



Universitas
Indonesia

Presentasi Sidang Skripsi

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

Dipresentasikan oleh :

MIKHAEL DEO BARLI

NPM 1906350572

Jurusan Ilmu Komputer

GARIS BESAR



Universitas
Indonesia

Latar Belakang

Rumusan masalah

Landasan Teori

Framework Aristotle

Translation to FOL

01

02

03

04

05

Decomposition to CNF

Search and Resolve

Desain Eksperimen

Hasil dan Analisis

Penutup

06

07

08

09

10

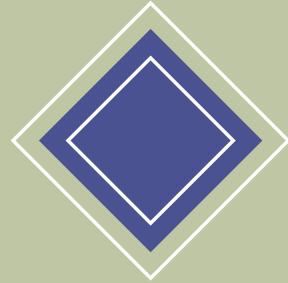


LATAR BELAKANG

1. 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 logis masih menghadapi kendala fundamental.
2. Adanya *shift* dari penggunaan LLM ke SLM karena keterbatasan sumber daya, khususnya pada negara berkembang seperti negara Indonesia, sehingga penggunaan SLM sangat diperlukan agar dapat menjalankan model secara lokal yang menjamin kedaulatan data dan privasi, serta menguntungkan secara ekonomi karena dapat mengurangi penggunaan energi yang signifikan jika SLM dibandingkan dengan LLM

Dari 2 poin tersebut, muncullah ide apakah SLM dapat melakukan penalaran logis pada dataset berbahasa Indonesia dengan *framework translation-decompose-search-resolve* pada penelitian ini



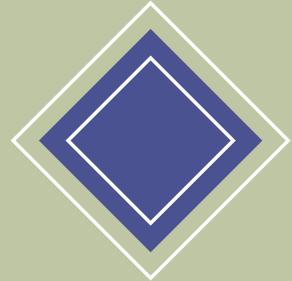


EVOLUSI PENALARAN LOGIS

METODE

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
2. ***Chain-of-Thought (CoT)***: 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)***: dikembangkan dari 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.



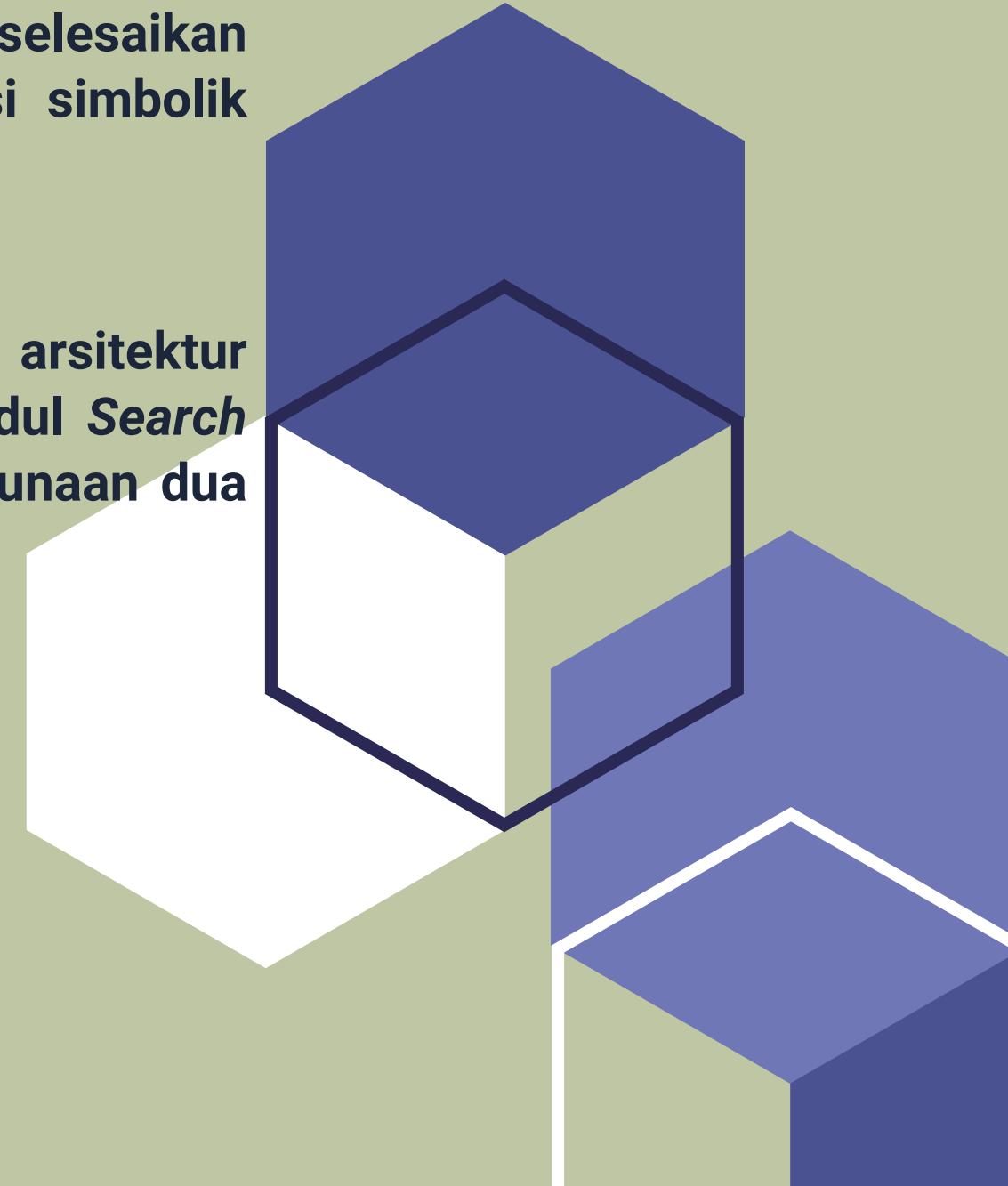


EVOLUSI PENALARAN LOGIS

METODE

4. **Neuro-Symbolic Approaches(*Logic-LM* & *SymbCoT*):** Untuk mencapai ketepatan logika yang strict, pendekatan *Neuro-Symbolic* mulai diadopsi. *Logic-LM* menggunakan LLM hanya sebagai penerjemah masalah ke dalam kode simbolik (seperti Prolog) , yang kemudian diselesaikan oleh solver deterministik. *SymbCoT* akan mencoba mengintegrasikan verifikasi simbolik langsung ke dalam rantai pemikiran LLM.

5. **Aristotle Framework:** Menyempurnakan pendekatan *Neuro-Symbolic* dengan arsitektur *Translation-Decompose-Search-Resolve*. Keunggulan utamanya adalah adanya modul *Search Router* yang memangkas ruang pencarian premis yang tidak relevan, serta penggunaan dua jalur pembuktian untuk meminimalkan halusinasi.





RUMUSAN MASALAH

Masalah Pertama

01

Bagaimana penalaran logis pada dataset berbahasa Indonesia dilakukan dengan menggunakan *framework translation-decompose-search-resolve*?

Masalah Kedua

02

Bagaimana perbandingan performa *framework translation-decompose-search-resolve* untuk penalaran logis berbahasa Indonesia antar *open-weight LLM* berparameter rendah atau SLM?

Masalah Ketiga

03

Bagaimana perbandingan performa *framework translation-decompose-search-resolve* dibandingkan naive prompting untuk penalaran logis berbahasa Indonesia pada *open-weight LLM* berparameter rendah atau SLM?



TUJUAN PENELITIAN

Tujuan Pertama

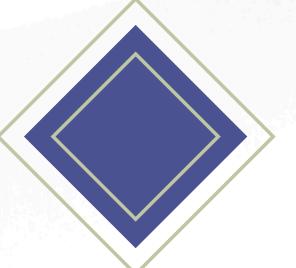
01 mengimplementasikan penalaran logis pada dataset berbahasa Indonesia dilakukan dengan menggunakan *framework translation-decompose-search-resolve*?

Tujuan Kedua

Mengevaluasi perbandingan performa *framework translation-decompose-search-resolve* untuk penalaran logis berbahasa Indonesia antar *open-weight LLM* berparameter rendah atau SLM?

Tujuan Ketiga

03 Mengevaluasi perbandingan performa *framework translation-decompose-search-resolve* dibandingkan naive prompting untuk penalaran logis berbahasa Indonesia pada *open-weight LLM* berparameter rendah atau SLM?



FIRST LOGIC

ORDER

Sintaks FOL dibangun dari komponen-komponen berikut menurut Brachman and Levesque (2004):

01

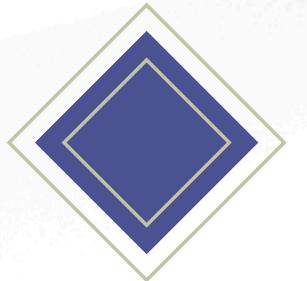
Simbol Logis

- Konektif: \neg (Negasi), \wedge (Konjungsi), \vee (Disjungsi), \rightarrow (Implikasi), \leftrightarrow (Bikondisional).
- Kuantor: \forall (Universal), \exists (Eksistensial).
- Variabel: x, y, z, \dots

02

Simbol Non-Logis

- Konstanta: Simbol yang merepresentasikan objek spesifik (misalnya, "Alice", "42").
- Fungsi ($f(x_1, \dots, x_n)$): Memetakan objek ke objek lain. Contoh: AyahDari(Budi).
- Predikat ($P(x_1, \dots, x_n)$): Fungsi yang memetakan tuple objek ke nilai kebenaran (True/False). Contoh: Ayah(Budi, True).



INTERPRETASI DAN KEBENARAN

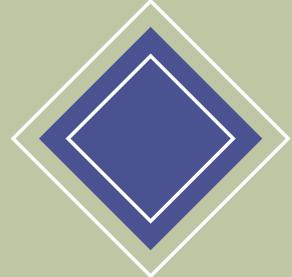
Interpretasi memetakan:

- Konstanta: Setiap simbol konstanta c dipetakan ke elemen spesifik di D.
- Predikat: Setiap simbol predikat n-ary dipetakan ke himpunan relasi n-ary di D (objek-objek mana yang memiliki sifat tersebut).
- Fungsi: Setiap simbol fungsi f dipetakan ke operasi konkret pada D. Misalnya, jika domainnya adalah bilangan bulat, simbol fungsi plus(x, y) dapat diinterpretasikan sebagai operasi penjumlahan $x + y$.



KONSEKUENSI LOGIS (*ENTAILMENT*)

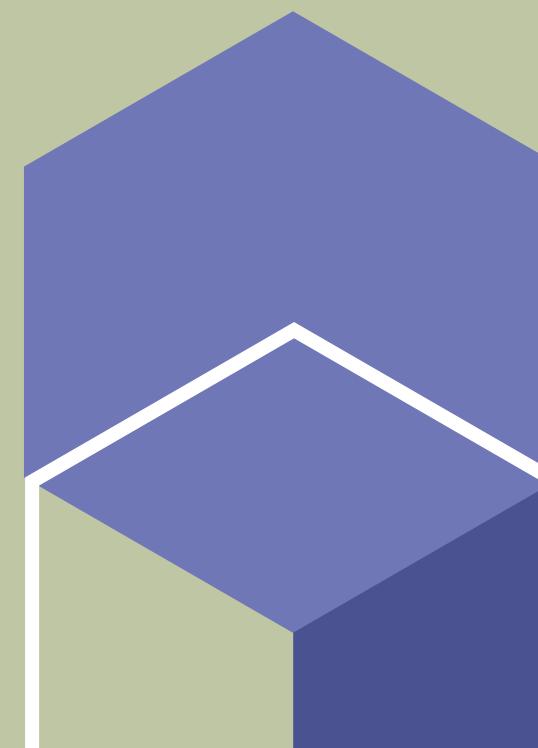
Dalam sistem logika, kumpulan fakta dan aturan yang kita yakini kebenarannya disebut sebagai Knowledge Base (KB). Menurut Huth and Ryan (2004) $KB \models a$ jika dan hanya jika untuk setiap interpretasi I di mana semua kalimat dalam KB bernilai Benar, maka a juga pasti bernilai Benar. Namun, memeriksa setiap interpretasi secara komputasi adalah mustahil karena jumlahnya bisa tak terbatas. Oleh karena itu, kita menggunakan algoritma inferensi sintaksis (seperti Resolusi) untuk membuktikan validitas tersebut secara otomatis.



RESOLUSI

Resolusi bekerja dengan prinsip Proof by Contradiction. Untuk membuktikan bahwa $KB \models a$, kita dapat menunjukkan bahwa himpunan $KB \cup \{\neg a\}$ adalah Unsatisfiable (tidak ada interpretasi yang memenuhi), maka terbukti secara logis bahwa a haruslah benar.

Tanda terjadinya kontradiksi adalah ketika algoritma berhasil menurunkan Klausu Kosong(\Box atau False) (Brachman and Levesque (2004)).





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: $\forall x(P(x)) \vee \forall x(Q(x))$ menjadi $\forall x(P(x)) \vee \forall y(Q(y))$).

KONVERSI KE CONJUNCTIVE NORMAL FORM (CNF)

4. Prenex Normal Form: Pindahkan semua kuantor ke depan formula.

5. Skolemisasi: Menghilangkan kuantor eksistensial (\exists). Variabel y yang terikat oleh \exists diganti dengan Fungsi Skolem ($f(x)$) yang bergantung pada variabel universal sebelumnya. Contoh: $\forall x \exists y (\text{Parent}(x, y))$ menjadi $\forall x (\text{Parent}(x, f(x)))$.

6. Distribusi & CNF: Gunakan aturan distributif untuk mendapatkan bentuk konjungsi dari klausa (AND of ORs).



ATURAN UNIFIKASI

RESOLUSI

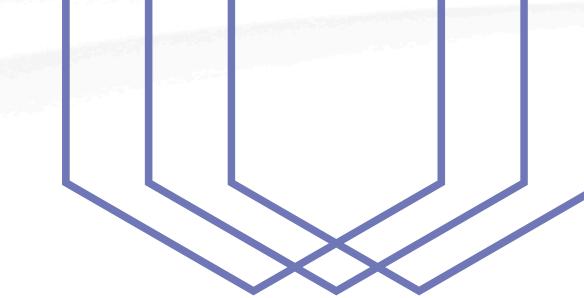
DAN

Aturan resolusi untuk logika predikat adalah:

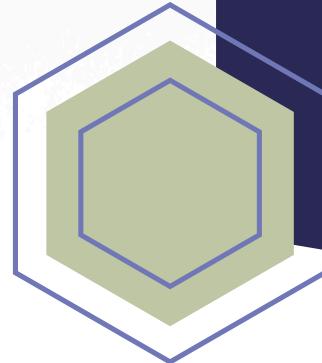
$$\frac{C_1 \vee L_1, \quad C_2 \vee L_2}{C_1 \vee C_2 \theta}$$

Di mana L_1 dan L_2 adalah literal yang saling berlawanan (misal $P(x)$ dan $\neg P(a)$). Agar mereka bisa saling menghilangkan (cancel out), argumen di dalamnya harus disamakan terlebih dahulu.

Proses penyamaan argumen ini disebut Unifikasi. Unifikasi mencari substitusi θ (disebut Most General Unifier / MGU) yang membuat dua atom menjadi identik secara sintaksis. Contoh: Unifikasi $P(x)$ dan $P(Alex)$ menghasilkan $\theta = \{x / Alex\}$. Tanpa unifikasi, resolusi pada FOL tidak mungkin dilakukan karena variabel pada premis yang berbeda harus diselaraskan.



DATASET PRONTOQA



Data tidak diambil dari internet, melainkan dibangkitkan dari aturan logika formal yang kemudian diterjemahkan ke bahasa alami karena sifatnya yang generatif dan sintetis.

Keunggulan: Kemampuannya mengisolasi kemampuan deduksi dari pengetahuan dunia (world knowledge) dengan menggunakan istilah fiktif (seperti "Wumpus" atau "Tumpus") mencegah model menjawab benar hanya karena hafal fakta (misalnya "Burung bisa terbang").

01

Sistem membuat graf ontologi fiktif

Wumpus → Jompus, Jompus → Tumpus

02

Sistem menurunkan rantai deduksi berdasarkan graf tersebut.

Setiap Wumpus adalah Jompus dan Setiap Jompus adalah Tumpus

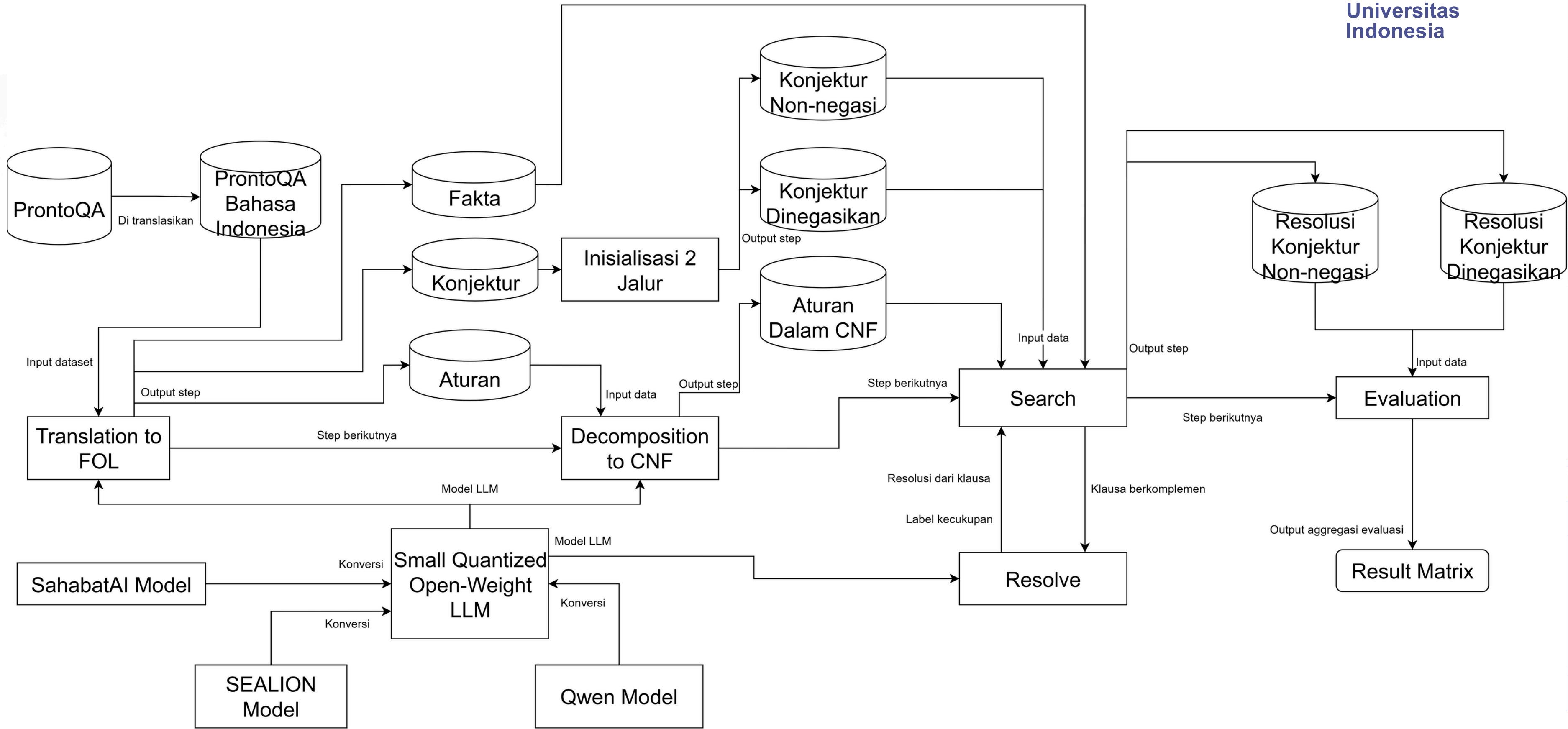
03

Simbol-simbol logika diganti dengan kata-kata fiktif atau nyata untuk membentuk kalimat bahasa alami.

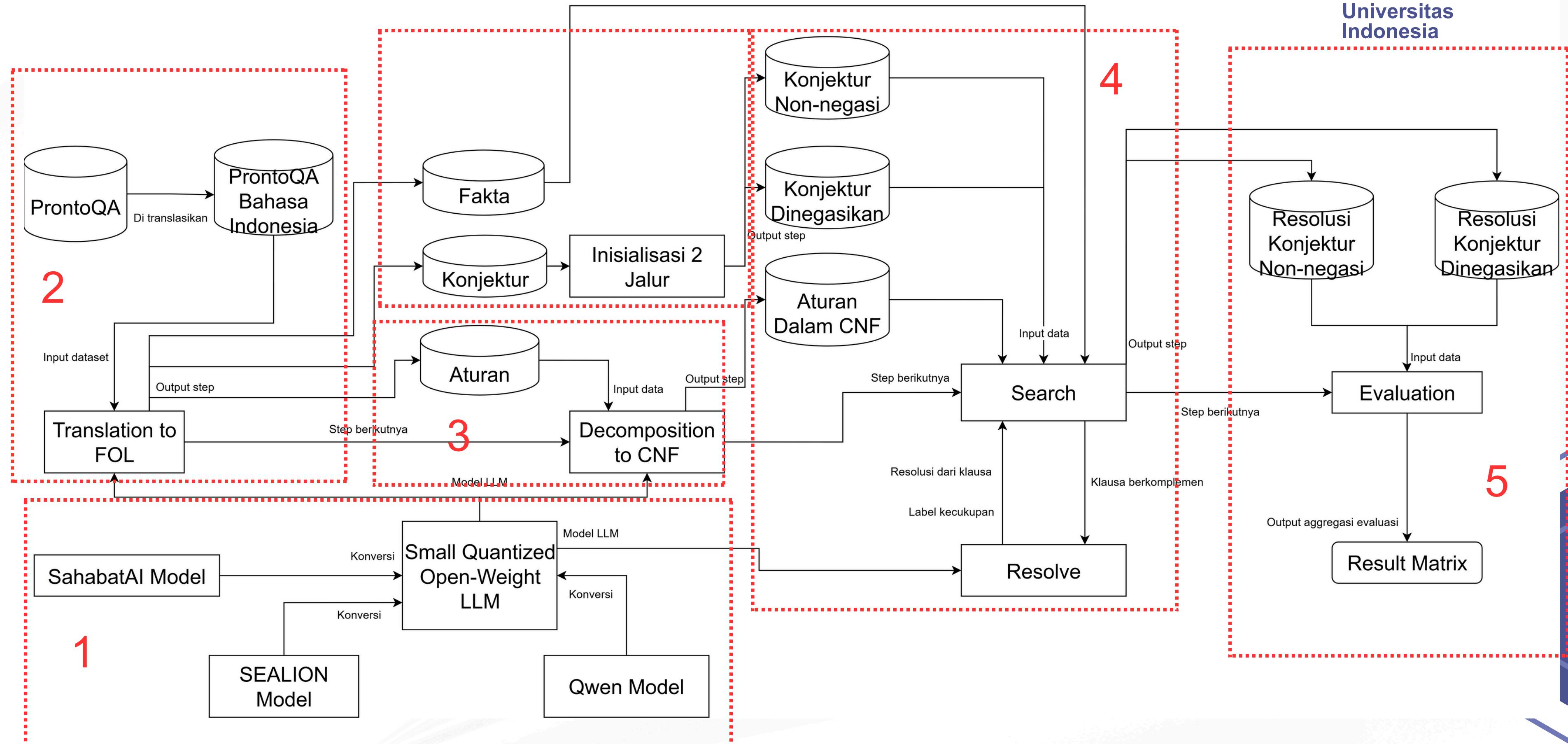
Setiap Wumpus adalah Merah. Wumpuses adalah Jompus. Jompus bersifat transparan. Semua Jompus adalah Tumpus

FRAMEWORK ARISTOTLE

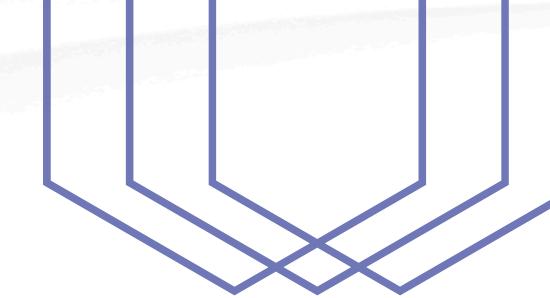
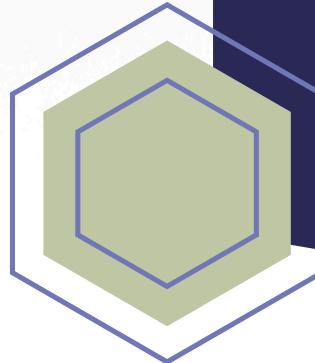
Universitas
Indonesia



FRAMEWORK ARISTOTLE



PEMILIHAN MODEL



Parameter untuk setiap model dibuat sama dengan ketentuan berikut:

- **Temperature:** Ditetapkan ke **0.0** sehingga model hanya memilih token dengan probabilitas tertinggi.
- **Max Tokens:** Adapun token yang dibutuhkan adalah **2500 tokens** 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.
n_ctx di set ke 0
- **n_gpu_layers:** Menentukan jumlah layer model yang dijalankan di GPU untuk akselerasi. **Nilai di set ke -1**

01

Qwen2.5-7B-IT-GGUF

Model ini dipilih untuk karena keunggulannya dalam benchmark logika matematika global. Model ini merepresentasikan kemampuan model global dalam dataset lokal ini

02

SEALIONv3-Llama-8B-IT-GGUF

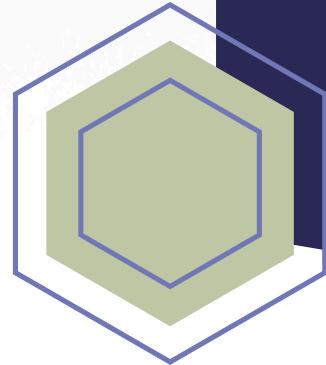
Model ini dipilih karena sudah melalui CPT(continued pre-training) dalam bahasa di Asia Tenggara. Model ini merepresentasikan kemampuan model regional dalam dataset lokal ini

03

SahabatAlv1-Llama-8B-IT-GGUF

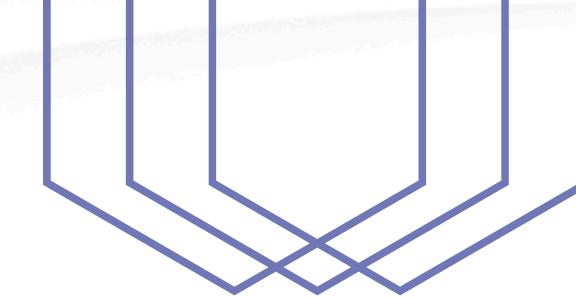
Model ini dipilih karena sudah melalui CPT(continued pre-training) dalam bahasa Indonesia. Model ini merepresentasikan kemampuan model lokal dalam dataset lokal ini

TRANSLATION TO FOL



Menerjemahkan narasi bahasa alami menjadi representasi First-Order Logic (FOL) yang terdiri dari Fakta, Aturan Implikasi, dan Konjektur.

Model diberikan few-shot prompt yang berisi contoh pasangan kalimat Indonesia dan format logika target.



Fakta

Format: $P(\text{Subjek}, \text{NilaiKebenaran})$.

Contoh: $\text{Wumpus}(\text{Max}, \text{True})$

01

Aturan

Format: $P(\text{Subjek}, \text{NilaiKebenaran}) \ggg P(\text{Objek}, \text{NilaiKebenaran})$.

Contoh: $\text{Wumpus}(\$x, \text{True}) \ggg \text{Manis}(\$x, \text{False})$

02

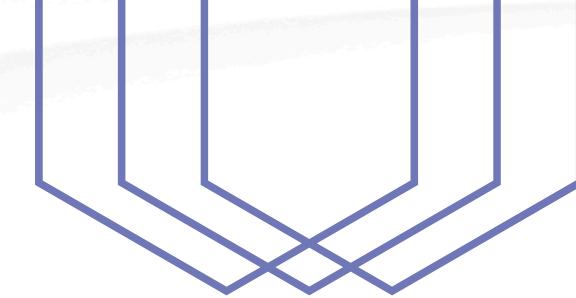
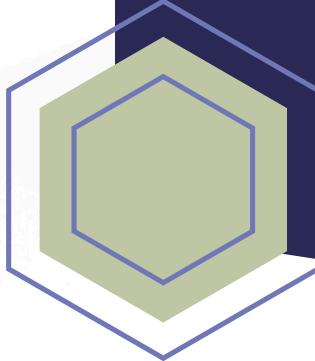
Konjektur

Format: $P(\text{Subjek}, \text{NilaiKebenaran})$.

Contoh: $\text{Manis}(\text{Max}, \text{True})$

03

DECOMPOSITION TO CNF



Universitas
Indonesia

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.

Hasil Decomposition to CNF:

Contoh:

Input: Wumpus(\$x, True) >>>

Manis(\$x, False)

Output: \left(Wumpus(x, False) \lor Manis(x, False) \right)

01

Prenex Normal Form (PNF)

Konversi FOL ke bentuk PNF, di mana semua kuantor diletakkan di awal formula, diikuti oleh matriks bebas kuantor.

02

Skolemisasi

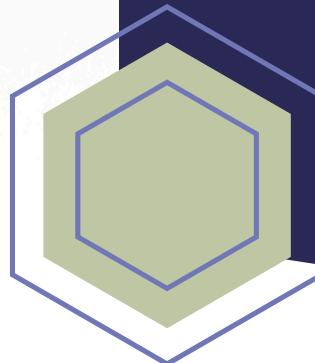
Penghapusan kuantor eksistensial (\exists) dengan mengganti variabel eksistensial dengan fungsi Skolem.

03

Conjunctive Normal Form (CNF)

Transformasi akhir ke CNF, di mana formula diekspresikan dalam bentuk konjungsi (AND) dari disjungsi (OR) literal.

SEARCH AND RESOLVE



Search Resolve merupakan sebuah modul dari *framework* dengan dua fungsi

01

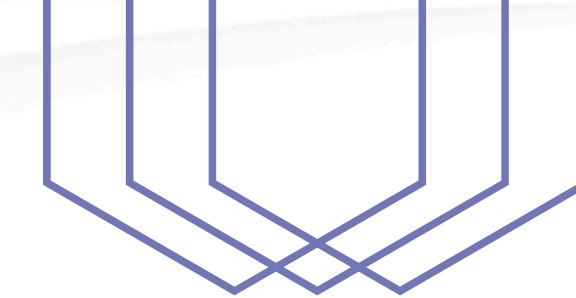
Logical Search Router

Memilih klausa mana yang relevan untuk diproses dan juga maksimum ronde pencarian atau ruang pencarian

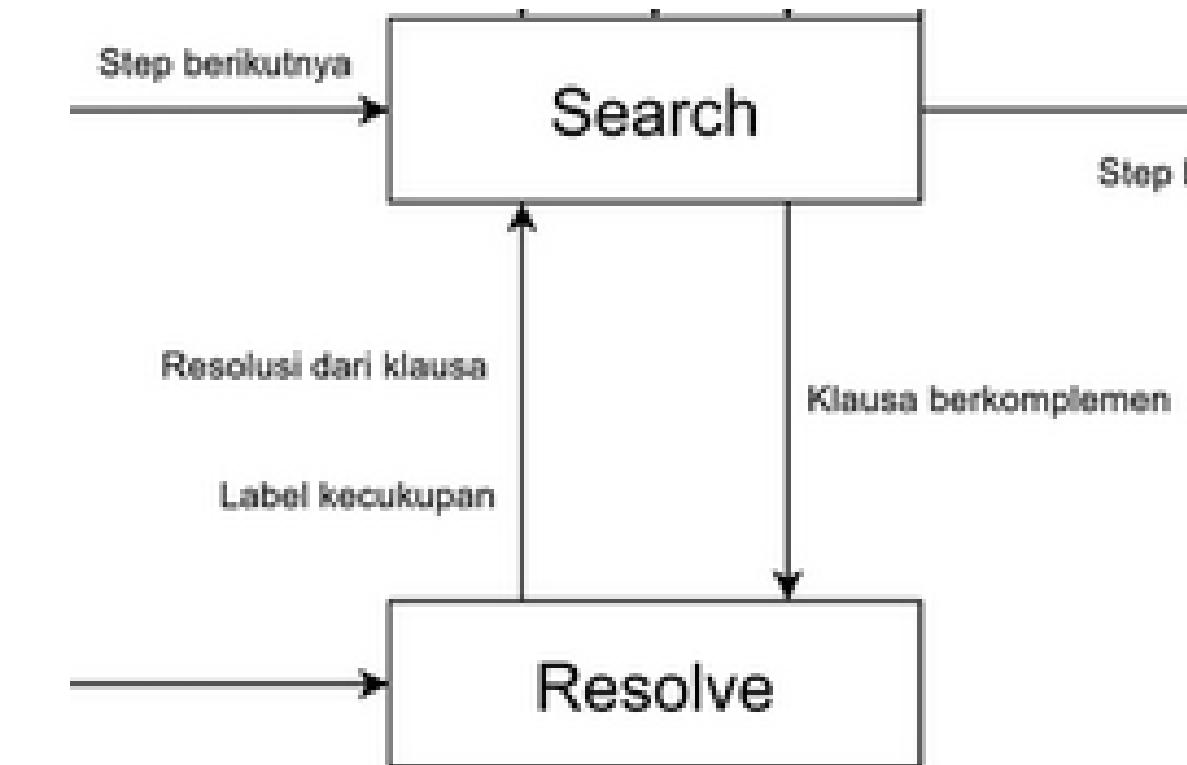
02

Logical Resolver

Menerapkan aturan resolusi secara iteratif hingga ditemukan kontradiksi (untuk pembuktian salah) atau hingga ruang pencarian habis.



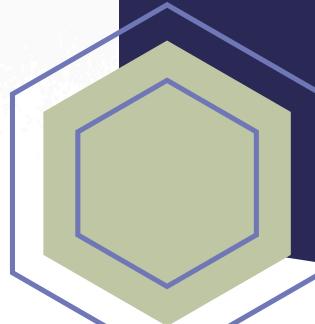
Universitas
Indonesia



Search Router akan mencari semua klausa yang berkomplemen dan akan dikirimkan ke SLM untuk di cari resolusinya.

Hasil dari SLM akan dikembalikan berupa klausa baru beserta apakah ditemukan kontradiksi dengan label kecukupan

SEARCH AND RESOLVE



Tabel 3.1: Matriks Keputusan untuk Penentuan Label Jawaban (Kesimpulan)

Kondisi (Skenario)	Jalur A: Buktikan S (Cari Kontradiksi di $\neg S$)	Jalur B: Buktikan $\neg S$ (Cari Kontradiksi di S)	Kesimpulan
1	Berhasil (Ada Kontradiksi)	Gagal	TRUE
2	Gagal	Berhasil (Ada Kontradiksi)	FALSE
3	Gagal	Gagal	UNKNOWN
4	Berhasil	Berhasil	SELF-CONTRADICTORY

01

Jalur A: Konjektur Dinegasikan

Asumsikan $\neg S$ (Negasi S), jika ditemukan kontradiksi (Klausa Kosong), maka S terbukti Benar.

02

Jalur B: Konjektur non-negasi

Asumsikan S , jika ditemukan kontradiksi, maka $\neg S$ terbukti Benar.

03

Jawaban Akhir

Apabila Jalur A berhasil ditemukan kontradiksi dan Jalur B tidak ditemukan kontradiksi, maka jawaban dari Konjektur tersebut adalah True dan akan di bandingkan dengan ground_truth / jawaban asli untuk data point tersebut



DESAIN EKSPERIMENTEN

Variabel dalam eksperimen ini:

Variabel Kontrol

1. Parameter setiap model dibuat sama dan dataset yang digunakan juga sama.
2. Setiap model terkuantisasi (4-bit GGUF)

Variabel Terikat

1. Akurasi (Accuracy): Ketepatan hasil akhir (Benar/Salah) yang diverifikasi terhadap ground truth.
2. Kepatuhan Format (Parsability): Kemampuan model menghasilkan sintaks logika (FOL/CNF) yang valid secara komputasi. Kegagalan menghasilkan format yang benar dianggap sebagai kegagalan penalaran.

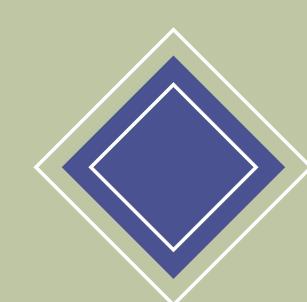
01

02

03

Variabel Bebas

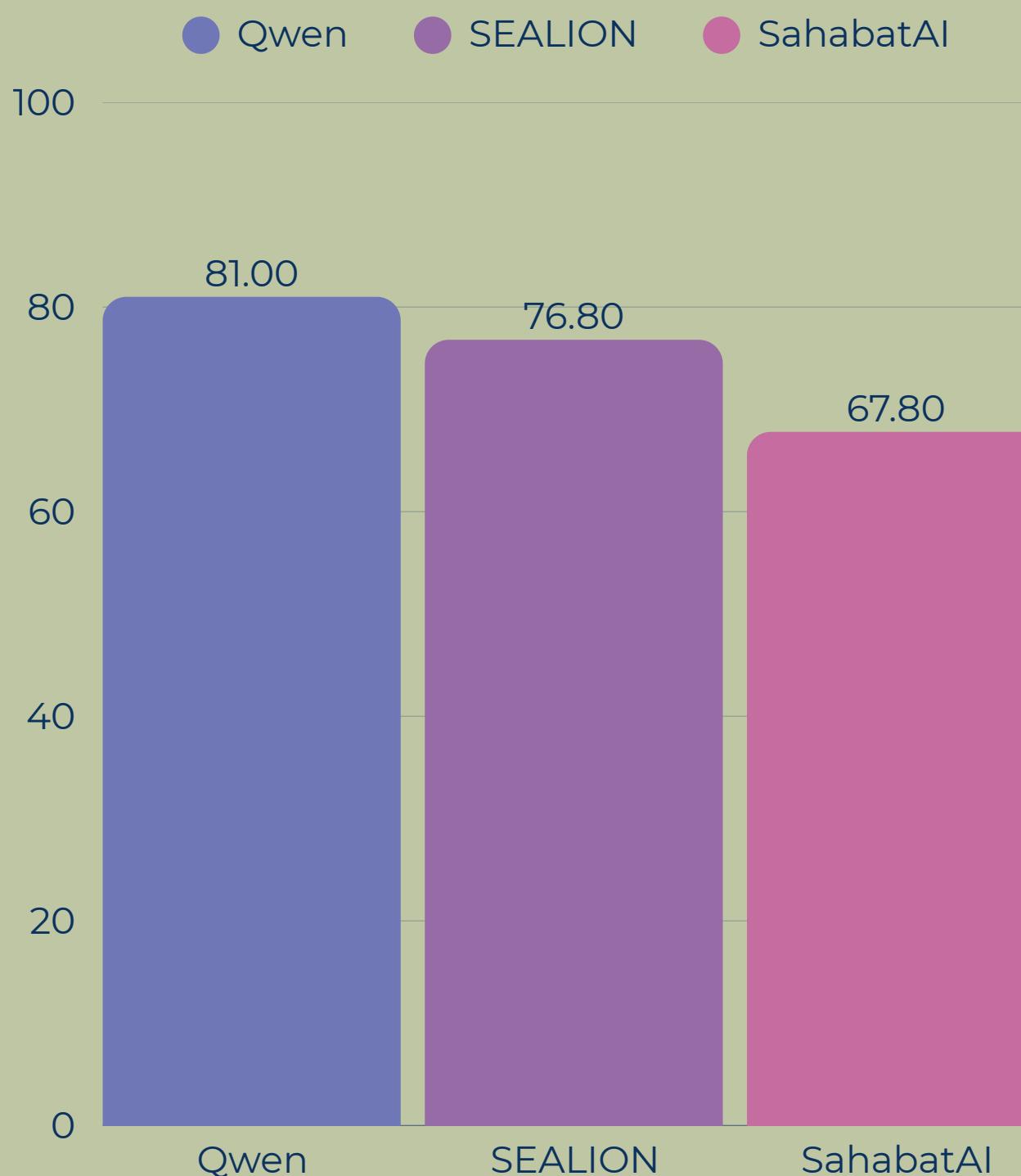
1. Model: Tiga kategori model yang mewakili spektrum pemahaman bahasa: Global (Qwen), Regional (SEA-LION), dan Nasional (SahabatAI).
2. Metode Prompting: Naive vs Aristotle Framework.



HASIL EKSPERIMENT

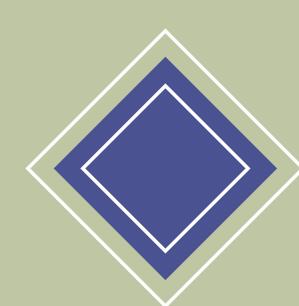


*Naive Prompting Explanation
Before Answer*



*Naive Prompting Explanation
After Answer*





HASIL EKSPERIMENT



Universitas
Indonesia

Ada beberapa poin yang bisa didapatkan dari hasil eksperimen tersebut:

1. Keunggulan Qwen pada CoT: Qwen2.5 menunjukkan lonjakan performa tertinggi (81%) ketika menggunakan metode Before Answer. Ini konsisten dengan literatur yang menyatakan bahwa model yang dilatih pada korpus kode/matematika besar (seperti Qwen) memiliki kemampuan Chain-of-Thought internal yang kuat (Wei et al. (2022)).

2. Kekuatan Semantik Sahabat-AI: Pada mode After Answer (yang lebih mengandalkan intuisi bahasa langsung), Sahabat-AI unggul (61.4%). Ini mengindikasikan bahwa model ini memiliki pemahaman Bahasa Indonesia yang lebih natural, sehingga intuisi "tebakan"-nya lebih akurat dibandingkan model global.

3. Naive Prompting dengan metode Explanation Before Answer menunjukkan hasil yang lebih baik dibanding Explanation Before Answer karena pada dasarnya SLM merupakan transformer yang menggunakan probabilitas untuk menentukan token selanjutnya ($P(w_t|w_{t-1}, \dots)$), sehingga dengan penjelasan sebelum jawaban akhir merupakan salah satu cara model melakukan Chain-of-Thought untuk menjelaskan bagaimana cara model mendapatkan jawaban tersebut

HASIL EKSPERIMENT

Ada beberapa penemuan menarik dari hasil eksperimen:

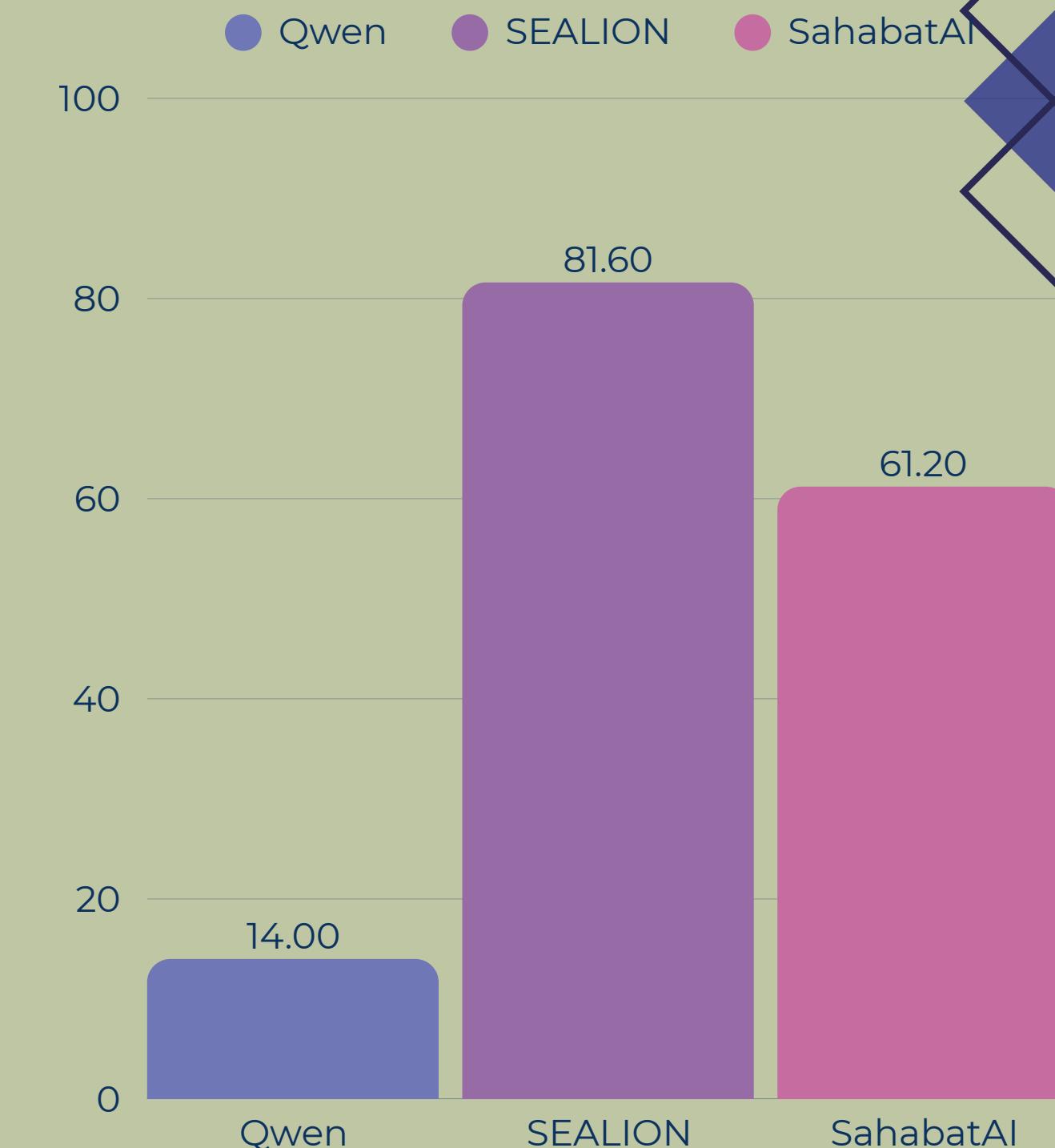
1. Anomali Qwen (14%): Meskipun unggul di Naive Prompting, Qwen gagal total dalam framework ini. Analisis log menunjukkan kegagalan ini bukan pada logika, melainkan pada kepatuhan format sintaks (parsability). Qwen sering menghasilkan output yang terlalu mengikuti format pada tahap search & resolve tanpa mengandalkan kemampuan penalaran logikanya, sehingga jawaban pada tahap tersebut menjadi tidak valid.

2. Dominasi SEA-LION (81.6%): SEA-LION v3 menunjukkan performa terbaik. Ini membuktikan hipotesis bahwa model regional yang dilatih dengan instruksi spesifik bahasa lokal (dan mungkin data multilingual alignment yang lebih baik) mampu menjadi "Penerjemah Logika" yang lebih patuh pada tahapan-tahapan framework Aristotle pada dataset berbahasa Indonesia.

3. Stabilitas Sahabat-AI: Sahabat-AI menunjukkan performa yang seimbang. Meskipun tidak setinggi SEA-LION, akurasinya (61.2%) sebanding dengan performa Naive-nya, menunjukkan konsistensi pemahaman.

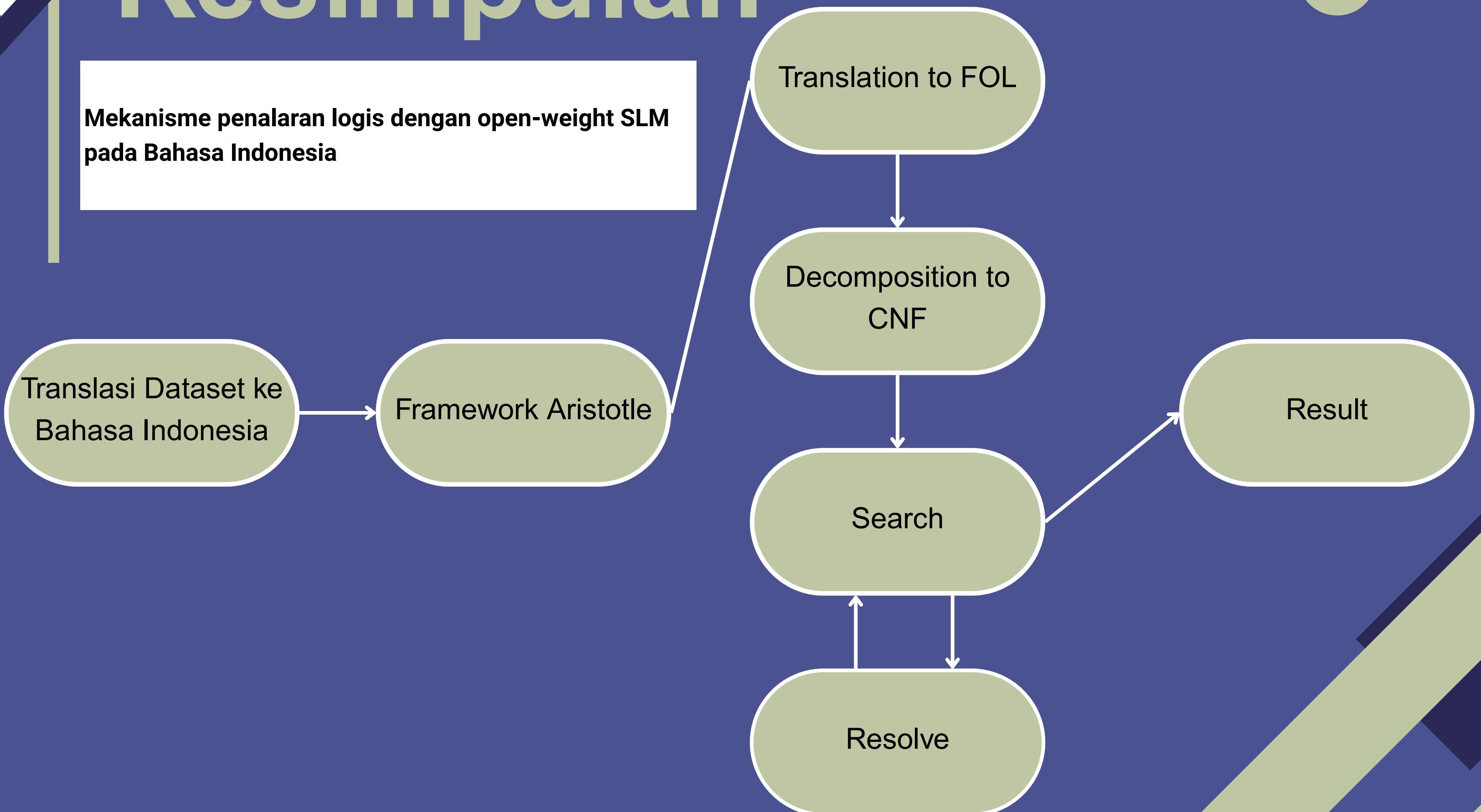


Framework Aristotle



Kesimpulan

Mekanisme penalaran logis dengan open-weight SLM pada Bahasa Indonesia



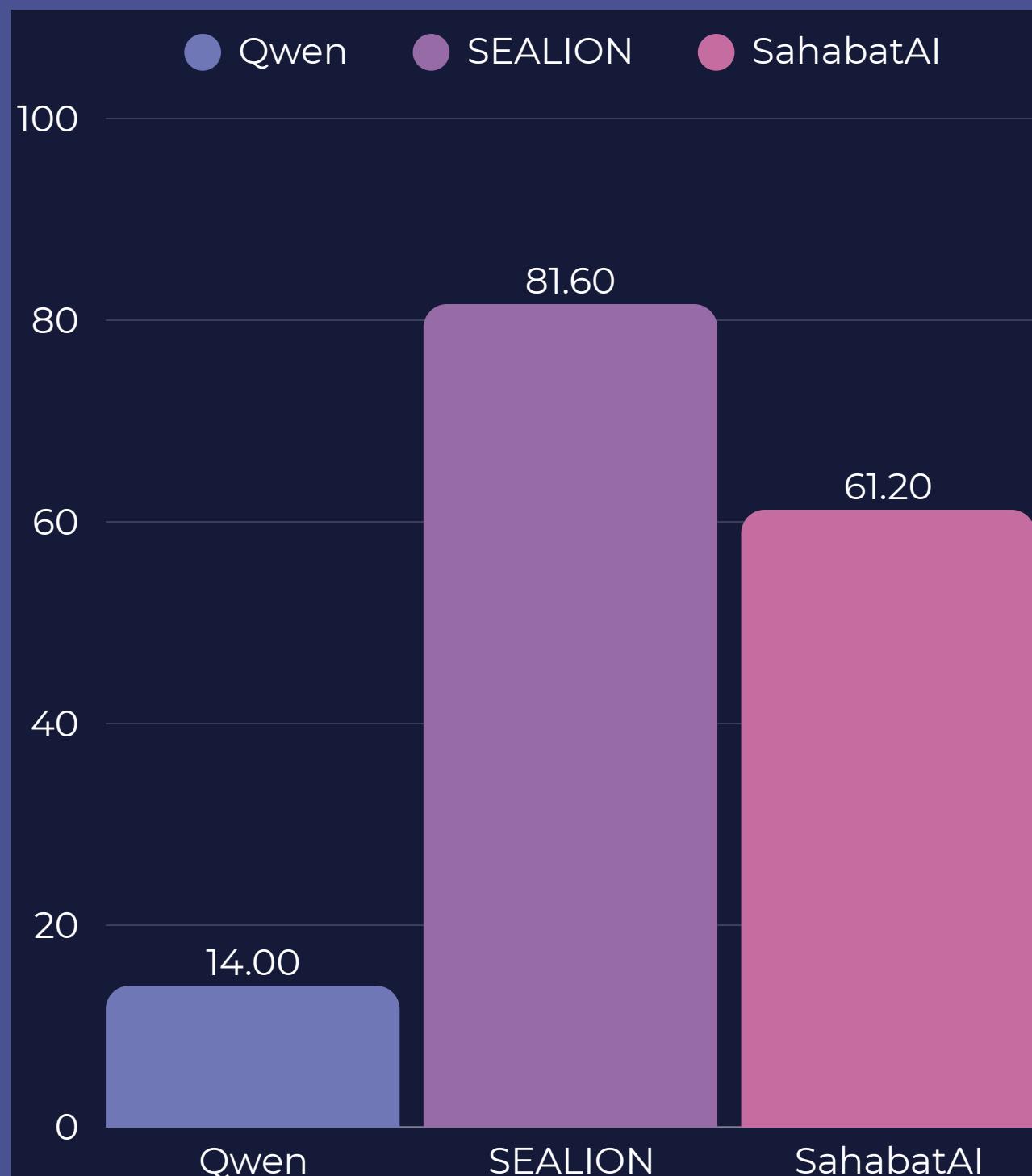
Kesimpulan

Perbandingan Performa Antar Model Open-Weight:

Best: SEALION v3
(Regional)

Worst: Qwen2.5
(Global)

Middle: SahabatAIv1
(Nasional)

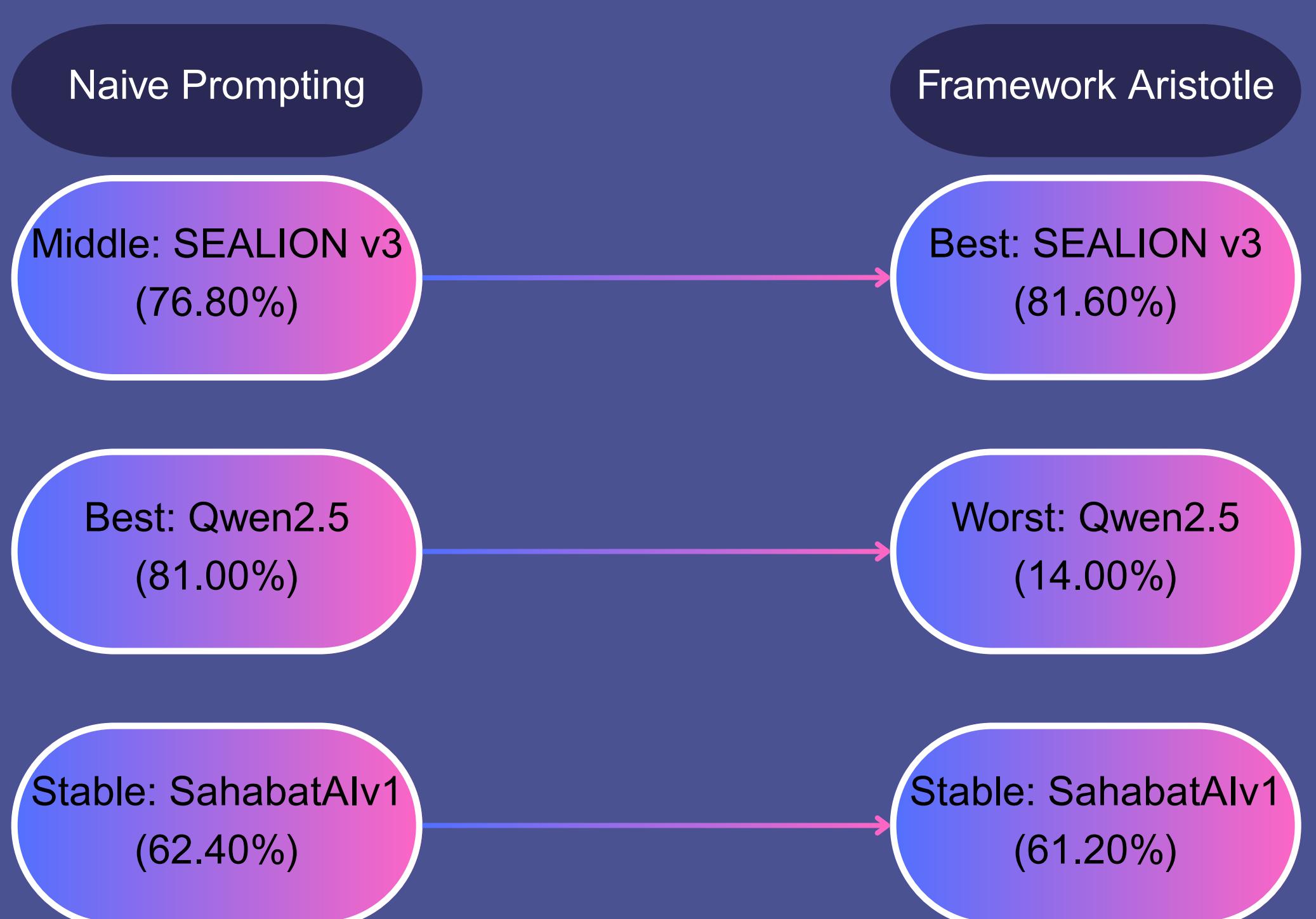


Universitas
Indonesia



Kesimpulan

Efektivitas Aristotle vs Naive Prompting (Explanation Before Answer):





Universitas
Indonesia

Sekian & Terima Kasih

Dipresentasikan oleh :

MIKHAEL DEO BARLI
NIM 1906350572
Jurusan Ilmu Komputer

INSPIRASI



BERANDA PROFIL PUBLIKASI MAPANRB JDIH PPID



Pemerintah Apresiasi Pengembangan Sahabat-AI sebagai LLM dengan Konteks Keindonesiaan

03 Jun 2025 Dilihat: 4976



Berita Terbaru

- 12 Des 2025 Penataan Kelembagaan dan Tata Laksana Dilakukan Secara Holistik dan Berorientasi Pada Efektivitas
- 12 Des 2025 Rapat Konsolidasi Penataan Bidang Kelembagaan dan Tata Laksana Instansi Pusat
- 12 Des 2025 Proses Aksesi Indonesia ke OECD, Momentum Penguatan Reformasi Birokrasi
- 12 Des 2025 Bilateral Meeting dengan Director of Public Governance OECD
- 11 Des 2025 Pertemuan ST ke-26, DWP Kementerian PANRB selenggarakan Bakti Sosial dan Bazar UMKM
- 11 Des 2025 es Aksesi Masuk Tahap Technical ew, Menteri PANRB: Agenda maksi Indonesia Selaras dengan ip Tata Kelola Publik OECD
- 11 Des 2025 en PANRB: JPT Madya Sebagai

Universitas
Indonesia



BERANDA PROFIL NASIONAL RAGAM LAYANAN BERITA GALERI LAPOR!



Kata Pencarian ..

Beranda / Sosial & Budaya / Wamenkomdigi: LLM Lokal Sahabat AI Sesuai Peta Jalan yang Sedang Dibuat Pemerintah

Wamenkomdigi: LLM Lokal Sahabat AI Sesuai Peta Jalan yang Sedang Dibuat Pemerintah

Administrator | Rabu, 4 Juni 2025 | 10:34 WIB



BERITA POPULER

Welli, Sosok Tangguh di Balik Lancarnya Posko Bencana di SDN 02 Cupak Tangah

08 Des 2025 | 14:52 WIB



Ketika Mesin Ekonomi Indonesia Bangkit Bersama

08 Des 2025 | 12:41 WIB



PASAR KEUANGAN INDONESIA

<https://www.menpan.go.id/site/berita-terkini/pemerintah-apresiasi-pengembangan-sahabat-ai-sebagai-llm-dengan-konteks-keindonesiaan>



Sahabat-AI Semakin Canggih:
Indosat dan GoTo Luncurkan
Model 70 Miliar Parameter
dengan Layanan Chat
Multibahasa

02 JUNE 2025



<https://www.menpan.go.id/site/berita-terkini/pemerintah-apresiasi-pengembangan-sahabat-ai-sebagai-llm-dengan-konteks-keindonesiaan>



ABLASI

Universitas
Indonesia