

MUHAMMAD FAHMI

— DS / AI / NLP ENGINEER

Portofolio berdasarkan pengalaman profesional selama ±5 tahun di bidang AI dan NLP.

Production ML



Distributed Processing



NLP



Observability

CORE OBJECTIVE

Fokus: pada desain arsitektur, reliability, dan hasil sistem di production.

Link Personal Website & Portfolio: <https://fahmi.vercel.app/>



System > Model

"Saya tidak fokus membuat model paling kompleks. Saya fokus membuat **sistem yang tetap benar** saat realitas tidak ideal."

⚠ DETECTED FAILURE MODES

ERR_01: DRIFT

Perubahan distribusi data

ERR_02: INPUT

Input tidak terstruktur / tidak valid

ERR_03: USAGE

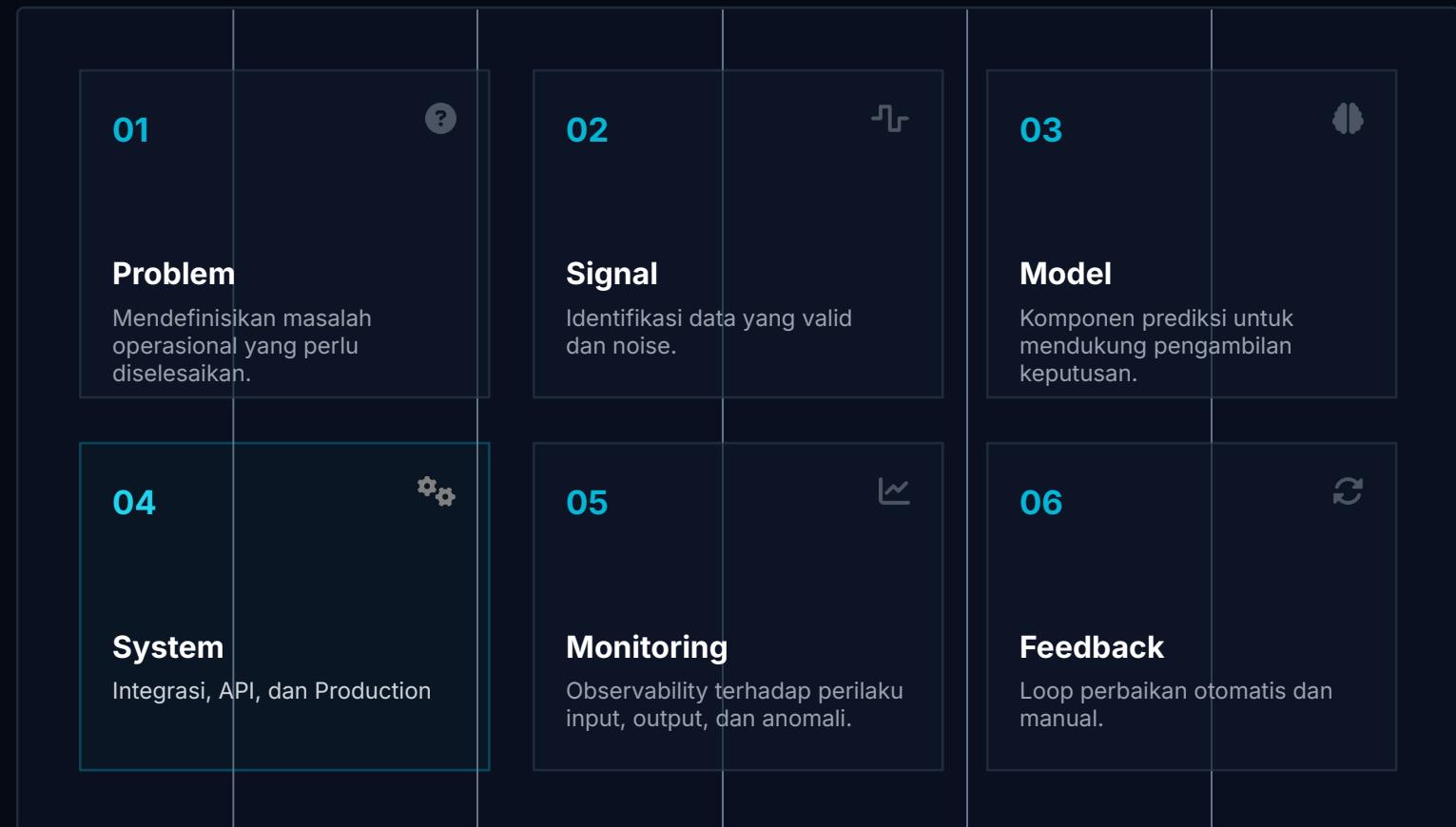
Asumsi sistem dilanggar oleh user

ERR_04: INFRA

Gangguan performa dan timeout

Engineering Lifecycle

PIPELINE STATUS: ACTIVE



REAL-TIME NLP PLATFORM

✖ OPERATIONAL PROBLEM

Respons terhadap opini publik sangat bergantung pada ketepatan waktu. Informasi yang terlambat membuat keputusan menjadi kurang relevan.

🚫 NAIVE APPROACH FAILURE

Pendekatan laporan periodik menyebabkan isu baru teridentifikasi saat situasi sudah berkembang.

SYSTEM ARCHITECTURE

STATUS: LIVE



✓ Streaming Event-Driven

✓ Tenant Isolation

✓ Time-Window Aggregation

🔧 ENGINEERING CHALLENGES

- ⚡ Data masuk tidak berurutan sehingga agregasi berbasis waktu harus tetap konsisten.
- ⚡ Lonjakan percakapan mendadak berpotensi membebani worker secara bersamaan.

⚖️ SOLUTION APPROACH

Consistency



Latency

- Event-driven processing untuk respons cepat terhadap perubahan percakapan
- Load distribution antar worker untuk mencegah bottleneck saat spike
- Time-window aggregation untuk mengurangi noise dan false alert

⌚ OPERATIONAL IMPACT

- Visibilitas situasi publik meningkat secara real-time
- Pengambilan keputusan menjadi lebih proaktif
- Risiko eskalasi isu dapat ditekan lebih awal

ABSA SENTIMENT ENGINE

✖ OPERATIONAL PROBLEM

Satu kalimat dapat mengandung opini berbeda terhadap beberapa entitas, sehingga label tunggal tidak cukup merepresentasikan makna sebenarnya.

Contoh
"Aplikasinya cepat, tapi CS-nya lama respon."

🚫 NAIVE APPROACH FAILURE

Sentimen general (klasifikasi satu data/artikel) membuat insight tidak bisa ditindaklanjuti karena menggabungkan aspect A dan aspect B menjadi satu skor sentimen.

🔧 ENGINEERING CHALLENGES

- 👥 Multi-entity dalam satu kalimat menyulitkan pemetaan sentimen ke target yang tepat
- ⚠️ Bahasa informal & ironi parsial menurunkan akurasi klasifikasi

PIPELINE_FLOW

MODEL: ASPECT-BASED



⚖️ SOLUTION APPROACH

Flexibility Consistency

- Fine-tuned IndoBERT (HuggingFace) untuk klasifikasi sentimen per-aspect
- Input dipasangkan (aspect, text) agar model memahami target opini
- Probabilistic calibration + confidence threshold untuk menjaga konsistensi prediksi

⌚ OPERATIONAL IMPACT

- Fine-tuned IndoBERT (HuggingFace) untuk klasifikasi sentimen per-aspect
- Input dipasangkan (aspect, text) agar model memahami target opini
- Probabilistic calibration + confidence threshold untuk menjaga konsistensi prediksi

LANGUAGE ROUTING FILTER

✖ OPERATIONAL PROBLEM

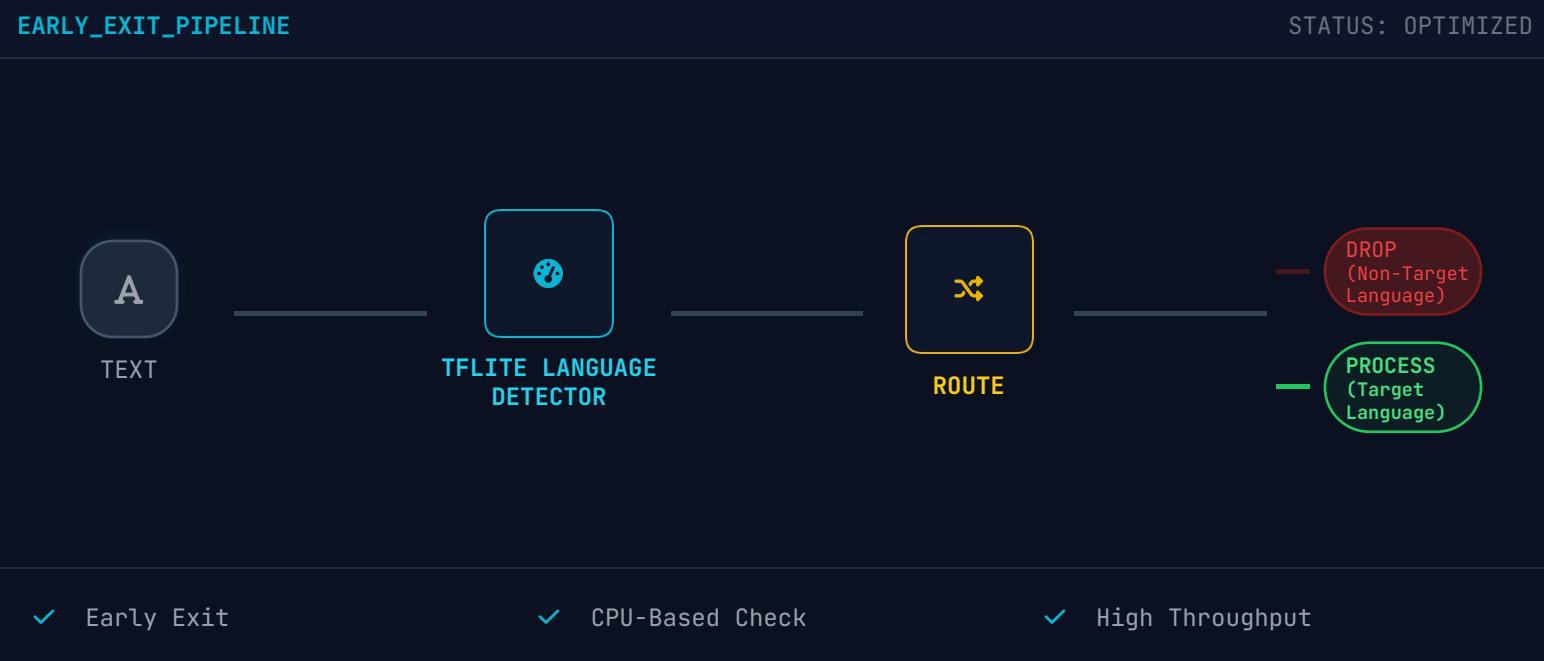
Sebagian data masuk tidak relevan dengan konteks proyek (misalnya bahasa asing), namun tetap diproses oleh model utama yang berat.

🚫 NAIVE APPROACH FAILURE

Filter dilakukan di akhir pipeline. Akibatnya, resource mahal (GPU) sudah terbuang untuk memproses data sampah sebelum akhirnya dibuang.

🔧 ENGINEERING CHALLENGES

- 💬 Short text (1-2 kata) sering tidak memiliki fitur bahasa yang cukup untuk klasifikasi akurat.
- 💬 Dominasi emoji dan slang dalam data sosmed mempersulit deteksi bahasa baku.



⚖️ SOLUTION APPROACH

Recall (Coverage) ————— Efficiency

- Menggunakan Google Language Detector (TFLite) sebagai detect language berbasis CPU
- Routing hanya bahasa yang sesuai konteks proyek (misal: Indonesia) ke model NLP utama
- Early-exit pipeline untuk mencegah inference model berat pada data tidak relevan

⌚ OPERATIONAL IMPACT

- Model utama hanya memproses bahasa relevan dengan project
- Beban komputasi model NLP berat berkurang signifikan
- Pipeline lebih stabil pada volume data tinggi

CLUSTERING & SNA ENGINE

✖ OPERATIONAL PROBLEM

Data percakapan besar sulit dianalisis secara manual, sehingga diperlukan ekstraksi topik dan pemahaman komunitas secara otomatis.

🚫 NAIVE APPROACH FAILURE

Visualisasi graph tanpa struktur hanya menghasilkan visual yang kusut dan tidak bermakna. Analis tidak bisa membedakan noise dari komunitas.

🔧 ENGINEERING CHALLENGES

- ✗ Cluster berubah seiring waktu
- ✗ Noise percakapan tinggi menyebabkan data un-cluster

PIPELINE_FLOW

METHOD: HYBRID



⚖️ SOLUTION APPROACH

Coverage



Precision

- BERTopic (HDBSCAN) untuk menemukan cluster percakapan tanpa label
- LLM (LangChain + OpenAI) memberi nama topik tiap cluster
- Social Network Analysis untuk menemukan komunitas interaksi
- LLM merangkum pembahasan utama tiap komunitas

⌚ OPERATIONAL IMPACT

- Topik pembicaraan publik dapat dipahami tanpa membaca manual
- Komunitas dan fokus diskusinya teridentifikasi otomatis
- Insight dapat dicari dalam hitungan menit

AUTOMATED REPORT GENERATION

✖ OPERATIONAL PROBLEM

Laporan operasional harian perlu dikirim ke klien dalam format presentasi, namun proses penyusunan masih manual.

🚫 NAIVE APPROACH FAILURE

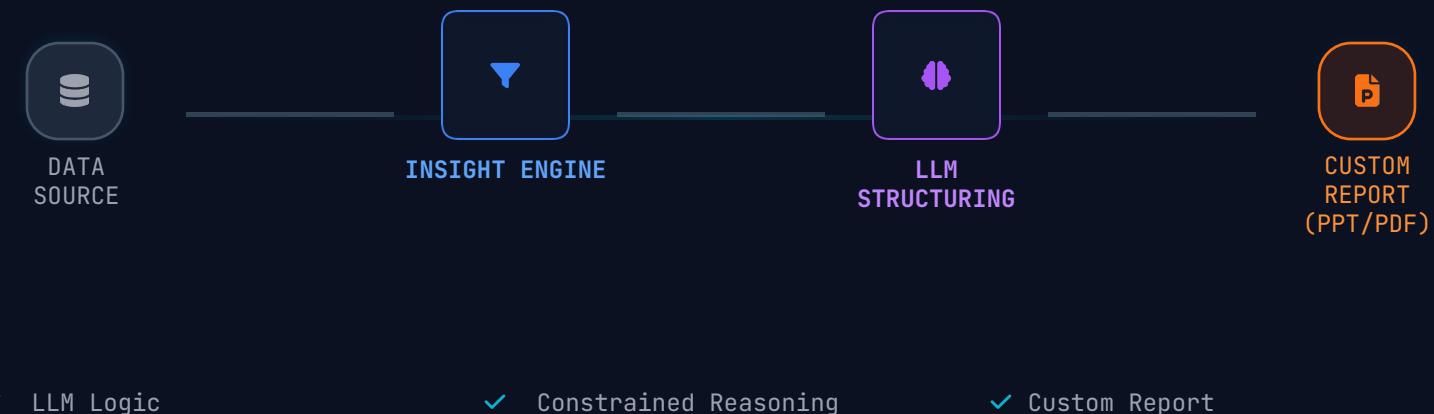
Menggunakan template statis tidak fleksibel terhadap variasi data, sementara penggunaan LLM bebas menghasilkan format tidak konsisten dan angka tidak terkontrol.

🔧 ENGINEERING CHALLENGES

- ⚠️ Hallucination: Memastikan LLM tidak mengarang angka yang tidak ada di data input.
- ⚡️ Format Stability: Output JSON dari LLM sering rusak (malformed) sehingga gagal di parsing.

GENERATION_PIPELINE

STATUS: ACTIVE



⚖️ SOLUTION APPROACH

Creativity ————— Consistency

- Data dikirim oleh DS dalam bentuk JSON/API terstruktur
- Rule-based insight extraction menentukan poin analisis
- LLM digunakan hanya untuk penyusunan narasi
- Engine menghasilkan dokumen PPT/PDF otomatis

⌚ OPERATIONAL IMPACT

- Analist fokus pada interpretasi, bukan penulisan laporan
- Laporan harian dapat dihasilkan otomatis dan konsisten
- Integrasi langsung dengan Dashboard / Product

HYBRID ML + LLM VERIFICATION

✖ OPERATIONAL PROBLEM

Model ML cukup akurat untuk kasus umum, namun gagal pada kalimat ambigu, sarkasme, dan konteks implisit.

🚫 NAIVE APPROACH FAILURE

Menggunakan LLM (seperti GPT-4) untuk memproses semua data sangat mahal dan memiliki latensi tinggi yang tidak praktis untuk skala besar.

HYBRID_REASONING_FLOW

STATUS: OPTIMIZED



✓ Cost Efficient

✓ High Precision

✓ Scalable Design

🔧 ENGINEERING CHALLENGES

- ⚠️ Menentukan ambang batas (threshold) confidence score yang tepat untuk memicu eskalasi ke LLM tanpa over-triggering.
- 💡 Mengelola latensi tambahan saat fallback ke LLM agar tidak memblokir proses utama.

⚖️ SOLUTION APPROACH

Simplicity ————— Complexity

- Model ML menangani mayoritas kasus secara real time
- Confidence threshold menentukan kasus ambigu / anomali
- Hanya prediksi ber-confidence rendah diverifikasi oleh LLM

⌚ OPERATIONAL IMPACT

- ±80% data diproses oleh model cepat
- ±20% kasus kompleks diverifikasi LLM
- Akurasi meningkat tanpa lonjakan biaya

MLOPS AUTO RETRAIN

✖ OPERATIONAL PROBLEM

Performa model di produksi menurun seiring perubahan data meskipun akurasi metric terlihat stabil.

🚫 NAIVE APPROACH FAILURE

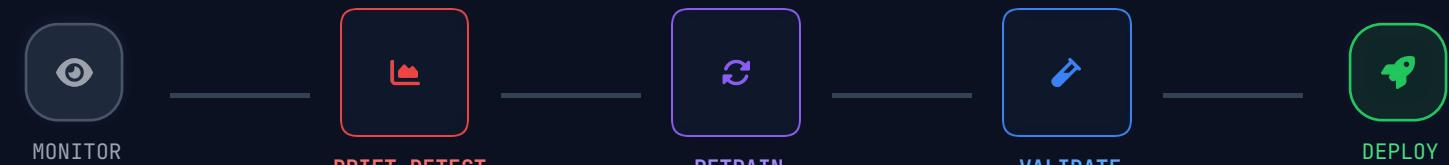
Retrain berkala (schedule-based) tidak mengikuti realitas. Kadang model diretrain padahal data stabil (boros), atau terlambat di retrain saat data berubah drastis.

🔧 ENGINEERING CHALLENGES

- ⌚ Label ground truth sering terlambat (delayed feedback), membuat deteksi drift akurasi menjadi sulit.
- ⌚ Menentukan threshold drift yang tepat agar tidak terlalu sensitif atau terlalu lambat.

RETRAINING_PIPELINE

STATUS: ACTIVE



✓ Drift Triggered

✓ Auto Evaluation

✓ Auto Re-Train Model

⚖️ SOLUTION APPROACH

Deploy Speed ————— Safety

- Monitoring distribusi data untuk mendeteksi drift
- Retrain hanya ketika drift terdeteksi
- Validasi otomatis sebelum deployment

⌚ OPERATIONAL IMPACT

- Model tetap relevan tanpa monitoring manual
- Mengurangi risiko penurunan performa
- Update model lebih aman

NER & STATEMENT EXTRACTION

✖ OPERATIONAL PROBLEM

Analisis narasi membutuhkan identifikasi aktor dan pernyataannya, bukan hanya kata kunci.

🚫 NAIVE APPROACH FAILURE

NER standar menemukan nama, tetapi tidak memahami hubungan pernyataan antar aktor.

EXTRACTION_PIPELINE

STATUS: ACTIVE



✓ Entity Normalization

✓ Speaker

✓ Quote Attribution

🔧 ENGINEERING CHALLENGES

📍 Kutipan tidak selalu berdekatan dengan aktor.

spep Menangani referensi implisit seperti "dia mengatakan" atau "pejabat itu membantah".

⚖️ SOLUTION APPROACH

Coverage



Precision

- Named Entity Recognition untuk identifikasi aktor
- Quote attribution untuk menentukan siapa mengatakan apa
- Entity normalization untuk konsistensi identitas

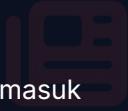
⌚ OPERATIONAL IMPACT

- Analisis narasi berbasis aktor dapat dilakukan otomatis
- Mendukung query seperti siapa mengatakan apa tentang siapa

MEDIA SUMMARIZATION

OPERATIONAL PROBLEM

Terlalu banyak artikel berita dan laporan yang masuk setiap hari melebihi kapasitas baca manual tim analis.



NAIVE APPROACH FAILURE

Ringkasan ekstraktif (memotong kalimat penting) sering kehilangan konteks nuansa dan menghasilkan informasi yang terpotong.



PIPELINE

STATUS: PRODUCTION



 Context Preservation

 Source Grounding

 Summarization

ENGINEERING CHALLENGES

-  Informasi antar sumber bisa saling bertentangan
-  Latensi meningkat karena multi-step

SOLUTION APPROACH

Speed (Latency)  Reliability

- Deduplication & chunking untuk menghindari informasi berulang
- Retrieval (vector search) sebelum LLM untuk menjaga grounding sumber
- LLM menyusun ringkasan berbasis konteks dokumen

OPERATIONAL IMPACT

- Menghasilkan Summary data dengan cepat
- Mendukung pengambilan keputusan cepat

TECH STACK

CORE_ENGINE.ML

MACHINE LEARNING & NLP

 **PyTorch**
Deep Learning Framework

 **Transformers**
Hugging Face

 **spaCy & NLTK**
Industrial NLP

 **ONNX**
Model Interchange

 **scikit-learn**
Classic ML Algorithms

 **LLM**
LangChain
OpenAI API

INFRA.OPS

INFRASTRUCTURE

 **Docker**
Containerization

 **SSH pm2**
Orchestration

 **Airflow**
Workflow Engine

 **FastAPI**
High Perf API

RUNTIME.LANG

</> LANGUAGE

 **Python**
Primary (3.11+)

 **Java / Scala**
Frontend/Node

DATA_PIPELINE.IO

DATA & STORE

 **Kafka**
Event Streaming

 **Vector DB**
Qdrant / Faiss

CLOUD.PROVIDER

CLOUD

 **AWS**
Notebook, SageMaker

 **GCP**
Vertex AI, BigQuery

● SYSTEM_SHUTDOWN_SEQUENCE_INITIATED

TERIMA KASIH

> Session terminated successfully.

FINAL_MESSAGE.TXT

"Mari berdiskusi tentang bagaimana membangun sistem AI yang **reliable, scalable**, dan memberikan **impact nyata** untuk bisnis kamu."

UPLOAD_COMPLETE

100%

PORTFOLIO_URL

<https://fahmi.vercel.app/>



DIRECT_MESSAGE

mfahmipamungkas123@gmail.com
WA : 081263299950



NETWORK

<https://www.linkedin.com/in/muhammad-fahmi17/>

