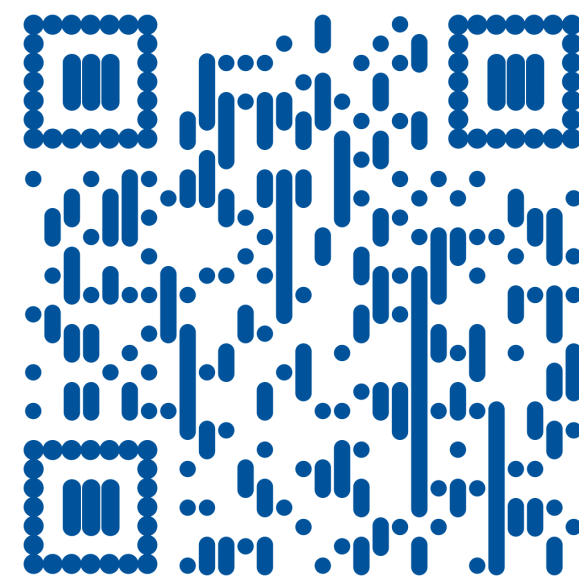


Labour and Material Cost Reduction using Agentic AI

P&G Cost Advisor · Ben Cat Plant



Team Cyberpunk

Phyu Phyu Shinn Thant · s4022136
Vo Nguyen Bao Ngoc · s3975091
Bui Phuong Dong Quan · s3975939

Nguyen Minh Khoi · s3978862
Le Minh Tuan · s4032012

Academic Supervisor: Dr. Truong Nguyen Xuan Vinh · Industry Supervisor: Tran Nguyen Cao Minh

100%

Cost Calculation Accuracy
With Deterministic Excel
Parity Checks

6

Optimization Goals for
Scenario Simulations

0

Cost Hallucinations and Fake
Numbers Produced by the
Advisor

Context

P&G's Ben Cat Plant packages Gillette razors at high volume. Labour and material cost is tracked per SKU **by hand** with analysts consolidating across several Excel workbooks and simulate scenarios in spreadsheets and PowerPoint.

Across **623 in-scope SKUs**, and against an 8.9% wage rise in 2025 with the fastest input-cost inflation in fifteen years, that manual process no longer scales.

Cost Advisor makes it continuous by projecting cost, **deterministically calculating costs** through the plant's own labour formula, **simulating various scenarios across different optimization goals** to compare and present opportunities to create savings, presenting these scenarios in a ranked orderly manner, and **monitoring volume swing** to close the gap and reduce repetitive manual effort.

Objectives

Reproduce the plant's own calculation. Rebuild the Excel labour and material calculations in fixed precision arithmetic, including the stepped manpower hierarchy. Verified against 20 supervisor-agreed set of SKUs.

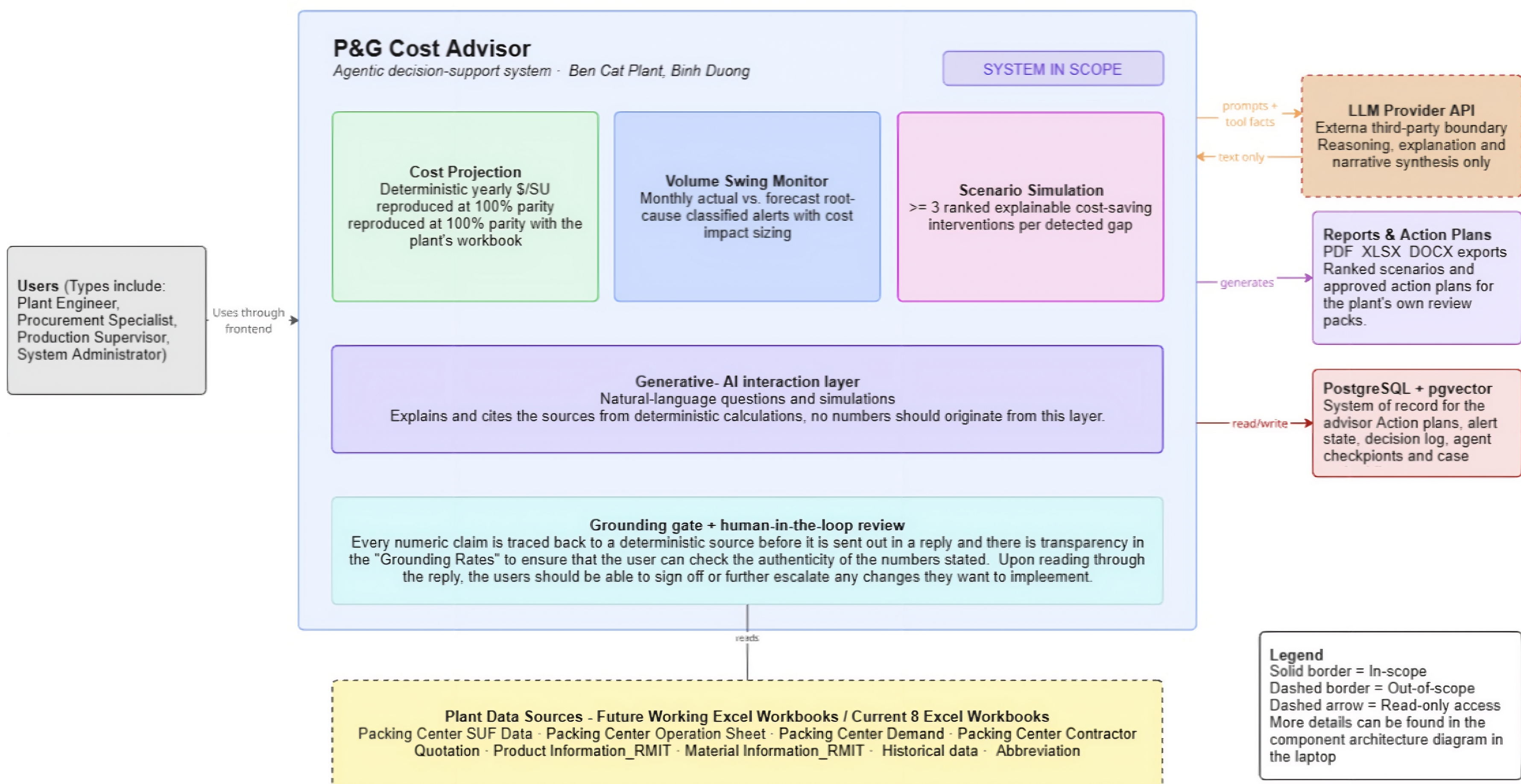
Turn a detected gap into ranked options. Rank at least 3 interventions across six goal categories, each naming the components or staffing changes involved, its projected annual saving alongside its potential trade-offs.

Refuse to assert hallucinations or unproven numbers. Route every numerical claim through verify_numbers.

Extended scope: Volume Swing Monitor. Monitor/detect deviations from a specific benchmark (5%) from the actual monthly volume swings

Methodology

The language model reasons and explains without ever calculating. Calculations are performed deterministically and verified before being sent out in a reply with users having the ability to check grounding rates.



1 Ingest and validate

Plant workbooks are mounted read-only and validated on load with pandera. Schema and unit drifts are rejected during ingestion.

2 Compute deterministically

A fixed-precision decimal.Decimal engine projects yearly \$/SU using the plant's own formula and models the stepped manpower hierarchy that makes labour cost non-linear against the throughput.

3 Reason, ground, escalate

The agent calls typed tools and retrieves comparable historical interventions. Any claim it cannot trace to a tool output is withheld and routed to a human reviewer.

C1 · NUMERIC AUTHORITY

Only the deterministic engine may originate a number

C2 · GROUNDING

No unverified claim reaches a user

C3 · HUMAN AUTHORITY

No action plan without human acceptance

C4 · SOURCE IMMUTABILITY

The system never writes to a plant source

Experiment & Results

Metric	Measured
Cost accuracy vs. Excel (20 SKUs)	100%
Scenario simulation (MAPE)	< 1%
Answer-grounding rate	>95%
Latency (median / p95)	10-15s

Stress testing with the industry supervisor allowed for consistency across usability, latency, and transparency of claims in the Agentic AI model. Cost accuracy was verified through weekly meetings and the simulation results were authenticated with the supervisor to ensure that the results were viable and made sense to the users at P&G. Full scope was achieved with the current prototype being tested by P&G.

Conclusion

Three findings shaped the system. The plant's KPI is cost per Statistical Unit rather than per item, and SU Factor can legitimately be zero, so every SU denominated metric needs a fallback. Manpower is added in discrete blocks, so cost per unit rises faster than volume falls and a linear labour model would misprice every scenario. Packaging changes on cheaper machines save labour while BOM simplification saves material, and merging the two produced recommendations the plant could not act on. The wider contribution is architectural: separating numerical authority from language generation produces a system its users can audit.

References

- [1] N. Bantilan, "pandera: Statistical data validation for pandas," Proc. SciPy, 2020, pp. 116–124.
- [2] G. Ke et al., "LightGBM," NIPS 30, 2017, pp. 3146–3154.
- [3] LangChain, "LangGraph 1.0 GA," 2025.
- [4] Python Software Foundation, "decimal — fixed-point arithmetic," 2025.
- [5] S. Makridakis et al., "M5 accuracy competition," Int. J. Forecasting, vol. 38, no. 4, 2022.