

Reducing Hallucinations in AI Agents for Legacy MES Data Analytics through Semantic Layer Design

Myunggun Kim, Sehee Moon, and **Hyunbo Cho***
Department of Industrial & Management Engineering.
Pohang University of Science & Technology (POSTECH)
City of Pohang, Korea

The Current MES Environment

- ❑ The MES dashboard provides visibility into current production status but lacks flexibility.
- ❑ Any dashboard modification requires IT involvement, introducing delays ranging from weeks to months.
- ❑ This limitation forces operators to manually construct SQL queries for ad hoc analysis.



AI Agent System

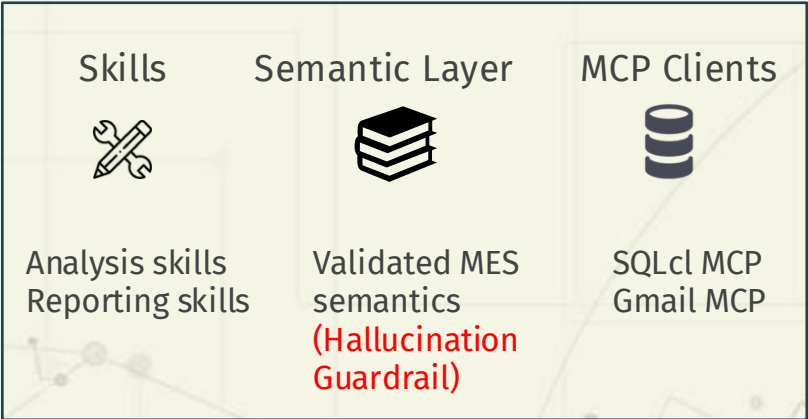
Operators



Prompts



CKP AI Agent



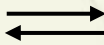
Cloud

MES Data Replica



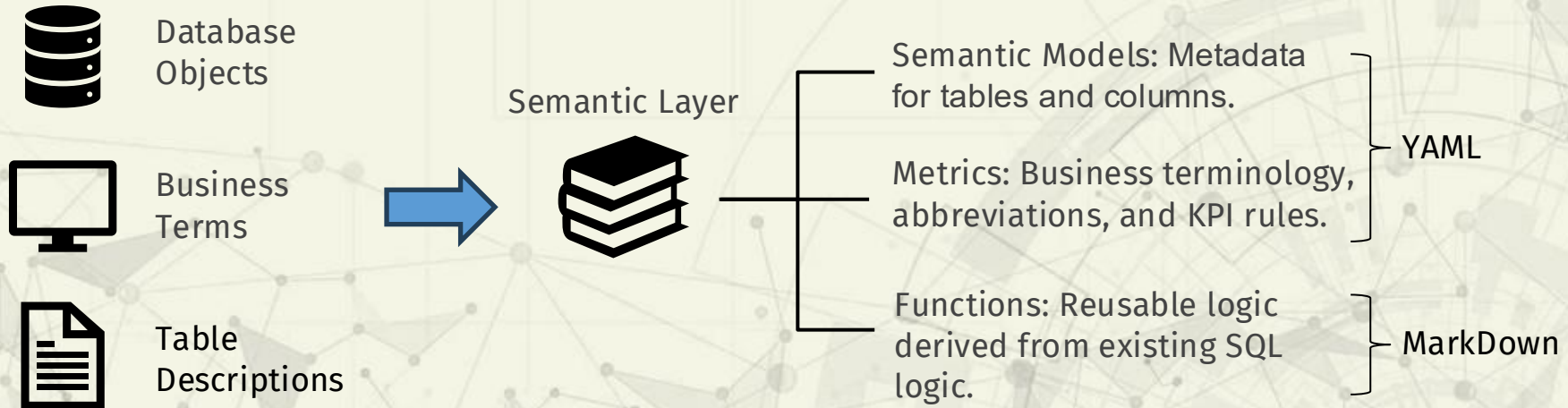
On-premise

MES data



How We Built the Semantic Layer

- ❑ We systematically collected MES database artifacts, including tables, queries, forms, and reports.
- ❑ Business terminology and KPI definitions were extracted from legacy SQL logic and dashboard metadata.
- ❑ This knowledge was formalized into YAML and Markdown to construct the semantic layer.



Experimental Design for Hallucination Evaluation

Prompt Complexity Factor

Level	Task	Output	Example
L1	Direct Lookup	Number	Retrieve yesterday's production amount for line A
L2	Grouped Aggregation	Table	Show a table of the size distribution by line on March 12
L3	KPI Reasoning	Table, Judgment	Identify the lines with an achievement rate below 70% on March 12
L4	Multi-metric Reasoning	Table, Derived Insights	Identify the model with the highest production on March 12, and show the achievement rate by line for that model.

Semantic Layer Configuration Factor

Level	Configuration	Total number of runs
V0	Baseline without the Semantic Layer	
V1	Semantic Models (Metadata for tables and columns)	$4 \times 4 \times 2$ (Prompt variations) \times
V2	V1 + Metrics (Business terminology, abbreviations, KPI rules)	10 (Replications) = 320
V3	V2 + Functions (Reusable logic derived from existing SQL logic)	

Semantic Layer Impact on Hallucination Reduction

- ❑ Each incremental enhancement of the semantic layer contributed to measurable gains in task accuracy and reliability.
- ❑ Hallucinations persisted in roughly 20% of experimental runs for V3; however, they can be mitigated by refining prompt specificity.
 - For instance, ambiguity can be reduced by explicitly specifying aggregation logic: “Show a table of the size distribution by line on March 12” → “Show a table of the size distribution by line on March 12, aggregated by total production quantity for each size category.”

