

A Tool for End-to-End Analytics: From Multi-driven Analytical Modeling to Goal-oriented Data Storytelling

Luís Pedro Baptista¹, Vânia Sousa^{2,1}, Pedro Guimarães^{2,1}, António Vieira^{1,2} and Maribel Yasmina Santos^{1,2}

¹ALGORITMI Research Center, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal

²CCG/ZGDV Institute, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal

Abstract

Industrial digital transformation generates a lot of data but turning it into useful insights for business strategy remains difficult. Traditional data-driven methods often overlook what decision-makers need, while requirements-driven methods can suggest solutions that are impractical due to insufficient data. To align these two perspectives, this paper presents a web-based prototype that uses a multi-driven analytical modeling approach supported by Large Language Models (LLMs). Our tool automatically proposes Data Warehouse (DW) schemas and dashboards by connecting physical data models to requirements structures from the iStar framework. Data engineers can review and improve the results generated by the LLM through a human-in-the-loop process. Used in real-world case studies from industry, this combined Human-AI approach accelerated DW design, improving the connection between strategic goals and analytics, and also helped identify key data gaps.

Keywords

Analytical Modeling, Large Language Models, Analytical Requirements, Data-driven Decision-Making

1. Introduction

Industrial decision-making increasingly depends on the ability to transform large volumes of operational data into actionable insights. While data-driven approaches have proven effective in supporting analytics and process optimization, they often focus on the available data rather than on the actual information needs of decision-makers. As a result, analytical solutions may be technically feasible but might fail to address questions that are most relevant to users.

To overcome this limitation, our previous work proposed a multi-driven and Large Language Models (LLMs) assisted analytical modeling methodology that combines data-driven and requirements-driven perspectives to derive conceptual and analytical models aligned with both available data and user requirements [1]. By leveraging LLMs, the methodology supports the generation of modeling artifacts while maintaining a human-in-the-loop process in which data engineers validate and refine the results. To support structuring analytical requirements aligned with the iStar framework [2] and enhance data storytelling in data-intensive contexts, the approach proposed in [3] links dashboards to analytical goals, improving both development efficiency and the decision-support process.

Although both approaches demonstrated promising results [1, 3], they still depend on multiple manual activities, including prompt preparation, model generation, artifact management, and iterative interactions between users, data engineers, and LLMs. These activities increase the effort required to apply them and can limit their scalability and adoption in industrial environments.

To address this gap, this paper presents a web-based prototype for a combined approach from [1] and [3]. The tool guides users throughout several steps, allowing verification and validation. By providing an integrated environment for analytical modeling, the proposed tool facilitates the systematic alignment

ER2026: Companion Proceedings of the 45th International Conference on Conceptual Modeling: Doctoral Consortium, Tutorials, Posters and Demos, October 5-8, 2026, St. John's, NL, Canada.

✉ pg61434@alunos.uminho.pt (L. P. Baptista); vania.sousa@ccg.pt (V. Sousa); pedro.guimaraes@ccg.pt (P. Guimarães); avieira@dsi.uminho.pt (A. Vieira); maribel@dsi.uminho.pt (M. Y. Santos)

🆔 0009-0000-2538-4723 (L. P. Baptista); 0009-0002-1279-6651 (V. Sousa); 0000-0003-3390-8528 (P. Guimarães); 0000-0002-1059-8902 (A. Vieira); 0000-0002-3249-6229 (M. Y. Santos)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

of data with user requirements while reducing the effort required to apply the methodology in practice. To assess its practical value and results, the tool was evaluated using the two industrial use cases [1, 3].

This paper is structured as follows. Section 2 briefly analyzes related works and positions the current work. Section 3 describes the proposed integrated methodology. Section 4 details the tool architecture and its prototype. Section 5 presents the conclusions.

2. Related Work

Recent research has demonstrated the potential of LLMs to support requirements engineering, domain and conceptual modeling, and Data Warehouse (DW) design tasks. In requirements engineering, LLM-based approaches have been used to assist requirements specification and knowledge extraction from unstructured sources, reducing manual effort and improving adaptability across domains [4, 5]. However, challenges related to domain expertise, hallucinations, and the interpretation of non-textual information remain significant obstacles to their reliable adoption in industrial settings [4]. In domain modeling, iterative LLM-driven frameworks have shown promising results in identifying entities, relationships, and higher-level modeling patterns, although limitations remain in handling abstraction and complex pattern detection [6]. Similarly, studies on conceptual modeling and DW design indicate that LLMs can effectively support model generation and refinement; however, the resulting artifacts may contain semantic inconsistencies, omit relevant elements, or fail to fully comply with modeling best practices, thereby requiring expert validation [7, 8, 9]. Recent studies have also explored the integration of LLMs into analytical modeling methodologies and decision-support design [1]. In a complementary direction, Sousa et al. [3] demonstrated how LLMs can support the transformation of analytical requirements into data models and visualizations within a goal-oriented methodology.

Overall, the literature highlights the potential of LLMs to support analytical modeling activities while emphasizing the need for human supervision, structured methodologies, and mechanisms that facilitate their practical adoption in industrial environments.

3. Integrated Approach for Multi-driven Analytical Modeling

The integrated approach resulting from [1] and [3] includes a hybrid methodology to align data availability with business requirements through a seven-step pipeline (Fig. 1). This pipeline operates in progressive stages, starting with the identification of the conceptual data model (in the form of an Entity-Relationship Diagram, ERD), considering the available data (phase Data-driven Conceptual Model, DDCM). In parallel, the Requirements Structuring and Refinement (RSR) phase can support the specification of the analytical goals in the form of an iStar model, in case the analytical requirements are specified in natural language and no iStar model is available [2, 10].

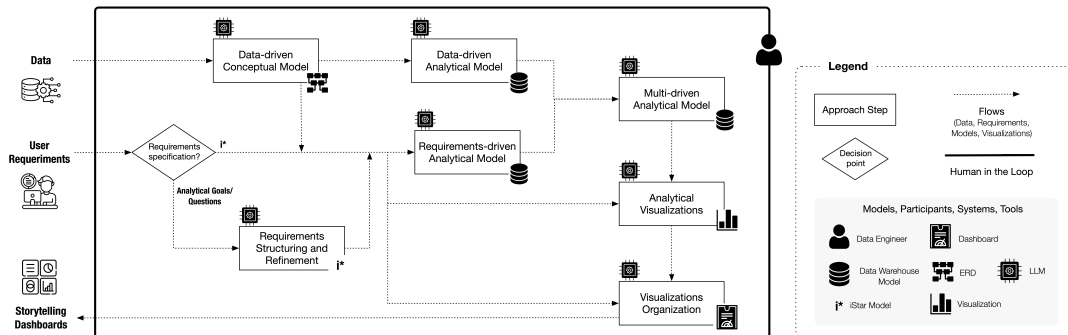


Figure 1: Integrated Approach for Multi-driven Analytical Modeling.

Once both ERD and iStar models are available, the LLM supports the identification of two different DW schemas. The Data-driven Analytical Model (DDAM) considers the available data, while the

Requirements-driven Analytical Model (RDAM) is based on requirements/decision-makers' goals.

Having these two models, the Multi-driven Analytical Model (MDAM) phase combines both perspectives, data and requirements, verifying overlapping concepts, identifying data gaps, and tracing back all tables/attributes to their corresponding analytical goals. With the MDAM, the tool supports the identification of the most suitable Analytical Visualizations (AV) and their organization (Visualizations Organization, VO) into dashboards that consider the strategic, decision and information goals [10] specified in the analytical requirements.

As relying only on automated decisions by the LLM is not a good practice due to hallucinations or misclassifications, the approach relies on a human-in-the-loop validation (Fig. 1). This validation is done by data engineers who review the outputs and fix errors before proceeding to the next stage.

4. Tool Architecture and Prototype

The proposed architecture (Fig. 2) is supported by a web-based prototype developed using a microservices architecture, with the system fully containerized using Docker Compose¹, ensuring scalability and facilitating deployment in industrial environments. The client-side is a single-page application that uses React² and Tailwind CSS³ for an interactive working space where complex models are rendered using React Flow⁴, such as network diagrams with ERD or iStar. Dashboards are generated by the backend using Recharts⁵. The global state of the system used to guide users throughout each step and enforce mandatory human checks is managed by Zustand⁶. The backend layer uses Python⁷ with FastAPI⁸ as the central orchestrator. It handles user requests asynchronously, manages prompt engineering, and uses the Mistral AI⁹ language model to generate JSON schemas. Resend API¹⁰ is used to send automated email notifications after processing. Data and models are stored in a MongoDB¹¹ NoSQL database.

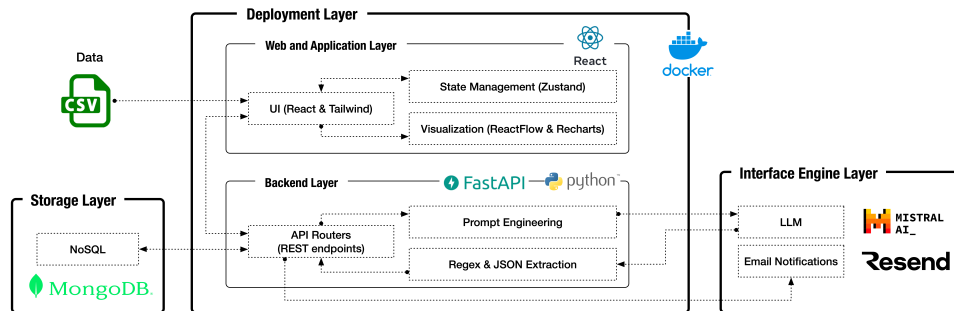


Figure 2: Tool Architecture Overview.

To protect the system against unpredictable formatting from language models, several safeguards were added to the backend. Strict prompt instructions and low-temperature settings were applied to ensure that the LLM only generates predefined JSON schemas. Responses from the LLM using Markdown¹² are cleaned up using regular expressions that extract valid JSON data. Rendering modules are wrapped in Global Error Boundaries¹³ to keep the frontend stable when the LLM mislabels concepts.

¹<https://docs.docker.com/compose/>, last accessed: 28.07.2026.

²<https://react.dev>, last accessed: 28.07.2026.

³<https://v3.tailwindcss.com/docs/guides/create-react-app>, last accessed: 28.07.2026.

⁴<https://reactflow.dev>, last accessed: 28.07.2026.

⁵<https://recharts.github.io>, last accessed: 28.07.2026.

⁶<https://zustand.docs.pmnd.rs/learn/getting-started/introduction>, last accessed: 28.07.2026.

⁷<https://www.python.org>, last accessed: 28.07.2026.

⁸<https://fastapi.tiangolo.com>, last accessed: 28.07.2026.

⁹<https://mistral.ai>, last accessed: 28.07.2026.

¹⁰<https://resend.com>, last accessed: 28.07.2026.

¹¹<https://www.mongodb.com>, last accessed: 28.07.2026.

¹²<https://www.markdownguide.org>, last accessed: 28.07.2026.

¹³<https://legacy.reactjs.org/docs/error-boundaries.html>, last accessed: 28.07.2026.

To validate this web-based prototype, two previous use cases from industrial settings were used [1, 3]. As shown in Fig. 3, the user can follow the predefined pipeline (stages of the approach presented in Section 3) and interact with the LLM to improve or validate the results. On the left of Fig. 3, the obtained MDAM is shown with colors highlighting the data-driven branch (blue), the requirements-driven branch (orange) and their overlap (green). To the right of Fig. 3, a dashboard organizing the proposed visualizations by a strategic goal and its decision and information goals is shown.

For accessibility and reproducibility, the code, data from the use cases, and videos demonstrating the tool are available at <https://github.com/luis-art-16/Multi-driven-Analytical-Tool>.

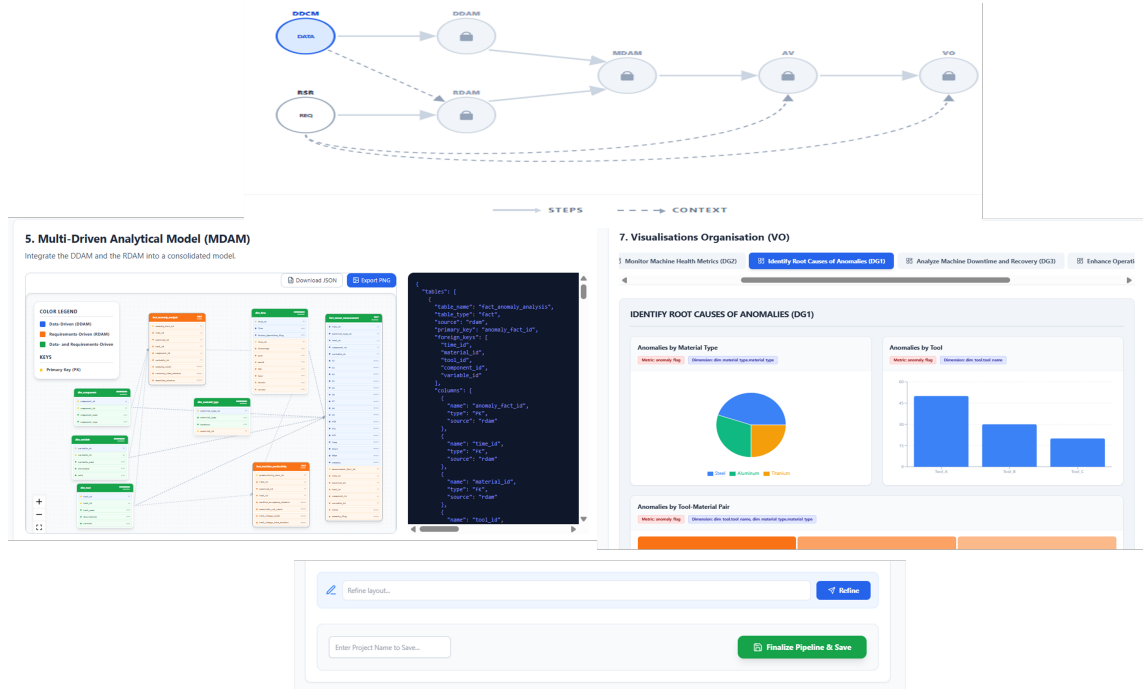


Figure 3: Tool frontend showcasing the pipeline stages (top), the generated MDAM and interactive Dashboard (middle), and the refinement interface (bottom).

5. Conclusions

This paper proposed an approach supported by a web-based prototype for the identification of multi-driven analytical models and the corresponding analytical visualizations. We have used LLMs to support both data-driven and requirements-driven views of conceptual models, data warehouse schemas, and goal-oriented dashboards. When tested in two industrial use cases, the human-in-the-loop capability was key to mitigating the effects of LLM hallucinations and ensuring that all outputs were constrained by domain rules and business knowledge.

As future work, the next phases should apply this approach to more industrial cases. To address data privacy and support sustainable Green IT, future research also considers using quantized Small Language Models (SLMs) instead of cloud-based LLMs. Running these optimized AI models locally would reduce reliance on APIs, cloud costs, and the carbon footprint of the analytics process.

Acknowledgments

This work has been supported by FCT – *Fundação para a Ciência e Tecnologia* within the R&D Unit Project Scope UID/00319/2025 - Centro ALGORITMI (ALGORITMI/UM)¹⁴ and the European Union

¹⁴<https://doi.org/10.54499/UID/00319/2025>, last accessed: 08.09.2026.

under the NextGenerationEU, through a grant of the Portuguese Republic's Recovery and Resilience Plan (PRR) Partnership Agreement, within the scope of the project BE.NEUTRAL – “*Agenda da Mobilidade para a neutralidade carbónica das cidades*”, aiming at the development, industrialization and operation of new carbon-zero mobility cyber-physical products that will accelerate the transition to carbon-neutral cities by 2030 (Project ref. nr. 35 - C644874240-00000016; Total project investment: 221.376.867,63 Euros; Total Grant: 128.619.400,50 Euros). This paper uses icons made available by www.flaticon.com.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly for sentence polishing and rephrasing and MistralAI for supporting the proposed approach. All generated content was reviewed and edited by the authors, who take full responsibility for the final text.

References

- [1] C. Vieira, V. Sousa, P. Guimarães, D. Rodrigues, A. C. Vieira, M. Y. Santos, Multi-driven and LLM-assisted Analytical Modeling: Evaluation in an Industrial Case, in: International Conference on Conceptual Modeling, 2025. URL: https://ceur-ws.org/Vol-4099/ER25_industrial_Vieira.pdf.
- [2] F. Dalpiaz, X. Franch, J. Horkoff, iStar 2.0 Language Guide, CoRR abs/1605.07767 (2016). URL: <https://doi.org/10.48550/arXiv.1605.07767>. arXiv: 1605.07767.
- [3] V. Sousa, C. Vieira, P. Guimarães, F. R. Pereira, M. de Sa, A. C. Vieira, M. Y. Santos, Goal-Oriented Data Storytelling from User Requirements: An LLM-Assisted Method for Industrial Analytics, in: International Conference on Conceptual Modeling, 2025. URL: https://ceur-ws.org/Vol-4099/forum_paper6.pdf.
- [4] A. Nouri, B. Cabrero-Daniel, F. Törner, H. Sivencrona, C. Berger, Engineering safety requirements for autonomous driving with large language models, in: 2024 IEEE 32nd International Requirements Engineering Conference (RE), IEEE, 2024, pp. 218–228. URL: <https://ieeexplore.ieee.org/document/10628478>.
- [5] W. Zhao, Q. Chen, J. You, LlmRe: A zero-shot entity relation extraction method based on the large language model, in: Proceedings of the 2023 7th International Conference on Electronic Information Technology and Computer Engineering, 2023. URL: <https://doi.org/10.1145/3650400.3650478>.
- [6] Y. Yang, B. Chen, G. Mussbacher, D. Varró, Multi-step Iterative Automated Domain Modeling with Large Language Models, ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems (MODELS Companion '24) 1 (2024) 587–595. URL: <https://doi.org/10.1145/3652620.3687807>.
- [7] J. Cámara, J. Troya, L. Burgueño, A. Vallecillo, On the assessment of generative AI in modeling tasks: An experience report with ChatGPT and UML, Software and Systems Modeling 22 (2023) 781–793. URL: <https://doi.org/10.1007/s10270-023-01105-5>.
- [8] K. Chen, Y. Yang, B. Chen, J. A. Hernández López, G. Mussbacher, D. Varró, Automated Domain Modeling with Large Language Models: A Comparative Study, in: 2023 ACM/IEEE 26th International Conference on Model Driven Engineering Languages and Systems (MODELS), 2023, pp. 162–172. URL: <https://doi.org/10.1109/MODELS58315.2023.00037>.
- [9] S. Rizzi, Using ChatGPT to refine draft conceptual schemata in supply-driven design of multidimensional cubes, in: International Workshop on Data Warehousing and OLAP, 2025. URL: <https://ceur-ws.org/Vol-3931/short4.pdf>.
- [10] A. Maté, J. Trujillo, X. Franch, Adding semantic modules to improve goal-oriented analysis of data warehouses using I-star, Journal of Systems and Software 88 (2014) 102–111. URL: <https://doi.org/10.1016/j.jss.2013.10.011>.