resolve) 16
3.6 Teknik Evaluasi dan Analisis Data 17
3.6.1 Parsing Bertingkat (Multi-stage Parsing) 17
3.6.2 Matrix Keberhasilan 21

4. HASIL EKSPERIMEN DAN ANALISA 22
4.1 Prompt Refining 22
4.1.1 *Translation to First Order Logic* 22
4.1.2 *Decomposition to Conjunctive Normal Form* 22
4.1.3 *Search Resolve* 22
4.2 Hasil Akhir 22
4.2.1 Naive Prompting 22
4.2.2 Aristotle 22

5. PENUTUP 23
5.1 Kesimpulan 23
5.2 Saran 23

DAFTAR REFERENSI 24
DAFTAR ISTILAH 1

DAFTAR GAMBAR

Gambar 3.1. Kerangka kerja Aristotle untuk inferensi logika menggunakan LLM dan metode neuro-symbolic (Xu et al. (2025)) 14

DAFTAR TABEL

Tabel 1.1. Perbandingan Penelitian Terkait Penalaran Logis dengan LLM 5

Tabel 4.1. Hasil Eksperimen dengan Naive Prompting 22
Tabel 4.2. Hasil Eksperimen dengan Aristotle Framework 22

DAFTAR KODE PROGRAM

Kode 3.1. Konfigurasi parameter deterministik untuk meniadakan halusinasi kreatif 13
Kode 3.2. Mekanisme injeksi data poin ke dalam template prompt Aristotle 16
Kode 3.3. Regex untuk ekstraksi blok Translasi 17
Kode 3.4. Regex untuk ekstraksi blok Dekomposisi 18
Kode 3.5. Regex untuk memvalidasi langkah Resolusi 20

DAFTAR LAMPIRAN

Lampiran 1. CHANGELOG 25
Lampiran 2. Judul Lampiran 2 28

BAB 1
PENDAHULUAN

Bab ini memaparkan latar belakang, permasalahan, tujuan, batasan, manfaat, ringkasan metodologi, serta sistematika penulisan penelitian ini. Penelitian ini berfokus pada evaluasi kemampuan penalaran logis oleh *Large Language Models* (LLM) yang bersifat *open-weight* ketika bekerja pada dataset berbahasa Indonesia dengan *framework translation-decomposition-search-resolve*

1.1 Latar Belakang

Perkembangan Large Language Model (LLM) telah mendorong kemajuan signifikan pada berbagai tugas pemrosesan bahasa natural seperti penerjemahan, ringkasan, dan tanya-jawab. Namun, kemampuan LLM untuk melakukan penalaran logisyaitu melakukan inferensi yang benar dari himpunan premis dan aturan formalmasih menghadapi kendala fundamental. Evaluasi yang ketat menunjukkan bahwa model sering kali berfungsi sebagai mesin probabilistik yang menebak pola statistik alih-alih melakukan deduksi deterministik. Fenomena ini dijelaskan oleh Saparov and He (2023) dalam penelitian mereka ”Language Models Are Greedy Reasoners”, yang menunjukkan bahwa LLM sering gagal dalam inferensi multi-langkah karena mereka mengambil jalan pintas heuristik daripada mengikuti rantai logika yang valid.

Untuk mengatasi kesenjangan antara kemampuan linguistik dan logika ini, berbagai metode prompting dan arsitektur telah dikembangkan. Berikut adalah sedikit tinjauan terhadap evolusi metode penalaran logis pada LLM yang menjadi landasan penelitian ini:

1. *Naive Prompting (Implicit Reasoning)*: Pendekatan paling dasar di mana model diminta langsung menjawab kesimpulan dari premis yang diberikan (misalnya, Zero-shot). Kelemahannya adalah ”Curse of Complexity”, di mana akurasi model menurun secara eksponensial seiring bertambahnya langkah logika karena model tidak memiliki memori kerja eksternal untuk menyimpan status inferensi perantara. Saparov and He (2023)
2. *Chain-of-Thought* (CoT): Diperkenalkan oleh Wei et al. (2023), CoT mendorong model untuk menghasilkan serangkaian langkah penalaran perantara sebelum memberikan

jawaban akhir. Metode ini terbukti meningkatkan performa pada tugas aritmatika dan simbolik secara signifikan dengan mengubah pemetaan Input-Output menjadi Input-Reason1-Reason2-Output. Namun, CoT rentan terhadap propagasi kesalahan, jika satu langkah penalaran salah (halusinasi), seluruh kesimpulan akan salah karena tidak ada mekanisme verifikasi eksternal.

3. Tree-of-Thoughts (ToT): Yao et al. (2023) mengembangkan konsep CoT menjadi struktur pohon, memungkinkan model untuk mengeksplorasi berbagai jalur penalaran, melakukan lookahead, dan backtracking saat menemui jalan buntu. Meskipun lebih kuat, ToT sangat mahal secara komputasi dan masih bergantung pada intuisi probabilistik model itu sendiri untuk mengevaluasi validitas setiap cabang pemikiran.
4. *Neuro-Symbolic Approaches* (Logic-LM & SymbCoT): Untuk mencapai ketepatan logika yang strict, pendekatan *Neuro-Symbolic* mulai diadopsi. Logic-LM oleh Pan et al. (2023) menggunakan LLM hanya sebagai penerjemah masalah ke dalam kode simbolik (seperti Prolog) , yang kemudian diselesaikan oleh solver deterministik. Xu et al. (2024) kemudian mengusulkan SymbCoT yang mencoba mengintegrasikan verifikasi simbolik langsung ke dalam rantai pemikiran LLM.
5. Aristotle *Framework*: Penelitian ini berfokus pada Framework Aristotle oleh Xu et al. (2025), yang menyempurnakan pendekatan *Neuro-Symbolic* dengan arsitektur Decompose-Search-Resolve. Keunggulan utamanya adalah adanya modul Search Router yang memangkas ruang pencarian premis yang tidak relevan, serta penggunaan dua jalur pembuktian (pembuktian S dan $\neg$S) untuk meminimalkan halusinasi.

Meskipun metode-metode di atas menunjukkan hasil positif, sebagian besar penelitian dilakukan pada dataset berbahasa Inggris. Studi terhadap kemampuan penalaran LLM pada bahasa lain, termasuk Bahasa Indonesia, masih terbatas. Perbedaan struktur linguistik (seperti ambiguitas subjek dalam Bahasa Indonesia) dan kualitas tokenisasi dapat memengaruhi performa model setelah adaptasi lintas bahasa. Selain itu, penelitian sebelumnya menggunakan proprietary LLM (seperti GPT-4), sehingga perbandingan performa *apple-to-apple* pada model open-weight dengan sumber daya terbatas belum banyak dikaji.

Dari gap penelitian yang sudah disebutkan, belum ada kajian mendalam mengenai efektivitas framework Aristotle pada dataset berbahasa Indonesia menggunakan open-weight LLM. Penelitian ini menjadi penting untuk mengevaluasi apakah model seperti Sahabat-AI

(yang dilatih khusus untuk Bahasa Indonesia) atau SEA-LION (regional ASEAN) ataupun Qwen(generik) dapat mengatasi tantangan penalaran logis.

1.2 Rumusan Masalah

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

1. Bagaimana penalaran logis pada dataset berbahasa Indonesia dilakukan dengan menggunakan *framework translation-decompose-search-resolve*?
2. Bagaimana perbandingan performa *framework translation-decompose-search-resolve* untuk penalaran logis berbahasa Indonesia antar open-weight LLM berparameter rendah?
3. Bagaimana perbandingan performa *framework translation-decompose-search-resolve* dibandingkan naive prompting untuk penalaran logis berbahasa Indonesia pada open-weight LLM berparameter rendah?

1.3 Tujuan Penelitian

Berikut adalah tujuan dari penelitian sesuai dengan latar belakang dan rumusan masalah

Tujuan:

1. Mengevaluasi penalaran logis pada dataset berbahasa Indonesia dilakukan dengan menggunakan *framework translation-decompose-search-resolve*
2. Mengukur perbandingan performa *framework translation-decompose-search-resolve* untuk penalaran logis berbahasa Indonesia antar open-weight LLM berparameter rendah
3. Mengukur perbandingan performa *framework translation-decompose-search-resolve* naive prompting untuk penalaran logis berbahasa Indonesia pada open-weight LLM berparameter rendah

1.4 Batasan Penelitian

Agar fokus dan ruang lingkup terukur, penelitian ini dibatasi sebagai berikut:

- **Dataset:** Fokus pada dataset diterjemahkan ke Bahasa Indonesia, yaitu ProntoQA saja
- **Model:** Eksperimen menggunakan model open-weight yang dapat dijalankan lokal

maupun server, khususnya dengan kuantisasi. Model tersebut antara lain: Qwen2.5-7B-IT-GGUF, SEALIONv3-LLama-8B-IT-GGUF, dan SahabatAIv1-Llama-8B-IT-GGUF

- **Evaluasi:** Metrik utama adalah akurasi jawaban akhir terhadap ground truth.
- Implementasi berbasiskan pada source code yang tersedia pada repository Aristotle yang merupakan implementasi *framework translation-decompose-search-resolve* Xu et al. (2025)

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan kontribusi yang bermakna bagi berbagai pihak:

- **Bagi pengembangan ilmu pengetahuan:** Penelitian ini dapat menambah pemahaman tentang kemampuan penalaran logis LLM pada bahasa Indonesia, sebuah aspek yang masih jarang dikaji. Temuan ini dapat menjadi fondasi bagi penelitian lanjutan dalam evaluasi model bahasa pada tugas-tugas penarlan logis kompleks dalam bahasa lokal.
- **Bagi praktisi dan pengembang:** Hasil analisis perbandingan model dan efektivitas framework dapat menjadi panduan dalam memilih model open-weight yang tepat dan merancang pipeline pemrosesan bahasa untuk tugas penalaran logis berbahasa Indonesia, terutama dengan sumber daya komputasi terbatas.
- **Bagi komunitas penelitian terbuka:** Dataset ProntoQA yang diterjemahkan ke bahasa Indonesia, skrip eksperimen, serta laporan hasil penelitian akan dibagikan kepada publik. Kontribusi ini memungkinkan peneliti lain untuk mereplikasi, memvalidasi, dan melanjutkan penelitian dalam domain yang sama tanpa perlu melakukan terjemahan dan persiapan data dari awal.

1.6 Posisi Penelitian

Penelitian ini mengisi gap dalam literatur saat ini dengan menerapkan kerangka kerja *Neuro-Symbolic* pada model berparameter rendah (*low-resource*) yang dikuantisasi, khusus untuk Bahasa Indonesia.

Peneliti (Tahun)	**Metode / Framework**	**Model / Bahasa / Dataset**	**Keterbatasan (Gap)**
Saparov & He (2022)	Naive Prompting & CoT	GPT-3 (175B) Inggris ProntoQA (Eng)	Hanya mengevaluasi model proprietary *high-resource*, rentan terhadap *greedy reasoning*.
Pan et al. (2023)	Logic-LM (Translate-Execute)	GPT-3.5 / GPT-4 Inggris ProofWriter, ProntoQA	Bergantung pada solver eksternal tanpa mekanisme *search* untuk memangkas premis tidak relevan.
Xu et al. (2025)	Aristotle (Translastion-Decompose-Search-Resolve)	Llama-2 / GPT-4 Inggris LogicNLI, ProntoQA, ProofWriter	Framework yang efektif, tetapi belum diuji pada model LLM *low-resource* dengan dataset berbahasa Indonesia.
Koto et al. (2023)	Evaluasi Benchmark	IndoGPT, XGLM Indonesia IndoMMLU, IndoNLI	Fokus pada evaluasi pengetahuan umum, bukan penalaran logika formal yang membutuhkan multi-langkah penyelesaian.
Penelitian Ini (2025)	Aristotle (Translastion-Decompose-Search-Resolve)	Qwen2.5, SEA-LION, Sahabat-AI Indonesia ProntoQA-ID	Menguji efektivitas *translastion-decompose-search-resolve* pada model *open-weight* kecil dan terkuantisasi.

Tabel 1.1: Perbandingan Penelitian Terkait Penalaran Logis dengan LLM

1.7 Sistematika Penulisan

Sistematika penulisan laporan adalah sebagai berikut:

- Bab 1 PENDAHULUAN

  Menguraikan motivasi penelitian, kesenjangan dalam literatur saat ini mengenai penalaran LLM bahasa Indonesia, rumusan masalah, dan tujuan spesifik validasi *framework* Aristotle.
- Bab 2 LANDASAN TEORI

  Menjelaskan prinsip-prinsip First-Order Logic (FOL) dan aturan inferensi berdasarkan literatur klasik (Rosen, Robinson), serta tinjauan mendalam mengenai paradigma *Neuro-Symbolic* dan evolusi dari Naive Prompting ke Aristotle.
- Bab 3 FRAMEWORK *TRANSLATION-DECOMPOSITION-SEARCH-RESOLVE*

  Mendeskripsikan desain eksperimen, proses adaptasi dataset ProntoQA ke Bahasa Indonesia, konfigurasi model, dan implementasi teknis *pipeline Translation-Decompose-Search-Resolve.*
- Bab 4 HASIL EKSPERIMEN DAN ANALISA

  Menyajikan data hasil akurasi antara Qwen, SEA-LION, dan Sahabat-AI. Bab ini akan menganalisis permasalahan pada tiap tahap *framework* serta membandingkan hasilnya dengan metode Naive Prompting.
- Bab 5 PENUTUP

  Merangkum temuan utama mengenai penggunaan model open-weight LLM untuk tugas logika formal dan memberikan rekomendasi untuk penelitian selanjutnya di bidang penalaran logis pada bahasa Indonesia.

BAB 2
LANDASAN TEORI

Bab ini membangun kerangka teoretis yang mendasari analisis kemampuan *Large Language Models* (LLM) dalam melakukan penalaran logika. Pembahasan mencakup prinsip formal dari *First-Order Logic* (FOL), prosedur konversi ke *Conjunctive Normal Form* (CNF), serta algoritma Resolusi yang menjadi mesin utama dalam *framework* Aristotle.

2.1 *First-Order Logic* (FOL)

First-Order Logic (FOL), atau Kalkulus Predikat, adalah sistem formal yang memperluas logika proposisi dengan memperkenalkan variabel, fungsi, dan kuantor untuk merepresentasikan objek dan relasi di dunia nyata.

2.1.1 Sintaks: Alfabet dan Aturan Pembentukan

Mengacu pada definisi standar dalam buku Discrete Mathematics and Its Applications karya Kenneth H. Rosen Rosen (2012), sintaks FOL dibangun dari komponen-komponen berikut:

- Simbol Logis:
  - Konektif: $\neg$ (Negasi), $\wedge$ (Konjungsi), $\vee$ (Disjungsi), $\rightarrow$ (Implikasi), $\leftrightarrow$ (Bikondisional).
  - Kuantor: $\forall$ (Universal), $\exists$ (Eksistensial).
  - Variabel: $x, y, z, \ldots$
- Simbol Non-Logis:
  - Konstanta: Simbol yang merepresentasikan objek spesifik (misalnya, ”Alice”, ”42”).
  - Fungsi ($f(x_1, \ldots, x_n)$): Memetakan objek ke objek lain. Contoh: $AyahDari(Budi)$.
  - Predikat ($P(x_1, \ldots, x_n)$): Fungsi yang memetakan tuple objek ke nilai kebenaran (True/False). Contoh: $Suka(Budi, Apel)$.
- Grammar (Tata Bahasa):
  - Term: Sebuah variabel, konstanta, atau fungsi yang diterapkan pada term lain.
  - Formula Atomik: Predikat yang diterapkan pada term. Ini adalah unit terkecil yang memiliki nilai kebenaran.
  - *Well-Formed Formulas* (WFFs): Formula yang disusun secara rekursif dari formula

atomik menggunakan konektif dan kuantor. Rosen menekankan pentingnya cakupan kuantor (scope), di mana variabel dalam cakupan kuantor disebut bound variable (terikat), sedangkan yang di luar adalah free variable (bebas).

2.1.2 Semantik: Interpretasi dan Kebenaran

Kebenaran sebuah kalimat FOL ditentukan oleh sebuah Interpretasi (I) atas Domain ($\mathcal{D}$) yang tidak kosong. Interpretasi memetakan:

- Konstanta ke elemen di $\mathcal{D}$.
- Memetakan predikat n-ary ke relasi n-ary di $\mathcal{D}$.
- Sebuah formula α dikatakan Benar di bawah interpretasi I (ditulis $I \models \alpha$) jika fakta di dunia nyata sesuai dengan struktur kalimat tersebut.

2.2 (Konsekuensi Logis *(Entailment)*)

Tujuan utama sistem berbasis pengetahuan adalah menarik kesimpulan baru dari premis yang ada. Konsep ini diformalkan sebagai Entailment ($KB \models \alpha$). Menurut Huth and Ryan (2004), $KB \models \alpha$ berlaku jika dan hanya jika untuk setiap interpretasi di mana KB bernilai benar, α juga pasti bernilai benar.Dalam konteks komputasi, memeriksa semua interpretasi adalah mustahil. Oleh karena itu, kita menggunakan Inference Rules (aturan inferensi) sintaksis untuk membuktikan validitas tanpa memeriksa semantik satu per satu.

2.3 Resolusi dan Unifikasi

Metode inferensi utama yang digunakan dalam modul ”Resolver” Aristotle adalah Resolusi. Teknik ini diperkenalkan oleh J.A. Robinson pada tahun 1965 dalam makalah seminalnya ”A Machine-Oriented Logic Based on the Resolution Principle”.Robinson (1965)

2.3.1 *Proof by Refutation* (Pembuktian Kontradiksi)

Robinson membuktikan bahwa untuk menguji $KB \models \alpha$, kita cukup membuktikan bahwa himpunan $KB \cup \{\neg\alpha\}$ adalah Unsatisfiable (tidak mungkin benar bersamaan). Jika kita dapat menurunkan klausa kosong ($\square$ atau False) dari himpunan tersebut, maka α terbukti benar.Robinson (1965)

2.3.2 Konversi ke *Conjunctive Normal Form* (CNF)

Agar aturan resolusi dapat diterapkan, formula logika harus diubah ke bentuk standar yang disebut *Conjunctive Normal Form* (CNF). Menurut Fitting (1996) merinci langkah-langkah algoritma konversi ini sebagai berikut:

1. Eliminasi Implikasi: Ubah $A \rightarrow B$ menjadi $\neg A \vee B$.
2. Geser Negasi ke Dalam: Gunakan hukum De Morgan dan aturan $\neg \forall x P \equiv \exists x \neg P$.
3. Standardisasi Variabel: Ubah nama variabel agar unik untuk setiap kuantor (misal: $\forall x Px \vee \forall x Qx$ menjadi $\forall x Px \vee \forall y Qy$).
4. Prenex Normal Form: Pindahkan semua kuantor universal ($\forall$) ke depan formula.
5. Skolemisasi: Menghilangkan kuantor eksistensial ($\exists$). Variabel y yang terikat oleh $\exists$ diganti dengan Fungsi Skolem fx yang bergantung pada variabel universal sebelumnya.Contoh: $\forall x \exists y Parent x, y$ menjadi $\forall x Parent x, fx$.
6. Distribusi & CNF: Gunakan aturan distributif untuk mendapatkan bentuk konjungsi dari klausa (AND of ORs).

2.3.3 Aturan Resolusi Robinson dan Unifikasi

Prinsip Resolusi Robinson menyederhanakan berbagai aturan inferensi klasik (seperti Modus Ponens dan Modus Tollens) menjadi satu aturan tunggal:

$$\frac{C_1 \vee L_1, \quad C_2 \vee L_2}{C_1 \vee C_2 \theta}$$

Di mana: L_1 dan L_2 adalah literal yang dapat dibuat saling berlawanan (komplemen). θ adalah substitusi variabel yang dihasilkan oleh algoritma Unifikasi.

Unifikasi adalah proses mencari Most General Unifier (MGU) θ yang membuat dua atom sintaksis menjadi identik. Contoh: Unifikasi Px dan $\neg PAlex$ menghasilkan $\theta = \{xAlex\}$. Tanpa unifikasi, resolusi pada FOL tidak mungkin dilakukan karena variabel pada premis yang berbeda harus diselaraskan.

BAB 3
CARA KERJA FRAMEWORK

Bab ini menguraikan rancangan operasional yang digunakan untuk menjawab rumusan masalah penelitian. Metodologi ini disusun untuk menguji secara empiris apakah pendekatan dapat mengatasi kelemahan penalaran probabilistik pada *open-weight Large Language Models* (LLM) dalam konteks Bahasa Indonesia.

Pendekatan penelitian ini bersifat kuantitatif-eksperimental. Fokus utamanya adalah mengukur akurasi penalaran logis yang diperoleh melalui manipulasi struktur *prompting* (dari Naive ke Neuro-Symbolic) dan dampaknya ketika model dijalankan dengan sumber daya terbatas (melalui kuantisasi).

3.1 Desain Eksperimen

Eksperimen dirancang dengan membandingkan performa model dalam dua skenario inferensi:

1. **Skenario Pertama (*Naive Prompting*):** Pada skenario ini, model diuji kemampuan ”intuisi”-nya. Masalah logika diberikan dalam bentuk narasi langsung dengan instruksi standar dan beberapa contoh pengerjaan (*few-shot*). Skenario ini merepresentasikan cara penggunaan LLM pada umumnya, di mana model dipaksa melakukan logika secara implisit di dalam *hidden states*-nya.
2. **Skenario Kedua (*Framework Aristotle*):** Skenario ini menerapkan (*pipeline*) yang memaksa model untuk mengeksternalisasi proses berpikirnya ke dalam simbol-simbol logika formal. Sesuai landasan teori, proses ini dipecah menjadi tahapan sekuensial: Translasi $\rightarrow$ Dekomposisi $\rightarrow$ Pencarian Bukti $\rightarrow$ Resolusi.

Struktur variabel penelitian didefinisikan sebagai berikut:

- **Variabel Bebas (*Independent Variables*):**
  - **Model:** Tiga kategori model yang mewakili spektrum pemahaman bahasa: Global (Qwen), Regional (SEA-LION), dan Nasional (SahabatAI).
  - **Metode Prompting:** *Naive* vs *Aristotle Framework*.
  - **Tingkat Presisi:** *Unquantized* (FP16/BF16) vs Terkuantisasi (4-bit GGUF)

- **Variabel Terikat (*Dependent Variables*):**
  - **Akurasi (*Accuracy*):** Ketepatan hasil akhir (Benar/Salah) yang diverifikasi terhadap *ground truth*.
  - **Kepatuhan Format (*Parsability*):** Kemampuan model menghasilkan sintaks logika (FOL/CNF) yang valid secara komputasi. Kegagalan menghasilkan format yang benar dianggap sebagai kegagalan penalaran.
- **Variabel Kontrol:** Untuk memastikan bahwa variasi hasil murni disebabkan oleh variabel bebas, seluruh eksperimen dijalankan secara deterministik dengan parameter yang tetap. Hal ini meniadakan faktor (*randomness*) dalam resolusi inferensi.

3.2 Instrumen Data dan Adaptasi Linguistik

Penelitian ini menggunakan dataset **ProntoQA** (*Prompting with Ontologies for QA*). Pemilihan dataset ini didasarkan pada kebutuhan untuk menguji kemampuan LLM dalam deduksi sintetis, terlepas dari pengetahuan dunia nyata dari LLM tersebut. ProntoQA menggunakan ontologi fiktif (misalnya: *"Setiap Wumpus adalah Jompus"*), sehingga model tidak dapat menjawab dengan mengandalkan hafalan dalam *pre-training*.

3.2.1 Proses Adaptasi ke Bahasa Indonesia

Proses penerjemahan dilakukan secara otomatis dengan *prompting* dan memeberikan beberapa contoh dalam proses tranlasinya. Penerjemahan dilakukan dengan *open-weight* LLM yaitu **Gemma-SEA-LION-v3-9B-IT**. Model dipilih karena kemampuan yang baik dalam translasi dari bahasa Inggris ke bahasa Indonesia sesuai dengan SEA-HELM *leaderboard*

- **Konsistensi Kuantor:** Frasa "Every X is Y" diterjemahkan menjadi "Setiap X adalah Y" atau "Semua X adalah Y" untuk menjaga pola *Universal Quantifier* ($\forall$).
- **Konsistensi Negasi:** Struktur "X is not Y" dipetakan menjadi "X bukan Y" atau "X tidak Y".
- **Konsistensi Plural:** Subjek seperti "Wumpuses", "Jompuses", dan subjek lainnya, tidak diterjemahkan imbuhannya untuk menjaga struktur kalimat.

3.3 Konfigurasi Model

Mengingat bahwa adanya batasan sumber daya komputasi, maka eksperimen dijalankan menggunakan pustaka `llama.cpp` Gerganov (2023) yang dioptimalkan untuk inferensi model.

3.3.1 Pemilihan Model

Model yang dipilih mewakili tiga tingkatan spesialisasi bahasa untuk menguji hipotesis bahwa pemahaman dalam bahasa lokal membantu proses dekomposisi logika:

1. **Qwen2.5-7B-Instruct (Representasi Global):** Model ini dipilih karena performanya yang unggul dalam benchmark logika matematika global, menjadi titik ukur batas atas kemampuan model *open-weight* saat ini.
2. **SEA-LION-v3-Llama-8B-Instruct (Representasi Regional):** Model yang telah melalui *continued pre-training* dengan data Asia Tenggara, diharapkan memiliki pemahaman semantik yang lebih baik terhadap struktur kalimat regional.
3. **SahabatAI-v1-Llama-8B-Instruct (Representasi Nasional):** Model yang dikhususkan untuk Bahasa Indonesia, juga sudah melalui *continued pre-training*, digunakan untuk melihat apakah spesialisasi bahasa yang mendalam dapat menyelesaikan penalaran lebih baik dalam tugas logika.

3.3.2 Parameterisasi Inferensi

Setiap model dijalankan dengan pengaturan parameter yang sama / konsisten untuk memastikan hasil eksperimen mencerminkan pengaruh variabel bebas, bukan perbedaan konfigurasi teknis. Strategi deterministik diterapkan untuk menghilangkan unsur *randomness* dalam proses inferensi. Parameter utama yang dikontrol adalah:

- ***Temperature*:** Ditetapkan ke 0.0 sehingga model hanya memilih token dengan probabilitas tertinggi tanpa penambahan noise acak (*greedy decoding*). Hal ini memastikan setiap prompt menghasilkan output yang identik.
- ***Max Tokens*:** Dibatasi sesuai dengan panjang output yang wajar untuk setiap tahap dalam pipeline, mencegah generasi yang berlebihan. Adapun token yang dibutuhkan untuk setiap tahap inferensi dalam *framework* antara lain: proses translasi ke FOL memerlukan output setidaknya 400 token, proses dekomposisi setidaknya memerlukan

output 700 token, dan proses pencarian bukti setidaknya memerlukan 1000 token., sehingga ***max tokens*** yang di set untuk semua proses adalah 2500 untuk mengatasi *overhead* atau halusinasi jika model tidak secara *strict* mengikuti format.

- ***n_ctx*:** Ukuran konteks diatur untuk menentukan berapa banyak token sebelumnya yang dapat dipertimbangkan model saat menghasilkan respons. Nilai yang lebih besar memungkinkan model memahami konteks yang lebih panjang, namun meningkatkan penggunaan memori. ***n_ctx*** di set ke 0, sehingga ukuran konteks akan sesuai dengan kemampuan model masing-masing.
- ***n_gpu_layers*:** Menentukan jumlah layer model yang dijalankan di GPU untuk akselerasi. Nilai yang lebih tinggi mempercepat inferensi tetapi memerlukan VRAM yang lebih besar. Nilai -1 berarti semua komputasi dilakukan di CPU.

Konfigurasi lengkap ini disajikan dalam kode berikut:

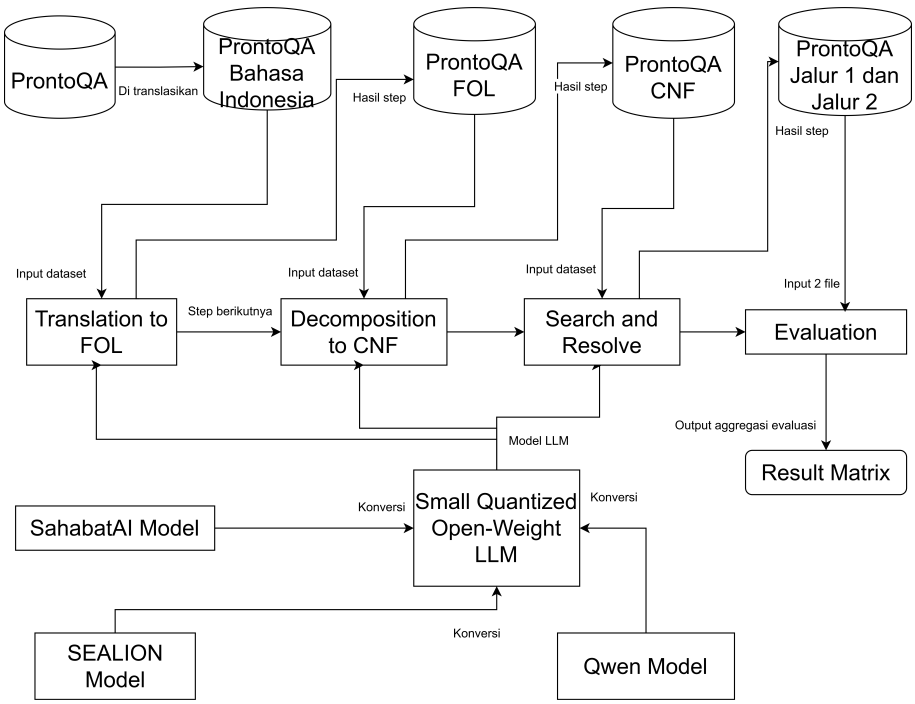
```
class LlamaCPPBackend:
    def __init__(self, local_model_path: str):
        """
        local_model_path: Must point to a specific .gguf FILE, not just a directory.
        """
        if not LLAMACPP_AVAILABLE: raise ImportError("llama-cpp-python not installed. Run 'pip install llama-cpp-python'")

        # n_gpu_layers=-1 means offload ALL layers to GPU
        self.llm = Llama(
            model_path=local_model_path,
            n_ctx=0,
            n_gpu_layers=-1,
            verbose=False
        )

    def generate(self, prompt: str, max_new_tokens: int = 512, temperature: float = 0.0, **kwargs) -> str:
        output = self.llm(
            prompt,
            max_tokens=max_new_tokens,
            stop=[],
            echo=True, # Return prompt + completion to match others
            temperature=temperature
        )
        return output['choices'][0]['text']
```

Kode 3.1: Konfigurasi parameter deterministik untuk meniadakan halusinasi kreatif

3.4 Kerangka Kerja Aristotle



Gambar 3.1: Kerangka kerja Aristotle untuk inferensi logika menggunakan LLM dan metode neuro-symbolic (Xu et al. (2025))

Penelitian ini mengimplementasikan *framework* **Aristotle** (Xu et al. (2025)) yang membagi proses inferensi menjadi empat modul utama, sesuai dengan gambar 3.1.

3.4.1 Logika Translasi (*Logical Translation*)

Modul ini menggunakan LLM *open-weight* untuk menerjemahkan premis bahasa natural menjadi FOL

- **Input:** Data poin dalam bahasa Indonesia
- **Proses:** Menerjemahkan data poin menjadi 3 bagian, sesuai dengan formula, yaitu premis dari konteks $\rightarrow$ fakta dan aturan, pertanyaan dari konteks $\rightarrow$ konjektur

3.4.2 Dekomposer (*Decomposer*)

Modul ini menggunakan LLM *open-weight* untuk dekomposisi aturan yang sudha berbentuk FOL ke dalam PNF, CNF, dan Skolemisasi

- **Input:** data poin yang sudah melewati Logical Translation dalam bentuk FOL

- **Proses:** Dekomposisi data poin yang sudha menjadi FOL ke dalam bentuk NF, lalu Skolemisasi, dan terakhir CNF.

3.4.3 Penyelesai Pencarian (*Search Router*)

Mesin inti inferensi terdiri dari dua sub-komponen:

1. ***Logical Search Router*:** Memilih klausa mana yang relevan untuk diproses. Hal ini mencegah terjadinya ledakan kombinatorial yang mengakibatkan proses komputasi yang lama dan pencarian bukti yang sesuai dengan pembuatan set untuk menyimpan premis yang sudah ditelusuri dan memberi batasan banyaknya ronde dalam pencarian bukti
2. ***Logical Resolver*:** Menerapkan aturan resolusi secara iteratif hingga ditemukan kontradiksi (untuk pembuktian salah) atau hingga ruang pencarian habis.

3.4.4 Resolusi (*Resolver*)

Sistem memverifikasi sebuah hipotesis S melalui dua jalur paralel dengan menerapkan pembuktian dengan kontradiksi (*Proof By Contradiction*)

- **Jalur 1 :** Asumsikan S. Jika ditemukan kontradiksi, maka S adalah **Benar**.
- **Jalur 2 :** Asumsikan $\neg S$ (Negasi S). Jika ditemukan kontradiksi, maka S adalah **Benar**.

3.5 Prosedur Eksekusi: Pipeline Aristotle

Berbeda dengan *Naive Prompting* yang hanya berupa satu kali pemanggilan (*single-turn*), implementasi kerangka kerja Aristotle menuntut beberapa pemanggilan model (*chain-of-calls*).

3.5.1 Tahap 1: Translasi Logika (Logical Translator)

Langkah pertama bertujuan memetakan kalimat Bahasa Indonesia yang sering kali implisit (tanpa penanda waktu atau subjek yang jelas) menjadi representasi *First Order Logic* (FOL). Model diberikan instruksi untuk mengidentifikasi:

- **Fakta Atomik:** Pernyataan spesifik tentang entitas (misal: $Pahit(Bernard, False)$).
- **Aturan Implikasi:** Hubungan sebab-akibat (misal: $Dermawan(x, False)$ ¿¿¿

$Murni(Axel, False)$).

3.5.2 Tahap 2: Dekomposisi dan Normalisasi

Pada tahap ini, output FOL diproses lebih lanjut untuk memenuhi standar mesin pembukti teorema. Model diinstruksikan untuk melakukan konversi ke *Prenex Normal Form* (PNF) dan kemudian ke *Conjunctive Normal Form* (CNF). Proses penting di sini adalah **Skolemisasi**, yaitu penghapusan kuantor eksistensial ($\exists$) yang sering menjadi sumber ambiguitas bagi model bahasa. Dalam dataset ProntoQA, tidak perlu dilakukan skolemisasi karena premis-premis dari dataset tersebut tidak memiliki kuantor eksistensial.

- **PNF:** Konversi FOL ke bentuk PNF, di mana semua kuantor diletakkan di awal formula, diikuti oleh matriks bebas kuantor.
- **Skolemisasi:** Penghapusan kuantor eksistensial ($\exists$) dengan mengganti variabel eksistensial dengan fungsi Skolem. Meskipun dataset *ProntoQA* tidak memerlukan langkah ini, tetpai inklusinya memastikan pipeline dapat menangani formula umum dengan kuantor campuran untuk dataset lain, seperti *ProofWriter* dan *LogicNLI*
- **CNF:** Transformasi akhir ke CNF, di mana formula diekspresikan dalam bentuk konjungsi (AND) dan disjungsi (OR) antar premis / kalimat.

3.5.3 Tahap 3: Pencarian dan Resolusi (Search & Resolve)

Ini adalah inti dari arsitektur neuro-symbolic. Alih-alih membiarkan model menebak jawaban akhir, model diminta bertindak sebagai mesin *Resolution Prover*.

1. **Clause Selection:** Model memilih premis yang relevan dari premis yang sudah didekomposisi dan disimpan ke memori.
2. **Resolution Step:** Model menerapkan aturan inferensi (seperti Modus Ponens atau Silogisme) untuk menurunkan klausa baru.
3. **Contradiction Check:** Proses berhenti ketika ditemukan kontradiksi (misal: P dan $\neg P$ hadir bersamaan), yang membuktikan bahwa asumsi awal salah / terjadinya kontradiksi.

Mekanisme injeksi instruksi ke dalam *prompt template* untuk setiap tahap diimplementasikan menggunakan kode Python berikut:

```
def construct_prompt_a(self, record, in_context_examples_trans):
    full_prompt = in_context_examples_trans
    if self.dataset_name == "LogicNLI":
```

```
        context = "\n".join(record['facts'] + record['rules'])
        question = record['conjecture']
    else:
        context = record['context']
        question = re.search(r'\?(.*)', record['question'].strip()).group(1).strip()
    full_prompt = full_prompt.replace('[[PREMISES]]', context)
    full_prompt = full_prompt.replace('[[CONJECTURE]]', question)
    return full_prompt
```

Kode 3.2: Mekanisme injeksi data poin ke dalam template prompt Aristotle

3.6 Teknik Evaluasi dan Analisis Data

Untuk menjamin objektivitas, evaluasi tidak dilakukan secara manual melainkan menggunakan sistem ekstraksi berbasis pola (*Regular Expression* / Regex).

3.6.1 Parsing Bertingkat (Multi-stage Parsing)

Sistem evaluasi dirancang untuk ”menangkap” struktur logika dari output teks model. Kegagalan model dalam mematuhi format yang diminta pada tahap apa pun akan langsung dianggap sebagai kesalahan (*failure*), tanpa upaya perbaikan manual. Ini adalah standar ketat yang diterapkan untuk menguji keandalan model sebagai komponen sistem logika.

- **Ekstraksi FOL:** Mengambil Fakta, Aturan, dan Konjektur sesuai bloknya

```
def extract_facts_rules_conjecture(self, content, context_sentence_count=None):
    # Clean invisible characters
    content = (content or "").replace('\u200b', '').replace('\ufeff', '')

    prompt_marker = re.compile(
        r'Di bawah ini(?:\s+adalah(?:\s+yang\s+perlu\s+Anda\s+terjemahkan)?)?:?',
        re.IGNORECASE
    )
    m_prompt = prompt_marker.search(content)
    search_start_pos = m_prompt.end() if m_prompt else 0

    block_header = re.compile(r'\*{0,3}Bentuk Akhir\*{0,3}', re.IGNORECASE)
    m_block = block_header.search(content, pos=search_start_pos)

    if m_block:
        area = content[m_block.end():]
    else:
        area = content[search_start_pos:]

    # Define Patterns for possible headers
    # Include the "End Markers" as a header type.
    patterns = {
```

```
        "facts": re.compile(r'(?:Fakta|Facts)\s*[:\-]?\s*', re.IGNORECASE),
        "rules": re.compile(r'(?:Aturan|Rules)\s*[:\-]?\s*', re.IGNORECASE),
        "conj": re.compile(r'(?:Konjektur|Conjecture)\s*[:\-]?\s*', re.IGNORECASE),
        "stop": re.compile(r'(?:\*{0,3}\s*Akhir Blok\s*\*{0,3}|###|```|-{3,})',
re.IGNORECASE)
    }

    def extract_section(target_key):
        start_match = patterns[target_key].search(area)
        if not start_match:
            return ""

        content_start_idx = start_match.end()

        # Find the next header
        next_indices = []
        for key, pat in patterns.items():
            m = pat.search(area, pos=content_start_idx)
            if m:
                next_indices.append(m.start())

        # If found upcoming headers, stop at the nearest one (min index).
        # If no headers found, go to end of string.
        if next_indices:
            cutoff_idx = min(next_indices)
            raw_text = area[content_start_idx:cutoff_idx]
        else:
            raw_text = area[content_start_idx:]

        return raw_text.strip()

    facts = extract_section("facts")
    rules = extract_section("rules")
    conjecture = extract_section("conj")

    return facts, rules, conjecture
```

Kode 3.3: Regex untuk ekstraksi blok Translasi

- **Ekstraksi CNF:** mengambil hasil akhir dari dekomposisi dari Aturan FOL.

```
def post_process_decompose(self, content, rules_count=None):

    content = (content or "").replace('\u200b', '').replace('\ufeff', '')

    marker_pattern = r'Di bawah ini adalah yang perlu Anda konversikan menggunakan
normalisasi.'
    marker_match = re.search(marker_pattern, content, flags=re.IGNORECASE)

    block_header = re.compile(r'\*{0,3}Bentuk Akhir\*{0,3}', re.IGNORECASE)
    m_block = block_header.search(content, pos=marker_match.end())
```

```
    area = content[m_block.end():]

    cnf_label_re = re.compile(
        r'(?:Aturan dalam CNF|Aturan CNF|Aturan|Rules)\s*[:\-]?\s*',
        flags=re.IGNORECASE
    )
    skolem_label_re = re.compile(
        r'(?:Aturan dalam Skolem|Skolemisasi|Skolem|Bentuk Akhir Setelah Skolemisasi|Skolemization)\s*[:\-]?\s*',
        flags=re.IGNORECASE
    )

    # Construct the pattern string explicitly to avoid parentheses nesting errors.
    boundary_pattern = (
        r'(?:'                                                  # Start outer group
            r'\r?\n\s*'                                         # Newline + whitespace
            r'(?:'                                              # Start inner grouping for headers
                r'(?:Aturan dalam CNF|Aturan CNF|Aturan \(CNF\)|Aturan|Rules)|'  # CNF headers
                r'(?:Skolemisasi|Skolem|Bentuk Akhir)|'         # Skolem headers
                r'(?:\*{0,3}\s*Akhir Blok\s*\*{0,3}|Final Form|###)' # End markers
            r')'                                                # End inner grouping
        r')'                                                    # End outer group
        r'|$'                                                   # OR End of String
    )

    boundary_re = re.compile(boundary_pattern, flags=re.IGNORECASE)

    def extract_after_label(label_re):
        """Finds label, returns text until next boundary."""
        lab_match = label_re.search(area)
        if not lab_match:
            return None
        start = lab_match.end()
        bound = boundary_re.search(area, pos=start)
        end = bound.start() if bound else len(area)
        return area[start:end].strip()

    cnf_raw = extract_after_label(cnf_label_re)
    skolem_raw = extract_after_label(skolem_label_re)

    def to_lines(raw):
        if not raw:
            return []
        return [ln.strip() for ln in raw.splitlines() if ln.strip()]

```

```
    cnf_lines = to_lines(cnf_raw)
    skolem_lines = to_lines(skolem_raw) if skolem_raw else None

    print(f"CNF Raw: {cnf_lines}")
    print(f"Skolem Raw: {skolem_lines}")

    return cnf_lines, skolem_lines
```

Kode 3.4: Regex untuk ekstraksi blok Dekomposisi

- **Ekstraksi Resolusi:** Mendeteksi klausa baru dan kesimpulan akhir dalam label kecukupan.

```
def post_process_logic_solver(self, response_d):
    content = response_d
    marker_pattern = r'(.*?)Dibawah ini tugas yang perlu Anda lakukan(.*?)'
    marker_match = re.search(marker_pattern, content, flags=re.IGNORECASE)
    search_area = content[marker_match.end():] if marker_match else content

    final_block_pattern = (
        r'\*{0,3}(?:Bentuk Akhir)\*{0,3}\s*'
        r'(.*?)'
        r'(?=(\*{0,3}(?:Akhir Blok)\*{0,3})|'
        r'$)' # or end of string
    )

    final_block_match = re.search(final_block_pattern, search_area, flags=re.DOTALL |
re.IGNORECASE)

    if not final_block_match:
        return [], None

    block = final_block_match.group(1)

    block_clean = block.strip()
    print(f"\n\nCHOSEN BLOCK:\n\n{block_clean}\n")
    print("END OF CHOSEN BLOCK\n\n")

    clause_pos = re.search(r'Clause\s*Baru', block_clean, flags=re.IGNORECASE)
    clause_after = block_clean[clause_pos.end():]
    m_new = re.search(r'\{(.*?)\}', clause_after, flags=re.DOTALL)

    if not m_new:
        raise ValueError(f"'Clause Baru:' with '{{...}}' not found in expected form.")

    new_clause = m_new.group(1).strip()

    label_pos = re.search(r'Label\s*Cukup', block_clean, flags=re.IGNORECASE)
    label_after = block_clean[label_pos.end():]
    m_label = re.search(r'\[(.*?)\]', label_after, flags=re.DOTALL)
    if not m_label:
```

```
38          raise ValueError(f"'Label Cukup' with '[...]' not found in expected form
    [True|False].")
39
40      sufficiency_label = m_label.group(1).strip()
41
42      return {
43          "new_clause": new_clause,
44          "sufficiency_label": sufficiency_label,
45      }
```

Kode 3.5: Regex untuk memvalidasi langkah Resolusi

3.6.2 Matrix Keberhasilan

Kinerja dari sebuah LLM diukur menggunakan ***Accuracy***. Berikut adalah matrix kebenaran dari sebuah jawaban:

$$A = \begin{cases} \text{True}, & P_n \vdash S_n \wedge P_n \nvdash \neg S_n \\ \text{False}, & P_n \nvdash S_n \wedge P_n \vdash \neg S_n \\ \text{Unknown}, & P_n \nvdash S_n \wedge P_n \nvdash \neg S_n \\ \text{Self-Contradictory}, & P_n \vdash S_n \wedge P_n \vdash \neg S_n \end{cases}$$

A merupakan output dari matrix kebenaran tersebut dan merupakan jawaban akhir dari sebuah data poin. P_n merupakan premis, S_n merupakan konjektur dan $\neg S_n$ merupakan konjektur yang dinegasi yang diinisialisasi dalam dua jalur pencarian bukti.

BAB 4
HASIL EKSPERIMEN DAN ANALISA

Bab ini menjelaskan tentang eksperimen dan hasil eksperimen dari penelitian

4.1 Prompt Refining

4.1.1 *Translation to First Order Logic*

4.1.2 *Decomposition to Conjunctive Normal Form*

4.1.3 *Search Resolve*

4.2 Hasil Akhir

4.2.1 Naive Prompting

Tabel 4.1: Hasil Eksperimen dengan Naive Prompting

	Qwen2.5 7B-Instruct-GGUF	**SEA-LION v3-Llama-8B-GGUF**	**SahabatAI v1-Llama-8B-GGUF**
Naive Prompting			
After Answer	51.40%	56.20%	61.40%
Before Answer	81.00%	76.80%	67.80%

4.2.2 Aristotle

Tabel 4.2: Hasil Eksperimen dengan Aristotle Framework

	Qwen2.5 7B-Instruct-GGUF	**SEA-LION v3-Llama-8B-GGUF**	**SahabatAI v1-Llama-8B-GGUF**
Aristotle	14.00%	81.60%	61.20%

BAB 5
PENUTUP

Pada bab ini, Penulis akan memaparkan kesimpulan penelitian dan saran untuk penelitian berikutnya.

5.1 Kesimpulan

Berikut ini adalah kesimpulan terkait pekerjaan yang dilakukan dalam penelitian ini:

1. **Poin pertama**
   Penjelasan poin pertama.
2. **Poin kedua**
   Penjelasan poin kedua.

Tulis kalimat penutup di sini.

5.2 Saran

Berdasarkan hasil penelitian ini, berikut ini adalah saran untuk pengembangan penelitian berikutnya:

1. Saran 1.
2. Saran 2.

DAFTAR REFERENSI

Fitting, M. (1996). *First-Order Logic and Automated Theorem Proving*. Springer, 2nd edition.

Gerganov, G. (2023). llama.cpp. `https://github.com/ggml-org/llama.cpp`. Accessed: December 5, 2025.

Huth, M. R. A. and Ryan, M. D. (2004). *Logic in Computer Science: Modelling and Reasoning about Systems*. Cambridge University Press, 2nd edition.

Pan, L., Albalak, A., Wang, X., and Wang, W. Y. (2023). Logic-lm: Empowering large language models with symbolic solvers for faithful logical reasoning.

Robinson, J. A. (1965). A machine-oriented logic based on the resolution principle. *J. ACM*, 12(1):23–41.

Rosen, K. H. (2012). *Discrete mathematics and its applications*. WCB/McGraw-Hill, 7th edition.

Saparov, A. and He, H. (2023). Language models are greedy reasoners: A systematic formal analysis of chain-of-thought.

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., and Zhou, D. (2023). Chain-of-thought prompting elicits reasoning in large language models.

Xu, J., Fei, H., Luo, M., Liu, Q., Pan, L., Wang, W. Y., Nakov, P., Lee, M., and Hsu, W. (2025). Aristotle: Mastering logical reasoning with A logic-complete decompose-search-resolve framework. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*.

Xu, J., Fei, H., Pan, L., Liu, Q., Lee, M.-L., and Hsu, W. (2024). Faithful logical reasoning via symbolic chain-of-thought.

Yao, S., Yu, D., Zhao, J., Shafran, I., Griffiths, T. L., Cao, Y., and Narasimhan, K. (2023). Tree of thoughts: Deliberate problem solving with large language models. *ArXiv*, abs/2305.10601.

LAMPIRAN

Lampiran 1: CHANGELOG

@todo

Silakan hapus lampiran ini ketika Anda mulai menggunakan *template*.

Template versi terbaru bisa didapatkan di `https://gitlab.com/ichlaffterlalu/latex-skripsi-ui-2017`. Daftar perubahan pada *template* hingga versi ini:

- versi 1.0.3 (3 Desember 2010):
  - *Template* Skripsi/Tesis sesuai ketentuan *formatting* tahun 2008.
  - Bisa diakses di `https://github.com/edom/uistyle`.
- versi 2.0.0 (29 Januari 2020):
  - *Template* Skripsi/Tesis sesuai ketentuan *formatting* tahun 2017.
  - Menggunakan BibTeX untuk sitasi, dengan format *default* sitasi IEEE.
  - *Template* kini bisa ditambahkan kode sumber dengan *code highlighting* untuk bahasa pemrograman populer seperti Java atau Python.
- versi 2.0.1 (8 Mei 2020):
  - Menambahkan dan menyesuaikan tutorial dari versi 1.0.3, beserta cara kontribusi ke template.
- versi 2.0.2 (14 September 2020):
  - Versi ini merupakan hasil *feedback* dari peserta skripsi di lab *Reliable Software Engineering* (RSE) Fasilkom UI, semester genap 2019/2020.
  - BibTeX kini menggunakan format sitasi APA secara *default*.
  - Penambahan tutorial untuk `longtable`, agar tabel bisa lebih dari 1 halaman dan header muncul di setiap halaman.
  - Menambahkan tutorial terkait penggunaan BibTeX dan konfigurasi *header/footer* untuk pencetakan bolak-balik.
  - Label ”Universitas Indonesia” kini berhasil muncul di halaman pertama tiap bab dan di bagian abstrak - daftar kode program.
  - *Hyphenation* kini menggunakan `babel` Bahasa Indonesia. Aktivasi dilakukan di `hype-indonesia.tex`.
  - Minor adjustment untuk konsistensi *license* dari template.
- versi 2.0.3 (15 September 2020):

  - Menambahkan kemampuan orientasi *landscape* beserta tutorialnya.
  - `\captionsource` telah diperbaiki agar bisa dipakai untuk `longtable`.
  - Daftar lampiran kini telah tersedia, lampiran sudah tidak masuk daftar isi lagi.
  - Nomor halaman pada lampiran dilanjutkan dari halaman terakhir konten (daftar referensi).
  - Kini sudah bisa menambahkan daftar isi baru untuk jenis objek tertentu (custom), seperti: ”Daftar Aturan Transformasi”. Sudah termasuk mekanisme *captioning* dan tutorialnya.
  - Perbaikan minor pada tutorial.
- versi 2.1.0 (8 September 2021):
  - Versi ini merupakan hasil *feedback* dari peserta skripsi dan tesis di lab *Reliable Software Engineering* (RSE) Fasilkom UI, semester genap 2020/2021.
  - Minor edit: ”Lembar Pengesahan”, dsb. di daftar isi menjadi all caps.
  - Experimental multi-language support (Chinese, Japanese, Korean).
  - *Support* untuk justifikasi dan word-wrapping pada tabel.
  - Penggunaan suffix ”(sambungan)” untuk tabel lintas halaman. Tambahan support suffix untuk `\captionsource`.
- versi 2.1.1 (7 Februari 2022):
  - Update struktur mengikuti fork template versi 1.0.3 di `https://github.com/rkkautsar/edom/ui-thesis-template`.
  - *Support* untuk simbol matematis `amsfonts`.
  - Kontribusi komunitas terkait improvement GitLab CI, atribusi, dan format sitasi APA bahasa Indonesia.
  - Perbaikan tutorial berdasarkan perubahan terbaru pada versi 2.1.0 dan 2.1.1.
- versi 2.1.2 (13 Agustus 2022):
  - Modifikasi penamaan beberapa berkas.
  - Perbaikan beberapa halaman depan (halaman persetujuan, halaman orisinalitas, dsb.).
  - *Support* untuk lembar pengesahan yang berbeda dengan format standar, seperti Laporan Kerja Praktik dan Disertasi.
  - Kontribusi komunitas terkait kesesuaian dengan format Tugas Akhir UI, kelengkapan dokumen, perbaikan format sitasi, dan *quality-of-life*.
  - Perbaikan tutorial.
- versi 2.1.3 (22 Februari 2023):

  - Dukungan untuk format Tugas Akhir Kelompok di Fasilkom UI.
  - Dukungan untuk format laporan Kampus Merdeka Mandiri di Fasilkom UI.
  - Minor *bugfix*: Perbaikan kapitalisasi variabel.
  - Quality-of-Life: Pengaturan kembali `config/settings.tex`.
  - Tutorial untuk beberapa *use case*.
- versi 2.2.0 (28 Agustus 2024):
  - Perbaikan format agar sesuai dengan format Tugas Akhir terbaru. Hal ini mencakup halaman judul, halaman pernyataan orisinalitas, header/footer, dan lampiran.
- versi 2.2.1 (16 Desember 2024):
  - *Bugfix*: isu *header* dan *footer* untuk halaman bolak-balik.
  - *Bugfix*: isu *auto-wrapping* pada kode yang tidak bisa berjalan sejak v2.2.0.
  - *Bugfix*: isu penomoran objek kustom yang tidak sesuai konvensi `[bab].[objek]`.
  - *Bugfix*: penomoran bab di Daftar Isi yang belum sesuai konvensi Tugas Akhir UI.
  - *Bugfix*: hal-hal lain pada *formatting* sesuai dengan permintaan dari Perpustakaan Fasilkom UI.
  - Perbaikan *formatting* untuk `landscape` dengan *library* `pdflscape`.
  - Perbaikan cara memasukkan sebuah persamaan ke dalam daftar persamaan.
  - Perbaikan penggunaan ”saya” menjadi ”kami” untuk dokumen-dokumen awal pada Tugas Akhir Kelompok.
  - Fitur baru: *Support* untuk *code highlighting* pada berbagai bahasa pemrograman yang tidak di-*support* secara *default* oleh *library* `listings`.
  - Fitur baru: *Support* untuk *glossary* (daftar istilah).
  - Perbaikan *major* pada tutorial, termasuk menampilkan contoh kode ke dalam PDF tutorial, dan pengaturan ulang subbab.

Lampiran hadir untuk menampung hal-hal yang dapat menunjang pemahaman terkait tugas akhir, namun akan mengganggu *flow* bacaan sekiranya dimasukkan ke dalam bacaan. Lampiran bisa saja berisi data-data tambahan, analisis tambahan, penjelasan istilah, tahapan-tahapan antara yang bukan menjadi fokus utama, atau pranala menuju halaman luar yang penting.

Subbab dari Lampiran 2

@todo
Isi subbab ini sesuai keperluan Anda. Anda bisa membuat lebih dari satu judul lampiran, dan tentunya lebih dari satu subbab.