Chatting With Your Data: LLM-Enabled Data Transformation for Enterprise Text-to-SQL

Vulqan.ai[2] **Sara Mani Kapoor**[1] **Dr. Soumya Banerjee**[1]

[1]University of Cambridge, Department of Computer Science and Technology

[2]VULQAN.ai

vikas.kapoor@vulqan.ai smk78@cam.ac.uk sb2333@cam.ac.uk

November 25, 2025

Abstract

Enterprise SQL databases are large, semantically irregular, and often poorly documented posing major challenges for natural language interfaces. Large Language Models (LLMs) achieve high accuracy on academic Text-to-SQL benchmarks but fail dramatically in real operational environments due to schema ambiguity and missing business context. We introduce MAIA (Management Abstraction and Intelligence Algorithm), an LLM-enabled data transformation and agentic reasoning system that converts fragmented multi-source schemas into semantically enriched Logical Data Models (LDMs) represented in Knowledge Graphs (KGs). This work demonstrates that improved schema intelligence and semantic representation is key to making enterprise Text-to-SQL practical.

Introduction

Real-world databases contain hundreds of entities, cryptic column names, implicit relationships, unnormalized list values, and missing foreign keys. Current Text-to-SQL methods operate directly on such schemas, forcing LLMs to infer business logic from frenetic representations. We instead reframe Text-to-SQL as a knowledge representation problem: raw Source Data Models (SDMs) are transformed into a semantically aligned LDM, then navigated with a structured multi-agent reasoning pipeline. Our central claim is that adding semantic structure enables LLMs to perform accurate schema linking, reduce hallucination, and generate business-valid SQL.

LLM	Use Case	Level 0	Level 1	Level 2	Level 3	Overall
Phi-4-14B	Ticket Management	83%	38%	0%	0%	14%
LLaMA-3-70B	Ticket Management	83%	0%	7%	0%	12%
GPTo3-mini	Ticket Management	100%	13%	7%	0%	15%

Table 1: LLM Accuracy by Difficulty Level (Baseline)

LLM	Use Case	Level 0	Level 1	Level 2	Level 3	Overall
Phi-4-14B	Ticket Management	86%	71%	75%	53%	69%
LLaMA-3-70B	Ticket Management	100%	100%	86%	47%	78%

Table 2: LLM Accuracy by Difficulty Level (MAIA)

Figure: A summary of LLM accuracy on the Ticket Management use case by difficulty level and language model.

Motivation and Background

Though enterprise users (particularly non-tehcnical analysts) may wish to pose their data inquiries as questions in natural language rather than writing SQL, state-of-the-art LLMs often produce incorrect or misleading queries in production. Prior work highlights a large performance gap between academic benchmarks (80%+) and enterprise environments (¡10%). We argue this gap arises not from LLM reasoning limitations, but from missing semantic structure that business users take for granted.

Current Approaches and Limitations

Most Text-to-SQL benchmarks—such as Spider, WikiSQL, and BIRD—are far simpler than real enterprise data. Yu et al. [2018], Zhong et al. [2017], Li et al. [2023] Their schemas are small, cleanly designed, and manually curated with explicit keys, intuitive names, and minimal heterogeneity. Natural language questions in these datasets are crafted to align closely with SQL structure, avoiding the ambiguity and business-context interpretation required in production settings. As a result, current models and agent pipelines often overestimate their real-world capability: they perform well on curated benchmarks but struggle with enterprise schemas that are fragmented across systems, inconsistently named, missing relationships, and semantically misaligned.

Our Approach: Knowledge Graph–based Schema Intelligence

MAIA introduces a transformation layer that maps SDM tables into a canonical business ontology using VULQAN.ai's Canonical Data Model (CDM). The resulting LDM is represented as a Neo4j KG containing:

- Core objects: *Ticket, Employee, Project, Release, Environment*
- Rich metadata: semantic taxonomy, enumerated values, format standardization
- Synthetic keys + primary and composite keys + role-based relationships

The KG serves as an explicit reasoning space for LLM agents, reducing ambiguity before SQL generation begins.

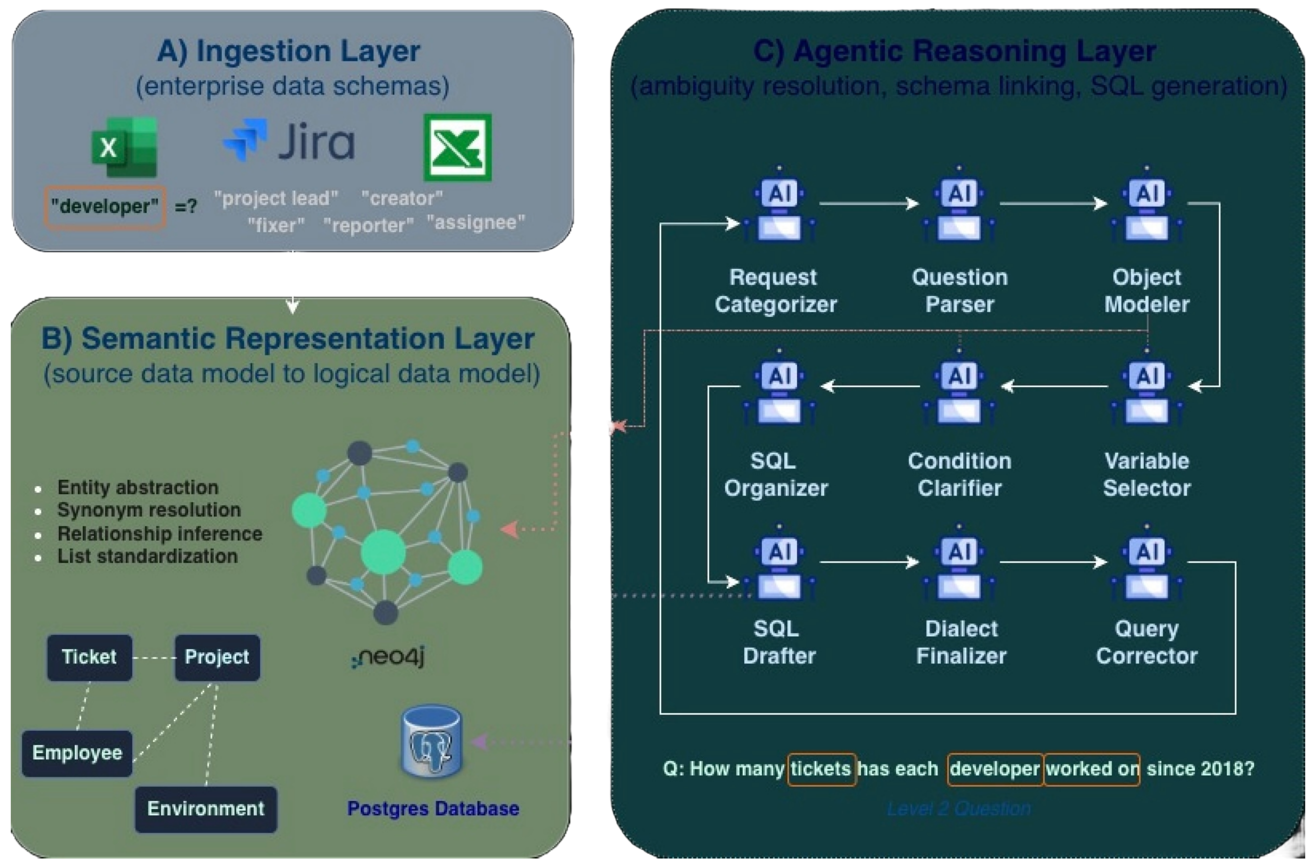

Figure: An overview of our MAIA framework.

Agentic Reasoning Pipeline

We design 9 specialized Query Reasoning Assistants that focus on different aspects of the schema linking process: **categorization** $\longrightarrow$ **object linking** $\longrightarrow$ **variable selection** $\longrightarrow$ **condition inference** $\longrightarrow$ **SQL assembly** $\longrightarrow$ **execution correction**. Agents operate *over the graph*, allowing the model to reason first and generate SQL only once the query is structurally defined.

Results and Validation

We validate MAIA on a real-world enterprise dataset from a software company managing its developer tickets over several years of growth. Baseline zero-shot prompting achieved only 10–15% semantic accuracy. MAIA improves to:

- **78% accuracy** with LLaMA-3-70B on complex enterprise questions
- Especially strong gains on joins, temporal logic, and ambiguous field names

Our findings demonstrate that **schema abstraction is more impactful than model scaling** in enterprise SQL reasoning.

Conclusions

MAIA establishes that enterprise Text-to-SQL requires business semantics in a unified representation. By injecting logical structure into the LLM's context and grounding reasoning in a graph-based LDM, the system becomes more interpretable, auditable, and controllable for enterprise deployment.

Future Directions and Impact

We envision extending MAIA toward:

- Automated alignment for enterprise data integration (Research currently under way at the University of Cambridge as a sequel to this work)
- Graph-guided self-reflection and execution-aware learning
- Scalable deployment in analytics platforms and BI tools

This work demonstrates a clear pathway toward trustworthy AI systems that enable non-technical users to *chat with their data*.

References

Jinyang Li, Binyuan Hui, Ge Qu, Jiaxi Yang, Binhua Li, Bowen Li, Bailin Wang, Bowen Qin, Rongyu Cao, Ruiying Geng, Nan Huo, Xuanhe Zhou, Chenhao Ma, Guoliang Li, Kevin C. C. Chang, Fei Huang, Reynold Cheng, and Yongbin Li. Can LLM Already Serve as A Database Interface? A BIg Bench for Large-Scale Database Grounded Text-to-SQLs, November 2023. URL `http://arxiv.org/abs/2305.03111`. arXiv:2305.03111 [cs].

Tao Yu, Rui Zhang, Kai Yang, Michihiro Yasunaga, Dongxu Wang, Zifan Li, James Ma, Irene Li, Qingning Yao, Shanelle Roman, Zilin Zhang, and Dragomir Radev. Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3911–3921, Brussels, Belgium, October-November 2018. Association for Computational Linguistics. doi: 10.18653/v1/D18-1425. URL `https://aclanthology.org/D18-1425/`.

Victor Zhong, Caiming Xiong, and Richard Socher. Seq2sql: Generating structured queries from natural language using reinforcement learning. *CoRR*, abs/1709.00103, 2017.

For more, please visit Vulqan.ai: `https://www.vulqan.ai/`