Addressing the Reliability and Security of LLMs usage with an Integrated Information Science and NLP Architecture

Bruno H. Brito[1], **Roney L. de S. Santos**[2]

[1]Departamento de Ciência da Informação (DCI) – Universidade Federal de São Carlos (UFSCar)
São Carlos, SP – Brazil

[2]Instituto de Ciência, Tecnologia e Inovação (ICTI) - Universidade Federal da Bahia (UFBA)
Camaçari, BA - Brazil

brunohenriquebrito@estudante.ufscar.br, roneysantos@ufba.br

Abstract. *This research addresses organizational data transformation by integrating Information Science (IS) and Natural Language Processing (NLP).We argue that Large Language Model (LLM) failures in data interaction tasks (instantiated via text-to-sql), such as hallucinations, stem from applications detached from information governance. We propose an interdisciplinary architecture employing a Knowledge Graph as an Intelligent Proxy to centralize metadata and ontologies. Validation through nine use cases demonstrates the architecture is functionally effective, generating accurate database queries matching ground-truth and identifying invalid relationships without exposing raw transactional data to the LLM.*

1. Introduction

The digital revolution has fundamentally transformed how organizations handle information. What was once limited to physical files has expanded into a vast digital ecosystem of documents, databases, and information systems. This explosion highlights an organizational paradox: despite the abundance of information, it rarely translates into actionable knowledge or tangible value. Although Natural Language Processing (NLP) tools, particularly Large Language Models (LLMs), are often presented as a solution, they frequently fail when applied to complex corporate environments.

We argue that this failure does not stem from technological limitations of NLP itself, but rather from a methodological shortcoming: the absence of multidisciplinarity. NLP systems are commonly developed within a computational silo, disregarding core principles from Information Science (IS) related to how knowledge is structured (Information Ecology), governed (DMBOK), and shared (SECI). This isolated approach introduces critical risks, namely: (1) Low reliability, in which LLMs hallucinate responses when confronted with semantic ambiguities; and (2) Security risks, as models often require direct access to sensitive data, thereby violating governance and compliance principles.

In this paper, we address this interdisciplinary gap by arguing that, for an LLM to generate actionable knowledge, it must interact with a knowledge layer curated and governed according to IS principles. The central objective of this work is to present an integrated IS+NLP architecture that operationalizes this perspective, employing a Knowledge Graph (KG) as an Intelligent Proxy. It is worth noting that while this research focuses

on the overarching challenge of secure, governed natural language interfaces to corporate data , the proposed architecture is empirically validated here using Text-to-SQL use cases as a structural baseline. To support community research and ensure reproducibility, the complete architectural implementation, source code and use cases are publicly available on Github [1].

To validate the proposed architecture, we conducted a rigorous qualitative evaluation **(detailed in Section 4)** based on nine use cases designed to simulate real-world corporate challenges. These use cases assess the system's performance in scenarios involving high query complexity, such as multiple JOIN operations across heterogeneous domains, semantic interpretation through ontological reasoning, and data governance enforcement via metadata-driven rules, including sensitive field masking. Additionally, controlled failure scenarios were included to evaluate the system's ability to recognize the absence of valid relationships between data sources and to explicitly refuse the generation of unsupported SQL queries. Together, these use cases cover complete data journeys, meta-knowledge querying about relational structures, and the generation of comprehensive analytical reports grounded in both ontological knowledge and relational integrity.

Our results demonstrate that the proposed IS+NLP architecture successfully achieves its objectives. Specifically, the system (1) resolves semantic ambiguity by translating business concepts into precise SQL queries; (2) ensures reliability by recognizing the absence of valid data connections and actively refusing to generate unsupported queries; and (3) assures security by restricting the LLM's interaction to a metadata proxy rather than exposing sensitive data.

The remainder of this article is organized as follows. **Section 2** reviews the theoretical foundations, contrasting the traditional silos of Information Science and NLP. **Section 3** details the proposed interdisciplinary architecture. **Section 4** presents the experimental evaluation and qualitative analysis of the results. Finally, **Section 5** discusses limitations, future research directions, ethical considerations, and concludes the article.

2. Theoretical Foundation and Related Work

The difficulty in addressing the Information Paradox does not arise from a lack of technological tools, but rather from a lack of integration between complementary knowledge domains. Current literature and industrial practice largely operate within two silos that rarely intersect: (1) the Information Science (IS) or business-oriented silo, and (2) the Natural Language Processing (NLP) silo. The analytical perspective that guides this work, illustrated in Figures 1 and 2, seeks to bridge these domains by establishing a shared conceptual foundation.

Identifying prior work that simultaneously addresses these conditions remains challenging. In practice, corporate solutions often prioritize immediate applicability without explicit theoretical grounding, while academic research tends to formalize foundations without operational deployment. As a result, there is a limited body of work that explicitly proposes and validates integrated solutions of this nature. In this context, the related work discussed in this section draws both from well-established academic frameworks and from accumulated experience in real-world corporate environments, positioning this article as a contribution aimed at narrowing this gap.

[1] Available at `https://tinyurl.com/38c48jkc`

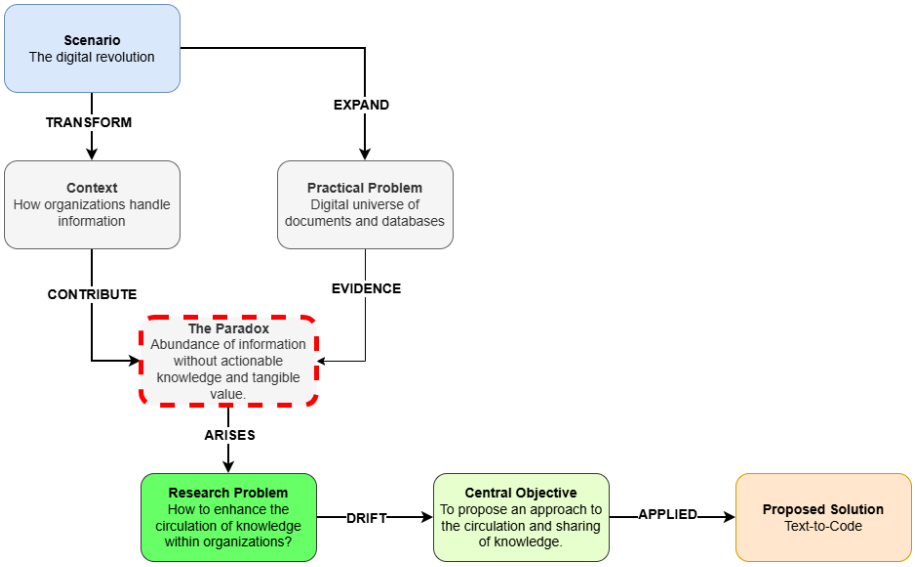

Figure 1. Problem analysis (Source: Author)

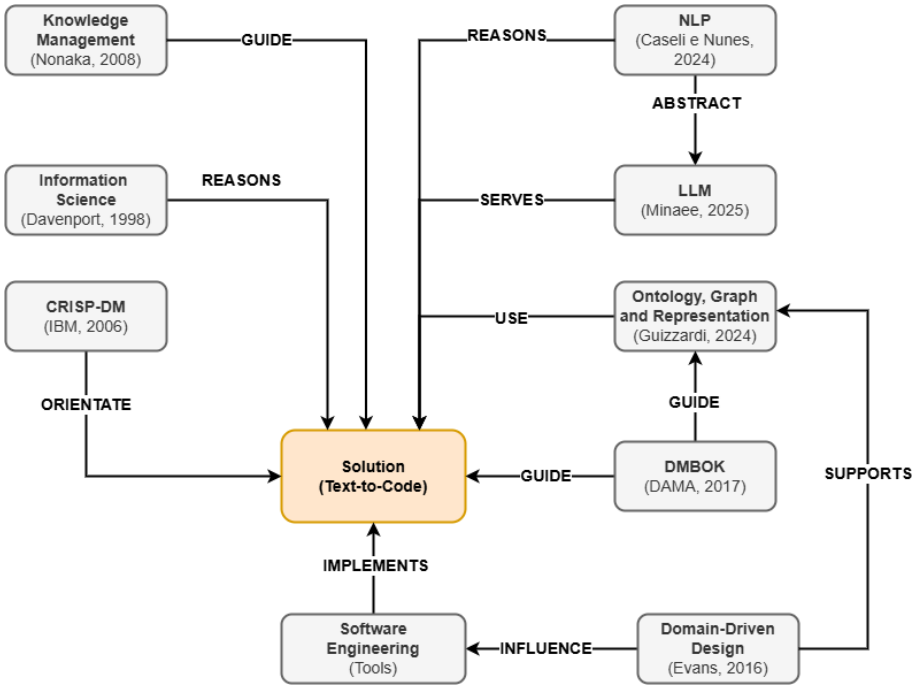

Figure 2. Problem statement and conceptual foundation (Source: Author)

2.1. The Information Science and Governance Silo: The Structure of Knowledge

Information Science provides essential models for understanding how knowledge is created, structured, and flows within organizations. The concept of Information Ecology [Davenport 1997] emphasizes that information cannot be dissociated from its human, organizational, and social context. The SECI model [Nonaka and Takeuchi 2008] formalizes the knowledge life cycle, encompassing Socialization, Externalization, Combination, and Internalization processes. In parallel, market-oriented frameworks such as DAMA-DMBOK [DAMA International 2017] and Domain-Driven Design [Evans 2016] operationalize these principles through governance, data quality, and domain modeling prac-

tices. However, when considered in isolation, these frameworks do not provide automated mechanisms for transforming structured knowledge into actionable outputs at scale.

2.2. The NLP Silo: Technical Tools without Anchorage

In parallel, the NLP field offers a rich set of technical tools [Caseli and Nunes 2024], culminating in the development of Large Language Models [Minaee et al. 2025]. Predominantly NLP-centric approaches, such as Text-to-SQL benchmarks, frame the problem as a statistical translation task focused on maximizing syntactic accuracy. However, as highlighted in the literature, these models primarily learn patterns of co-occurrence rather than explicit semantic structures. This lack of semantic anchoring manifests as hallucination. In corporate environments, a hallucinated SQL query represents not only a technical flaw, but also a significant operational and governance risk. To mitigate LLM hallucinations limitations, Retrieval-Augmented Generation (RAG) is commonly implemented. RAG enhances model capabilities by querying an external data source to retrieve relevant context before appending it to the user's prompt, ensuring grounded responses. In this research, this concept is adapted to a structured metadata environment, where the retrieval stage queries a Knowledge Graph instead of unstructured text to support the Text-to-SQL task.

2.3. The Gap

The existing literature reveals a clear lack of approaches that effectively integrate these two silos. We argue that improvements in reliability will not arise solely from increasingly larger models, but from interdisciplinary architectures that anchor NLP systems in the conceptual and governance foundations of Information Science. In this context, Semantic Web technologies, including ontologies, knowledge graphs, and formal knowledge representation [Guizzardi and Guarino 2024], play a central role. This work addresses this gap by proposing and validating such an architecture, demonstrating that the frameworks discussed in **Section 2.1** are not merely theoretical constructs, but essential components of a robust engineering solution.

3. Methodology

To address the interdisciplinary gap identified in **Section 2.3**, this work adopts a design-oriented methodology focused on the specification and validation of an integrated IS+NLP architecture. Rather than proposing a purely algorithmic solution, the methodology emphasizes architectural design grounded in Information Science principles and evaluated through controlled use cases. **Figure 3** [2] presents an overview of the proposed solution.

The proposed architecture is not conceived as a linear processing pipeline, but as an anchored knowledge system. Its core component is the Intelligent Proxy, which combines a Knowledge Graph responsible for semantic representation and governance, and an Orchestrator responsible for controlling interactions between the user, the knowledge layer, and the language model. This separation of responsibilities is essential to ensure reliability, semantic grounding, and data security. To implement this architecture, the methodology is organized into two complementary stages.

[2]Resized image available at `https://tinyurl.com/23jdsb7k`

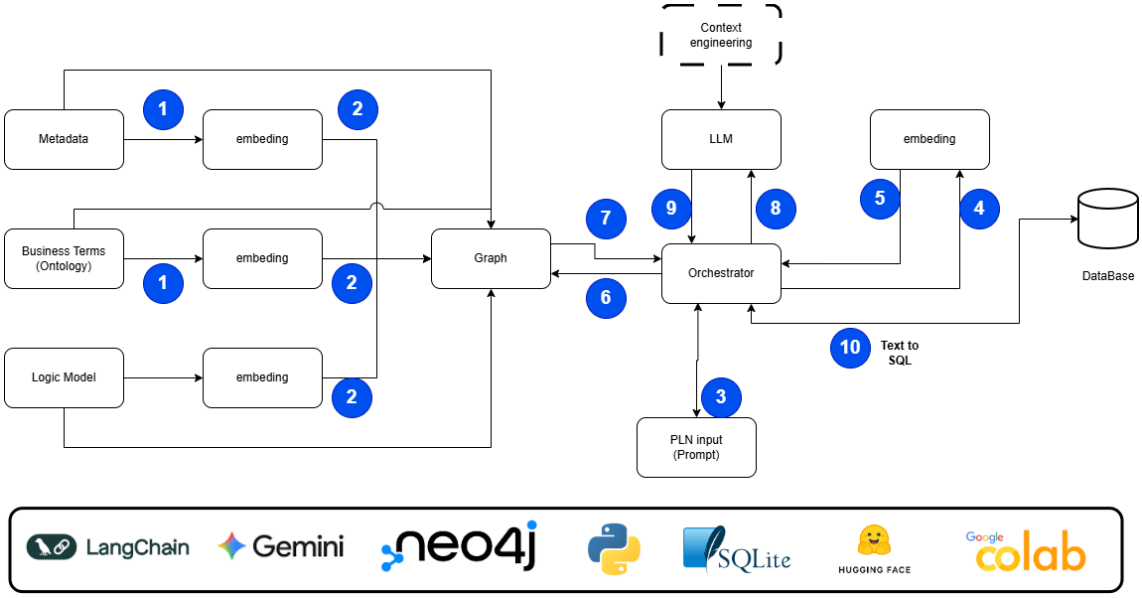

Figure 3. Solution proposal and architectural overview (Source: Author)

3.1. Stage 1: Knowledge Structuring - The IS Cognitive Load

The first stage focuses on translating organizational knowledge, resulting from the Socialization and Externalization processes of the SECI model, into a machine-processable representation. This stage concentrates the cognitive and governance load defined by Information Science principles.

1. **Inputs:** Organizational knowledge artifacts are collected **(Step 1 in Figure 3)**, including the logical data model (database schema), business terms expressed as ontologies and semantic definitions, and metadata representing governance rules derived from DMBOK principles.
2. **Graph construction:** These artifacts are not directly exposed to the LLM. Instead, they are processed, vectorized using sentence-level embeddings, and integrated into a Knowledge Graph implemented in Neo4j[3] **(Step 2)**. This graph constitutes the Intelligent Proxy, providing a governed, semantic, and queryable representation of organizational knowledge.

3.2. Stage 2: Knowledge Activation

The second stage instantiates the Combination and Internalization processes of the SECI model and is managed by the Orchestrator, implemented using the LangChain framework[4]. This stage governs how user intent is translated into controlled interactions with the knowledge layer and the language model, following the sequence illustrated in **Figure 3**.

1. **User input (Step 3):** The process begins when the user submits a query in natural language through the NLP interface.

[3]Available at `https://neo4j.com/`.
[4]Available at `https://www.langchain.com/`.

2. **Schema awareness (Steps 4 and 5):** In parallel, the Orchestrator retrieves the physical database schema and its corresponding embeddings, ensuring that structural constraints are available for subsequent reasoning.
3. **Prompt interception (Step 6):** The Orchestrator intercepts the user prompt and assumes full control of the interaction flow, preventing direct access to the language model.
4. **Semantic grounding via the Knowledge Graph (Step 7):** The Orchestrator queries the Knowledge Graph using embedding-based similarity search to retrieve the most relevant metadata, business terms, and relational structures associated with the user's semantic intent.
5. **Context construction (Step 8):** Based on the retrieved graph context and the database schema, the Orchestrator constructs a controlled and semantically grounded context to guide the language model.
6. **LLM interaction (Step 9):** The Orchestrator sends an augmented prompt, combining the original user query with the constructed context, to the language model, Google Gemini [Google DeepMind 2024]. The LLM has access only to this mediated information.
7. **Text-to-SQL generation and execution (Step 10):** The language model generates a Text-to-SQL statement, which is validated and executed by the Orchestrator against the SQLite database.

This methodology deliberately constrains the operational boundaries of the LLM. At no point does the model access raw data, whether sensitive or non-sensitive. Instead, it reasons exclusively over governed metadata and semantic representations mediated by the Intelligent Proxy. This design choice directly supports the evaluation criteria presented in the next section, where the architecture is assessed in terms of semantic reliability, controlled failure behavior, and data governance compliance.

4. Experimental Evaluation and Results

To validate the central thesis that an interdisciplinary IS+NLP architecture addresses the limitations of pure NLP approaches, we conducted a qualitative experimental evaluation. The objective of this evaluation was to assess the architecture's effectiveness in complex corporate scenarios, with particular emphasis on reliability, semantic interpretation, and data security.

It is important to note that this work does not aim to outperform existing Text-to-SQL models in terms of quantitative accuracy metrics. Instead, the evaluation focuses on architectural properties that are not adequately captured by standard quantitative benchmarks, such as semantic grounding, controlled failure behavior, and governance compliance. In this context, a qualitative evaluation is more appropriate, as the primary research contribution lies in demonstrating how interdisciplinary design choices influence system reliability and safety rather than in optimizing statistical performance.

4.1. Experimental Setup

The experimental evaluation was conducted using a functional prototype of the proposed architecture, as described in Section 3, and a set of nine use cases designed to simulate realistic business questions. These use cases represent different levels of complexity and

semantic ambiguity and are summarized in Figure 4. The evaluation procedure consisted of comparing the SQL queries generated by the system, along with their explanations, against predefined golden queries that represent the expected ground truth.

Use Case	Goal	Expected Result	Sucess?
Use Case 1	Test multiple JOINs between tables from different domains.	Generate a SQL query that consolidates data from postings, accounts, and cost centers.	Yes
Use Case 2	Test the system with a non-existent data connection between domains.	The LLM must recognize the absence of a connection and not generate the SQL query.	Yes
Use Case 3	Validate again the ability to recognize the lack of connection between tables.	The LLM must indicate the impossibility of generating the SQL query due to the lack of a relationship.	Yes
Use Case 4	Demonstrate the use of the ontology to interpret business concepts.	The LLM must use the ontology to translate the concept "flying products" into a functional query.	Yes
Use Case 5	Test data governance capability (field masking).	The system must mask a sensitive field based on metadata.	No
Use Case 6	Test a complete data journey, with multiple JOINs for a 360° view.	The system must generate a complex SQL query with LEFT JOINs that includes all products.	Yes
Use Case 7	Test the ability to query meta-knowledge.	The LLM must reason about the Product table's relationships and return a count.	Yes
Use Case 8	Test the ability to query meta-knowledge for all relationships.	The system must provide a complete overview of the Product table's relationships.	No
Use Case 9	Test the ability to use the ontology and JOINs for a complete product report.	The LLM must generate a query with JOINs based on the ontology for a complete report.	Yes

Figure 4. Overview of the evaluated use cases (Source: Author)

4.2. Learning 1: Reliability and Recognition of Absence of Connections

A critical limitation of pure LLM-based systems is their tendency to hallucinate answers when confronted with plausible questions that lack a valid representation in the underlying data model. To assess this behavior, we evaluated the system using queries that appeared semantically reasonable but did not correspond to any explicit relationship in the logical schema. In Use Cases 2 and 3, the system was queried about non-existent relationships, such as associations between employees and cost centers.

Concrete Result: As illustrated in Figure 5 [5], the system correctly identified the absence of an explicit relationship path in the Knowledge Graph. In accordance with the architectural constraints, the LLM responded by explicitly refusing to generate a SQL query, stating that it was not possible to establish a direct connection between the requested data entities.

This result demonstrates the architecture's ability to prioritize reliability over superficial accuracy. Rather than producing a syntactically valid but semantically incorrect query, the system explicitly acknowledges the absence of sufficient information, thereby avoiding operational and governance risks.

4.3. Learning 2: Semantic Efficiency (Intent Translation)

The second learning focuses on the system's ability to translate user intent, expressed in business terms, into a precise technical action. This capability reflects the integration between Information Science concepts and NLP mechanisms. In Use Case 4, the user query requested a list of "flying products," a term that does not exist explicitly in the SQLite database[6] schema.

[5]Resized image available at `https://tinyurl.com/4p6jv3uu`
[6]Available at `https://www.sqlite.org/`

```
--- SQL QUERY GENERATED BY THE LLM ---

IT WAS NOT POSSIBLE TO GENERATE THE SQL QUERY. BASED ON THE AVAILABLE SCHEMA,
THE CONNECTION BETWEEN THE DATA IS NOT DIRECT.

Please, review the logical model or the question.


--- END OF QUERY ---
```

Figure 5. LLM response indicating the absence of a valid data connection (Source: Author)

Concrete Result: The Orchestrator leveraged the embedding representation of the user prompt to query the Knowledge Graph (Step 7) and identified relevant ontology-based concept nodes, including "Flying Equipment" and "Flying Backpack." By injecting this semantic context into the augmented prompt, the system generated the correct SQL query, correctly inferring that the expression "flying products" corresponded to categories matching the condition Nome_Categoria LIKE '%Voador%'. Figure 6 [7] illustrates this behavior in the prototype.

```
--- SQL QUERY GENERATED BY THE LLM ---

SELECT T1.Nome_Produto, T1.Descricao_Produto
FROM Produto AS T1
INNER JOIN Categoria_Produto AS T2
    ON T1.ID_Categoria_Produto = T2.ID_Categoria_Produto

WHERE T1.Nome_Produto LIKE '%Voador%';



--- END OF QUERY ---
```

Figure 6. LLM-generated query for semantic interpretation in Use Case 4 (Source: Author)

This result highlights the role of the Knowledge Graph in enabling semantic reasoning beyond literal schema matching, allowing the system to bridge the gap between business language and technical implementation.

[7] Resized image available at `https://tinyurl.com/hruw35w7`

4.4. Learning 3: Security and Privacy by Design (Metadata Proxy)

The architecture was explicitly designed to ensure that the language model never accesses raw data, regardless of its sensitivity. Throughout all nine evaluated use cases, the LLM interacted exclusively with metadata and semantic representations stored in the Knowledge Graph.

Concrete Result: Across all nine use cases, the LLM interacted exclusively with the Knowledge Graph and the metadata embeddings stored within it. The Orchestrator was responsible for generating the SQL query, which was executed separately on the database only after its construction (Step 10). This architectural separation constitutes the core security and privacy contribution of the proposed approach, as it relies solely on metadata for reasoning. By design, the architecture mitigates the risk of data exfiltration and preserves privacy, since the LLM operates over governed metadata rather than accessing the underlying data itself.

5. Conclusions

This work originated from the question of how to enhance the circulation of knowledge within organizations in the presence of the so-called Information Paradox. While the dominant approach in the NLP literature has relied on computational scale through Large Language Models, this strategy, when applied in isolation, fails to guarantee reliability and security in corporate environments. In response, we proposed and validated an interdisciplinary architecture that supports a different thesis: the effective generation of actionable knowledge does not depend on increased computational force, but on the integration of NLP technologies with the robust methods, processes, and governance frameworks of Information Science, including SECI, DMBOK, and ontological modeling.

We conclude that Information Science should not be treated as an auxiliary component, but as the foundational layer for reliable NLP-based systems. By using IS principles to model, govern, and structure organizational knowledge prior to any interaction with the language model, the proposed Intelligent Proxy architecture demonstrated its ability to translate semantic intent, preserve data privacy, and, critically, recognize the limits of its own knowledge. In doing so, this work answers its research question by affirming that knowledge potentiation emerges when the expressive power of NLP is anchored in the discipline and governance of Information Science.

Reinforcing the arguments presented in the introduction, the main contributions of this research can be summarized as follows:

1. **Effective translation of semantic intent:** We demonstrated that Knowledge Graphs provide the missing bridge between Information Science and NLP. By anchoring the LLM in a graph enriched by ontological structures, the system was able, in Use Case 4, to translate tacit knowledge and business terminology, such as “flying products”, into actionable SQL queries that could not be derived from the relational schema alone.
2. **A reliability paradigm based on controlled failure:** The proposed architecture ensures reliability by explicitly recognizing the absence of valid data relationships. As demonstrated in Use Cases 2 and 3, instead of producing hallucinated queries, a common failure in pure NLP approaches, the system actively refused to respond, prioritizing correctness and trustworthiness over false accuracy.

3. **Security and privacy by design:** We presented and validated an architectural model in which the LLM reasons exclusively over governed metadata and never accesses transactional data, whether sensitive or not. The separation between semantic reasoning, performed by the proxy, and query execution, performed at the database layer, constitutes a practical and ethical contribution that supports compliance with data protection regulations such as the Brazilian General Data Protection Law (LGPD) and facilitates the responsible adoption of LLMs in corporate environments.

5.1. Limitations and Future Work

Despite the qualitative success of the proposed interdisciplinary architecture, this work presents limitations that define directions for future research. The primary limitation concerns the scope of the experimental evaluation. While the nine use cases are representative and demonstrative, they do not constitute an exhaustive quantitative assessment, nor do they include large-scale comparisons against human-generated golden queries. In addition, the experiments did not explore scenarios involving very high query complexity or fully active data governance mechanisms, such as dynamic masking, which was only partially addressed.

Future work will address these limitations through three main directions: (1) the execution of rigorous quantitative evaluations using larger and more diverse datasets; (2) the integration of unstructured data sources, such as documents and organizational wikis, to further enrich the Knowledge Graph; and (3) the expansion of active governance mechanisms, potentially through the integration of advanced policy frameworks such as ODPS (Open Data Product Specification)[8].

5.2. Ethical and Security Considerations

The adoption of Large Language Models in corporate environments raises significant ethical and security concerns, particularly with respect to data privacy and regulatory compliance. This work contributes directly to this discussion by proposing an architectural approach that mitigates these risks by design. Unlike pure NLP approaches, which often require direct access to sensitive transactional data, the proposed IS+NLP architecture restricts the LLM's interaction to a governed metadata layer.

Query execution occurs exclusively within a controlled database environment, fully separated from the reasoning process. We argue that this architectural separation between semantic reasoning and data execution constitutes a fundamental ethical requirement for the responsible use of language models in organizational contexts, particularly in environments subject to strict data protection regulations.

References

Caseli, H. M. and Nunes, M. G. V. (2024). *Processamento de Linguagem Natural: Conceitos, Técnicas e Aplicações em Português*. BPLN, 2 edition.

DAMA International (2017). *DAMA-DMBOK: Data Management Body of Knowledge*. Technics Publications, 2 edition.

[8]Available at `https://opendataproducts.org/`.

Davenport, T. H. (1997). *Information Ecology*. Oxford University Press.

Evans, E. (2016). *Domain-Driven Design*. Addison-Wesley.

Google DeepMind (2024). Gemini: A family of highly capable multimodal models. Accessed: 2026-01.

Guizzardi, G. and Guarino, N. (2024). Explanation, semantics, and ontology. *Data & Knowledge Engineering*.

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., and Gao, J. (2025). Large language models: A survey. arXiv:2402.06196.

Nonaka, I. and Takeuchi, H. (2008). *The Knowledge-Creating Company*. Harvard Business School Press.