From Conceptual to Logical Schema: An LLM-based Approach

Hudson Afonso Batista da Silva [1], Vitória Soares dos Santos [1], Carina Friedrich Dorneles [1], Ronaldo dos Santos Mello [1]

[1]Instituto de Informática e Estatística – Universidade Federal de Santa Catarina
Florianópolis – Santa Catarina – Brazil

Abstract. *The use of language models for solving computational tasks has grown in recent years. Generating logical models from conceptual modeling (ER) is a specialized task with well-defined rules, performed by database specialists. This paper presents an experimental investigation into the capabilities of large language models (LLMs), including GPT, Gemini, and Claude, in generating logical models from conceptual representations. The results demonstrate that the Claude model exhibits superior results in addressing the proposed task.*

1. Introduction

Large Language Models (LLMs) have emerged as a promising approach for a variety of tasks, due to their ability to generalize and adapt across multiple domains. These models operate over large volumes of data and have shown consistent performance in applications involving interpretation, transformation, and generation of information at different levels of structure [Yao et al. 2024]. This flexibility continues to motivate research into new application domains where structured data plays a central role.

Among these domains, the database area has attracted increasing attention. [Li et al. 2024] highlights several opportunities to apply LLMs in data management activities, including data processing, query optimization, and analytical tasks. Other studies propose frameworks for translating natural language into SQL queries, aiming to simplify user interaction with database systems [Visperas et al. 2023, Liu et al. 2025, Singh et al. 2025].

Database design is a fundamental activity for any data-centric application. It accomplishes a process that guides the modeling of the application data aiming at producing a much suitable as possible data schema that provides efficient data storage and access. Elmasry define a simplified overview of the classic database design process in 4 steps: *i) analyzing the requirements,ii) conceptual design or Entity-Relationship Diagram, iii) database schema or logical model and iv) physical design* [Elmasri and Navathe 2016].

[Mylopoulos et al. 2025] show the necessity of Understanding how to leverage LLMs as a tool to support different phases of the conceptual modeling process is still an open issue. In this work, we evaluate the ability of LLMs in the task of ER mapping, generating logical schemas *(step iii)* from ERD image input *(step ii)*. This paper is structured as follows: Section 1 presents the Introduction, Section 2 describes related works, Section 3 discuses the methodology, Section 4 show the experiments and evaluation, Section 5 discusses the conclusions, and Section 6 lists the references.

2. Related Works

In order to identify related work to our proposal, we conducted a systematic mapping of the literature using the following search string: *'LLM' OR 'Large Language Models' AND 'Conceptual Model' OR 'ER model' OR 'entity-relationship' OR 'conceptual design' AND 'Logical Schema' OR 'schema'*. The search was performed in the Scopus research database, complemented by forward and backward snowballing. We analyzed a total of 45 articles: 15 retrieved from the research database and 25 identified through snowballing, of which 3 were selected (Table 1). We included articles written in English, published in the last five years in conferences, journals and preprints (arXiv). Dissertations, theses, and undergraduate final projects were excluded.

[Salem et al. 2025] proposes the generation of database schemas from requirement specifications based on natural language processing. However, it does not provide clear evaluation metrics nor a code repository. Moreover, it does not specify which LLMs models were used, nor the prompts and parameters, making reproduction of the results unfeasible. The study uses five different datasets (ER diagrams) but does not detail the prompting techniques applied. [Avignone et al. 2025] sugest a framework employing LLMs to analyze a logical schema and generate textual descriptions of the initial problem requirements. The models used include GPT, Mistral, Gemini, and LLaMA. Evaluation metrics comprise F1-score, perplexity, and ROUGE. A total of 13 datasets were employed. Prompting techniques included zero-shot, n-shot, and chain-of-thought. Finally, [Härer 2023] addresses UML modeling from textual requirements. It employs GPT-4 and LLaMA 2, with the output represented as a UML diagram. The evaluation is generalist in nature, consisting of a simple comparison between the results of the two models, without well-defined metrics.

Table 1. Comparative Analysis of LLM-based Approaches

Work	Input → Output	Models Used	Evaluation Metrics	Datasets
[Salem et al. 2025]	Text requirements or ERD image → Schema (entities, attributes, relationships)	Not reported	Total number of correct predictions	5
[Avignone et al. 2025]	ERD → Text (requirements specifications)	GPT, Mistral, Gemini, LLaMA	F1-score, Perplexity, ROUGE	13
[Härer 2023]	Text → UML Schema	GPT-4, LLaMA 2	General comparison across models	1
Our Work	**ERD image → Schema (entities, attributes, relationships)**	**GPT-4o, GPT-5, Gemini, Claude**	**F1-score, Accuracy**	**1**

Our main contribution to the research field lies in the evaluation of several LLMs models for generating logical schemas from an input image of an ERD, providing evaluation metrics such as F1 Score and Accuracy.

3. Methodology

We employed three language models (GPT, Gemini and Claude) through each model's API, using an ERD image as input and generating a JSON output (entities, attributes, primary keys, and foreign keys), as illustrated in Figure 1. Two prompts were executed: i) without using mapping rules that consider relationship cardinalities, and ii) with the addition of such rules. The code and prompts can be accessed at https://github.com/hudsonsilva12/eramia.

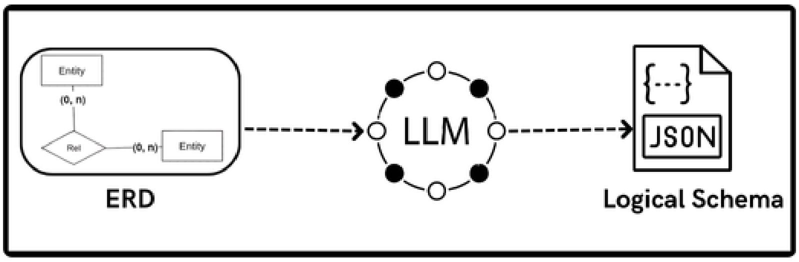

Figure 1. Proposed Approach

4. Experiments and Results

For the evaluation of the models, we analyzed the outputs of the JSON files. We calculated the total number of entities, attributes, primary keys (PK), and foreign keys (FK) ($N_{entities}$, $N_{attributes}$, N_{PK}, N_{FK}) returned by each model. These values were compared against the corresponding ground-truth counts reported in company Schema [Elmasri and Navathe 2016]. Based on these comparisons, we computed two performance metrics: the F_1 Score, and Accuracy. These metrics quantitatively assess the models' ability to generate logical schemas consistent with the expected structures, as summarized in Table 2.

Table 2. Evaluation of LLM outputs for schema generation.

Prompt / Model	**F1-Score / Accuracy**
Prompt 1 - gpt-4o	77.26% / 63.01%
Prompt 2 - gpt-4o	95.50% / 92.16%
Prompt 1 - gpt-5	94.86% / 91.07%
Prompt 2 - gpt-5	94.90% / 91.07%
Prompt 1 - claude-opus-4-1-20250805	**96.77% / 94.12%**
Prompt 2 - claude-opus-4-1-20250805	96.77% / 94.12%
Prompt1- gemini-2.5-flash	94.86% / 91.07%
Prompt2- gemini-2.5-flash	95.83% / 92.59%

5. Conslusion

LLMs are effective tools for ER mapping. The Claude Opus 4.1 model demonstrated superior capability compared to the other models, while the GPT-4o model showed the lowest accuracy when using a simple prompt without rules. Therefore, there is strong feasibility for the use of LLMs in specific database design tasks, in this study, specifically for ER mapping.

This work represents research in its initial stages. Experiments were conducted with a limited number of samples and models. In future work, we intend to expand this research by incorporating a larger number of samples and LLMs models, evaluate generation of Physical Design and NoSQL logical models.

References

Avignone, A., Tierno, A., Fiori, A., and Chiusano, S. (2025). Exploring large language models' ability to describe entity-relationship schema-based conceptual data models. *Information*, 16(5):368.

Elmasri, R. and Navathe, S. B. (2016). Fundamentals of database systems seventh edition.

Härer, F. (2023). Conceptual model interpreter for large language models. *arXiv preprint arXiv:2311.07605*.

Li, G., Zhou, X., and Zhao, X. (2024). Llm for data management. *Proceedings of the VLDB Endowment*, 17(12):4213–4216.

Liu, X., Shen, S., Li, B., Ma, P., Jiang, R., Zhang, Y., Fan, J., Li, G., Tang, N., and Luo, Y. (2025). A survey of text-to-sql in the era of llms: Where are we, and where are we going? *IEEE Transactions on Knowledge and Data Engineering*.

Mylopoulos, J., Guizzardi, G., and Guarino, N. (2025). Conceptual modeling: Foundations, a historical perspective, and a vision for the future. *Data & Knowledge Engineering*, 160:102483.

Salem, N., Al-Tarawneh, K., Hudaib, A., Salem, H., Tareef, A., Salloum, H., and Mazzara, M. (2025). Generating database schema from requirement specification based on natural language processing and large language model. *COMPUTER*, 16(7):1703–1713.

Singh, A., Shetty, A., Ehtesham, A., Kumar, S., and Khoei, T. T. (2025). A survey of large language model-based generative ai for text-to-sql: Benchmarks, applications, use cases, and challenges. In *2025 IEEE 15th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 00015–00021. IEEE.

Visperas, M., Adoptante, A. J., Borjal, C. J., Abia, M. T., Catapang, J. K., and Peramo, E. (2023). On modern text-to-sql semantic parsing methodologies for natural language interface to databases: A comparative study. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pages 390–396. IEEE.

Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., and Zhang, Y. (2024). A survey on large language model (llm) security and privacy: The good, the bad, and the ugly. 4(2):100211.