From Questions to Answers: A Natural Language Interface for DATASUS Hospitalization Data

Maicon Moraes[1,2], Isadora Figueiredo[2], Victória Marques[2], Juliano Santos[2], Isabel H. Manssour[2]

[1]UFRGS - Federal University of Rio Grande do Sul, Porto Alegre, Brazil

[2]PUCRS - Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, Brazil

maicon.kevyn@ufrgs.br, isabel.manssour@pucrs.br

Abstract. *To democratize access to Brazil's DATASUS hospitalization data, we built a conversational agent that provides natural language responses to Portuguese questions. The agent uses a LangGraph and Llama 3.1:8B workflow to translate questions into SQL, achieving 100% end-to-end success and 82.9% execution accuracy on a 52-query benchmark. However, performance on complex queries dropped to 58.3%, revealing limitations in the LLM's multi-step reasoning for multi-table analytical operations. Our work validates this stateful agentic approach and establishes a reproducible baseline for future extensions, including model benchmarking, domain-specific fine-tuning, and query decomposition strategies.*

Resumo. *Para democratizar o acesso aos dados de hospitalização do DATASUS, criamos um agente conversacional que fornece respostas em linguagem natural para perguntas em português. O agente usa um fluxo com LangGraph e Llama 3.1:8B para traduzir perguntas em SQL, alcançando 100% de sucesso de ponta a ponta e 82,9% de acurácia de execução em um benchmark de 52 consultas. Contudo, o desempenho em consultas complexas caiu para 58,3%, revelando limitações no raciocínio de múltiplos passos para operações analíticas multi-tabela. Nosso trabalho valida esta abordagem agêntica com estado e estabelece uma baseline reproduzível para extensões futuras, incluindo benchmarking de modelos, fine-tuning de domínio específico e estratégias de decomposição de consultas.*

1. Introduction

Access to reliable health data is essential for supporting research, guiding public policies, and improving healthcare management. In Brazil, DATASUS[1] provides one of the most comprehensive open data repositories on hospitalizations, mortality, and epidemiological indicators. However, interacting with these databases requires technical knowledge or navigating complex data portals, which limits their accessibility to non-technical users.

In this context, recent initiatives aim to bring process analysis closer to non-technical users through more accessible interfaces. [Yeo et al. 2022] highlight the use of natural language processing for process mining queries in healthcare, eliminating the need for advanced programming knowledge. In Brazil, for example, the chatbot Ana,

[1]https://datasus.saude.gov.br/

proposed by [Chagas et al. 2023], had as its main functions symptom screening and providing reliable information to the population, combating misinformation regarding the COVID-19 pandemic.

Although conversational systems based on Artificial Intelligence (AI) have the potential to expand access to complex information, their application in healthcare still faces significant obstacles. [Kavaz et al. 2023] point out that most interfaces using natural language processing remain limited to simplified scenarios, achieving only specific results and lacking the flexibility for more sophisticated analyses. In healthcare, [Laymouna et al. 2024] show that, despite gains in accessibility and efficiency, chatbots still face barriers linked to usability, user trust, and ethical dilemmas related to privacy and equity.

Despite these limitations, there is still a lack of systems that enable direct, flexible access to structured health data in Brazil. To address this gap, we propose a conversational agent that allows natural language querying of hospitalization records obtained from DATASUS. We used a Large Language Model (LLM) to translate queries into SQL, achieving promising accuracy in simple queries, though complex ones remain challenging, highlighting the need for improved reasoning strategies. Our proposed approach is described in Section 2. Section 3 details the obtained results. The last section outlines our contributions, limitations, and future directions.

2. Proposed Solution

2.1. Data gathering and processing

We developed an ETL pipeline to process 23,792,498 SIH/SUS records from Rio Grande do Sul (2008–2023), extracted via PySUS.

Using Polars[2] for transformation, data were organized into a Star Schema with Admissions as the fact table and dimensions including Hospital, Municipalities, CID10, and Obstetrics. Official DATASUS and IBGE[3] dictionaries (procedures, diagnoses, municipality codes) were incorporated for semantic consistency.

Loading used PostgreSQL with batch insertion and primary/foreign key constraints to ensure referential integrity. Figure 1 shows a partial view of the schema: SIH/DATASUS tables in white and support tables in blue.

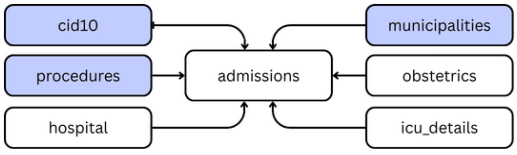

Figure 1. Partial database schema. Admissions fact table and its dimensions.

2.2. Solution Architecture

To address the accessibility challenge of querying DATASUS data, we developed a conversational agent that employs an agentic workflow with LangGraph providing stateful

[2]https://pola.rs/

[3]https://www.ibge.gov.br/

orchestration to translate Portuguese queries into SQL, execute them against PostgreSQL, and return natural language responses (Figure 2).

The pipeline comprises six stages: (1) **Table Discovery**, (2) **Table Selection** via LLM to reduce schema context, (3) **Schema Retrieval** of metadata for selected tables, (4) **SQL Generation** with domain-aware prompts, (5) **Self-Correction** using error feedback loops, and (6) **Execution & Response Generation** in natural language.

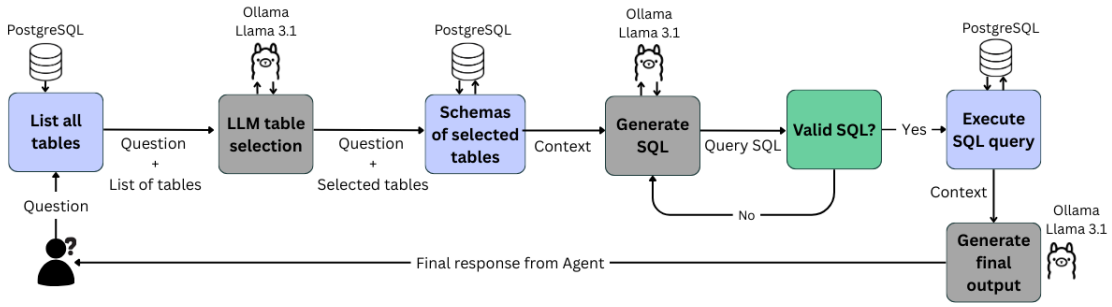

Figure 2. Agent workflow: LLM-driven tasks (gray), database operations (blue), and validation (green).

3. Experiments

The evaluation benchmark consists of 52 queries by complexity[4]: simple, single-table; moderate, 1–2 tables with ICD-10/temporal functions; complex, 2–4 table JOINs. Performance was measured using three metrics: **Success** (valid SQL executed without errors), **Component Matching (CM)**, and **Execution Accuracy (EX)** [Yu et al. 2018]. Table 1 shows overall 100% success, 82.9% EX, and 67.9% CM.

Table 1. Performance on 52-query evaluation set.

Tier	Queries	CM	EX	Success
Simple	22	89.5%	94.7%	100%
Moderate	17	61.9%	85.7%	100%
Complex	12	41.7%	58.3%	100%
Overall	**52**	**67.9%**	**82.9%**	**100%**

4. Final Remarks

This work presents a conversational agent using Llama 3.1:8B and LangGraph to query DATASUS hospitalization data, achieving 82.9% execution accuracy overall and 58.3% on complex queries, revealing limitations in multi-step reasoning for multi-table operations. Four extensions are planned: (i) *LLM benchmarking* comparing multiple models using in-context learning techniques [Zhang et al. 2023]; (ii) *Domain fine-tuning* on Portuguese NL→SQL pairs grounded in DATASUS semantics to internalize domain-specific priors [Sun et al. 2024, Lee et al. 2023]; (iii) *Query decomposition* using sub-query verification to reduce cascading errors [He et al. 2024, Wang et al. 2023]; and (iv) *Expert-validated ground truth* with public-health professionals following benchmark best practices [Lee et al. 2023].

[4]The code and benchmark of 52 queries are available at https://tinyurl.com/2kcnbrbz.

Acknowledgments

The authors gratefully acknowledge the financial support of the CNPq Scholarship PIBIC and Technological Development Productivity (303208/2023-6), and the FAPERGS 22/2551-0000390-7 (RITE CIARS Inteligência Artificial Aplicada à Saúde).

Referências

Chagas, B. A., Pagano, A., Prates, R., Praes, E., Ferreguetti, K., Vaz, H., Reis, Z., Ribeiro, L., Ribeiro, A. L., Pedroso, T., Beleigoli, A., Oliveira, C., and Marcolino, M. (2023). Evaluating user experience with a chatbot designed as a public health response to the covid-19 pandemic in brazil: Mixed methods study. *JMIR Hum Factors*, 10:e43135.

He, X. et al. (2024). Din-sql: Decomposed in-context learning of text-to-sql with self-correction. *arXiv preprint arXiv:2402.16347*.

Kavaz, E., Puig, A., and Rodríguez, I. (2023). Chatbot-based natural language interfaces for data visualisation: A scoping review. *Applied Sciences*, 13(12).

Laymouna, M., Ma, Y., Lessard, D., Schuster, T., Engler, K., and Lebouché, B. (2024). Roles, users, benefits, and limitations of chatbots in health care: Rapid review. *J Med Internet Res*, 26:e56930.

Lee, G., Hwang, H., Bae, S., Kwon, Y., Shin, W., Yang, S., Seo, M., Kim, J.-Y., and Choi, E. (2023). Ehrsql: A practical text-to-sql benchmark for electronic health records. *arXiv preprint arXiv:2301.07695*.

Sun, Y. et al. (2024). Sql-palm: Improved large language models for text-to-sql via few-shot learning and instruction fine-tuning. *Transactions on Machine Learning Research (TMLR)*. arXiv:2302.05965.

Wang, B., Li, B., Miao, Z., Zhang, Y., Ge, C., Xue, F., Li, H., Zan, H., Sun, B., and Yang, Y. (2023). DIN-SQL: Decomposed in-context learning for text-to-SQL. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 6468–6485, Toronto, Canada. Association for Computational Linguistics.

Yeo, H., Khorasani, E., Sheinin, V., Manotas, I., An Vo, N. P., Popescu, O., and Zerfos, P. (2022). Natural language interface for process mining queries in healthcare. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 4443–4452.

Yu, T., Zhang, R., Yang, K., Yasunaga, M., Wang, D., Li, Z., Ma, J., Li, I., Yao, Q., Roman, S., Zhang, Z., and Radev, D. (2018). Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-sql task. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3911–3921. Association for Computational Linguistics.

Zhang, H., Cao, R., Chen, L., Xu, H., and Yu, K. (2023). Act-sql: In-context learning for text-to-sql with automatically-generated chain-of-thought. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 3501–3532. Association for Computational Linguistics.