An LLM-Assisted Financial Risk Monitoring System

Data Engineering - Project Proposal – Fall 2026

Marc David Sutjipto: 49006528

Jan Peter Saal: 49006517

Luis Alan Morales Castillo: 49006499

Yassine Essakkati: 49006454

Agenda

1. Problem Definition & Solution
2. System Architecture
3. Data Exploration
4. Schedule

Problem Definition

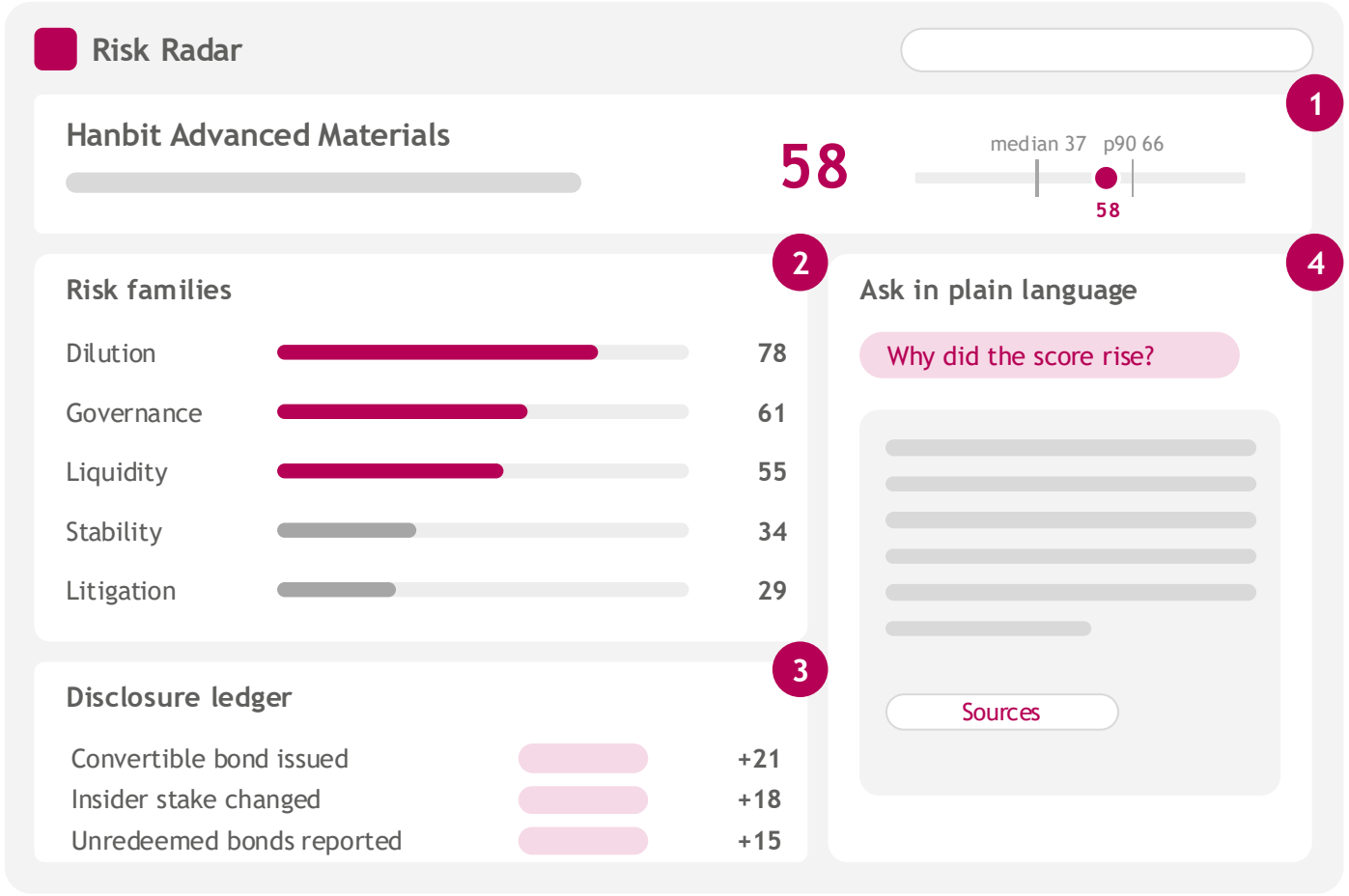
(1) Problem Definition & Solution

Hidden financial risks can impact returns, but there is no practicable risk monitoring for the Korean market

Dilution & Overhang	Governance	Liquidity & Debt	Business Stability	Litigation & Sanctions
• Convertible bond issued • Warrant bond issued • Rights issue decided	• Large-holding report • Top shareholder changed • Insider stake changed	• Default occurred • Unredeemed Bonds	• Operations suspended • Core assets sold • Division sold off	• Lawsuit filed • Rehabilitation filed • Creditor bank workout

Aspired Solution

(1) Problem Definition & Solution



1 Composite Score
One score from 0 to 100 per company, with its 12-month trend and quantiles of the industry

2 Risk Families
Five risk types scored side by side instead of in separate files

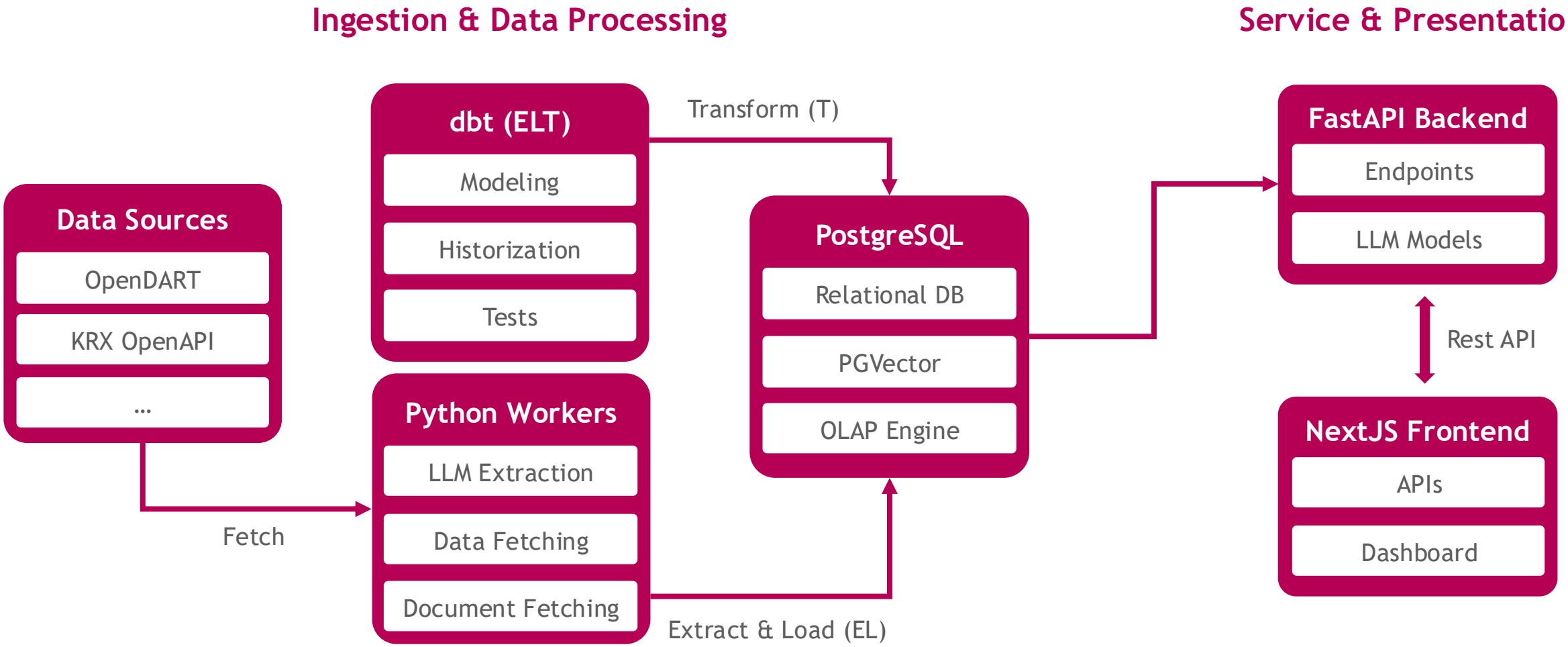
3 Disclosure Ledger
Every filing that moved a score, with date and impact

4 Ask in Plain Language
Answers in plain words, each linked to the source filings

Figure 1. Simplified view of the dashboard mockup. Mockup generated with Claude (Anthropic, 2026) from the authors specification, prompt and timestamp in AI Usage. All data shown are fictional.

Architecture Outline

(2) System Architecture



Data Collection

(3) Data Exploration

OpenDART API

corp_name	stock_code	report_nm	rcept_dt	corp_cls
Kakao Games	293490	Rights issue	2025-10-01	KOSDAQ
KR Motors	000040	Convertible bond	2025-10-01	KOSPI
YC Chem	112290	Warrant bond	2025-09-29	KOSDAQ
KBI Dongyang Cheolgwan	008970	Rights issue (corrected)	2025-10-01	KOSPI

Table 1. Example and simplified data excerpt from the OpenDart API

Data Pre-Processing

(3) Data Exploration

1 Type-Cast Numbers

Data type casting

```
"9,562,408"
"-"
```

↓

```
9562408
0
```

→ Text to numbers

2 Normalize Dates

3 endpoints, 3 different formats

```
"2024년 03월 04일"
"20221129"
"-"
```

↓

```
2024-03-04
2022-11-29
NULL
```

→ One standard date format

3 LLM Extraction

From text data to structured outputs

…보통주식 4,885,993주, 1주당 발행가액 1,228원

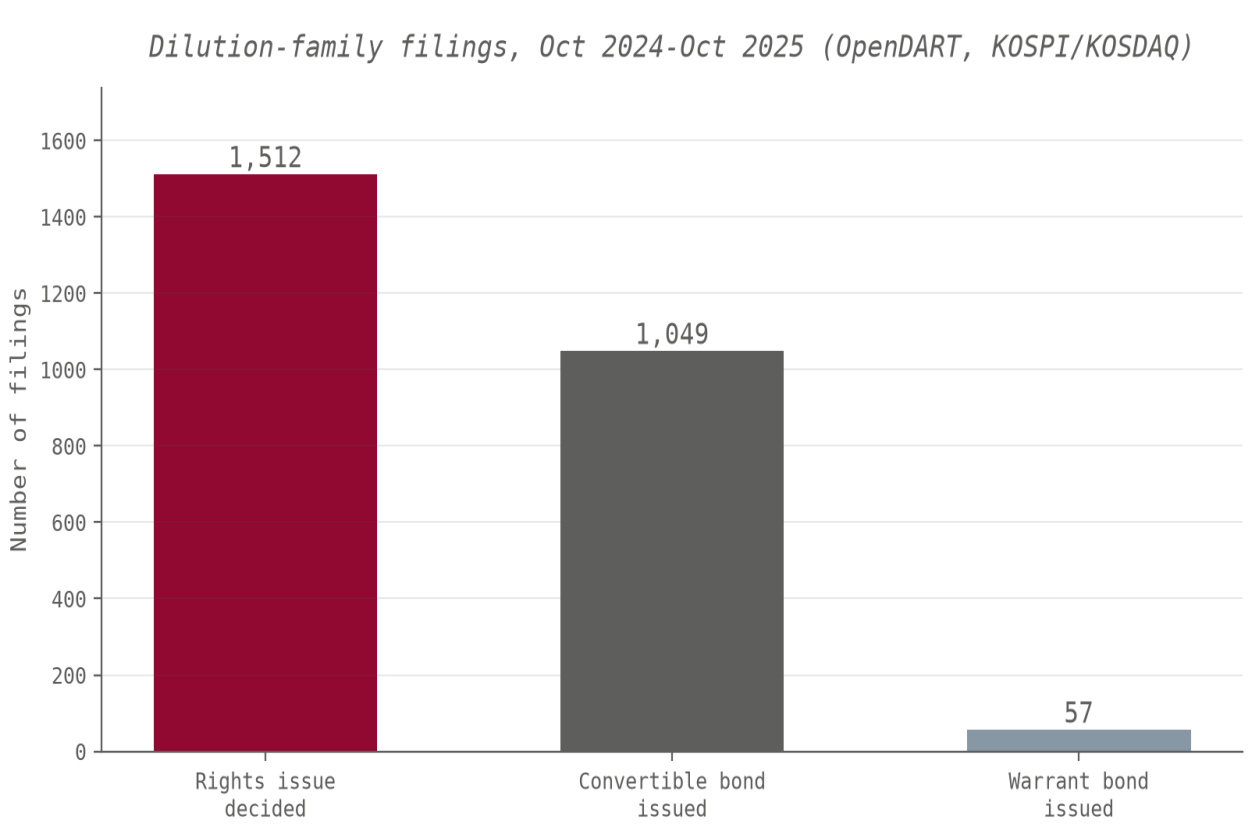
↓ LLM

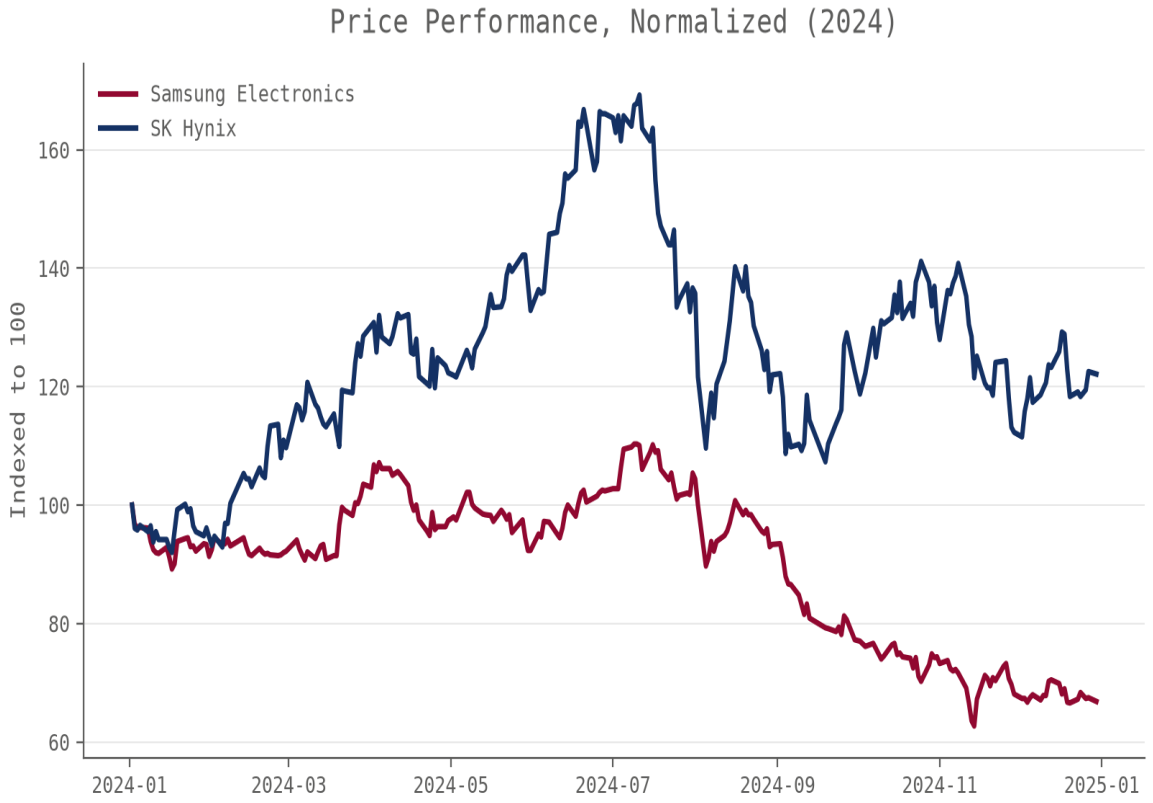
```
shares:  4885993
price:   1228 KRW
links:   Dilution
```

→ Structured outputs

Data Exploration Examples

(3) Data Exploration & Architecture





Roadmap

(4) Schedule

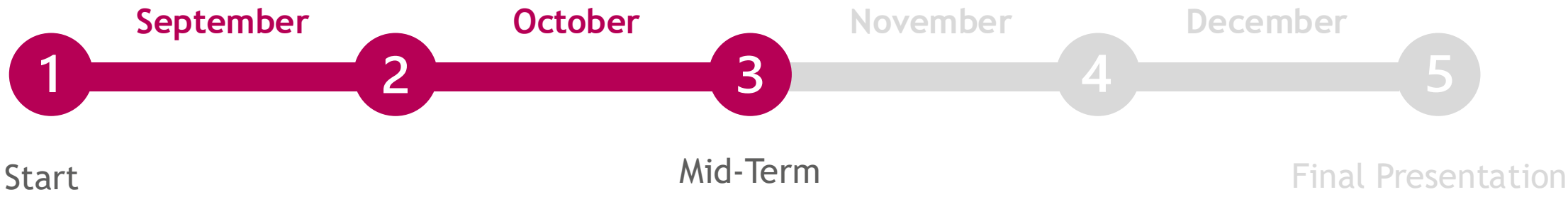
Phase 1
Collect
Find and collect data

Phase 2
Explore
Analyze and explore data

Phase 3
Modeling
Database schema and storage

Phase 4
Build
Build the application

Phase 5
Demo
Demonstrate the results



Roadmap

(4) Schedule

Phase 1
Collect
Find and collect data

Phase 2
Explore
Analyze and explore data

Phase 3
Modeling
Database schema and storage

Phase 4
Build
Build the application

Phase 5
Demo
Demonstrate the results



AI Usage

- Anthropic. (2026). *Claude* (Opus 5) [Large language model]. https://claude.ai
  Prompt: „Das ist unser aktueller Stand. Baue mir einen Musterscreenshot davon, wie das fertige Produkt aussehen könnte“ (submitted together with the file *group2_data_engineering_proposal.pptx*). Generated 18 October 2026, 16:35 KST.