Multi-driven and LLM-assisted Analytical Modeling: Evaluation in an Industrial Case

Cristiana Vieira[1,2,*], Vânia Sousa[1,2], Pedro Guimarães[1,2], Diogo Rodrigues[3], António Vieira[2,1] and Maribel Y. Santos[2,1]

[1]CCG/ZGDV Institute, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal
[2]ALGORITMI Research Center, University of Minho, Campus de Azurém, 4800-058 Guimarães, Portugal
[3]Colep Packaging, 3730-423 Vale de Cambra, Portugal

Abstract

Decision-making in industrial environments increasingly depends on the ability to align operational data with strategic goals. Data-driven approaches often focus on the available data while overlooking the specific informational needs of decision makers. To bridge this gap, this paper proposes a multi-driven approach that integrates both data-driven and requirements-driven perspectives, supported by the use of Large Language Models (LLMs). Through prompt-based interactions, LLMs generate conceptual and analytical models from metadata and user requirements, accelerating the modeling process while ensuring relevance and coherence through the integration of both data and analytical requirements. The approach is applied to a real-world industrial case characterized by operational complexity, high variability, and technological heterogeneity. The results show how the combined use of LLMs and a structured modeling approach can support the development of analytical systems that are technically feasible and strategically aligned.

Keywords

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

1. Introduction

In industrial environments, decision-making processes increasingly rely on the ability to extract actionable insights from operational data. Modern manufacturing settings employ advanced analytics to monitor operations in real time, detect anomalies, and optimize workflows - with Industry 5.0 studies reporting improvements in decision outcomes of up to 46% following the implementation of data-driven methods [1]. This data-driven approach focuses on analyzing event logs and transactional records to build analytical models that can inform business decisions. While effective in many scenarios, this perspective often neglects an essential component: the actual informational needs of decision-makers (i.e., a requirements-driven perspective). This paper proposes an approach that combines both, hereafter referred to as 'multi-driven'. Consequently, analytical systems may reconstruct available data but fail to address the key questions that truly matter to users.

To address this gap, this paper proposes an approach that combines a multi-driven perspective, integrating the strengths of both data-driven and requirements-driven approaches. The goal is to align what data reveals with what users need, ensuring that the resulting analytical models are both feasible and meaningful. This dual orientation enables organizations to capture not only the structure of the available datasets but also the strategic, decision, and information goals of stakeholders. Although such alignment can be performed manually, it is often labor-intensive and error-prone, particularly in complex industrial contexts characterized by diverse, fragmented data sources. To overcome this challenge, we propose leveraging Large Language Models (LLMs) to assist in constructing analytical

ER2025: Companion Proceedings of the 44th International Conference on Conceptual Modeling: Industrial Track, ER Forum, 8th SCME, Doctoral Consortium, Tutorials, Project Exhibitions, Posters and Demos, October 20-23, 2025, Poitiers, France
cristiana.vieira@ccg.pt (C. Vieira); vania.sousa@ccg.pt (V. Sousa); pedro.guimaraes@ccg.pt (P. Guimarães); diogo.rodrigues@colep-pk.com (D. Rodrigues); avieira@dsi.uminho.pt (A. Vieira); maribel@dsi.uminho.pt (M. Y. Santos)
0009-0000-5128-3571 (C. Vieira); 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)

models. Hence, a major contribution of this work is the proposal and evaluation of a multi-driven and LLM-assisted analytical modeling approach. This contribution have been grounded in a research process guided by the Design Science Research Methodology for Information Systems [2].

Recent work has demonstrated that LLM agents can automate multiple stages of industrial analytics, including process understanding, concept extraction, and iterative self-refinement, effectively accelerating the end-to-end data-to-dashboard pipeline [3]. In our approach, LLMs use metadata and user-specified requirements to produce initial versions of conceptual and analytical models, which are then validated and refined by data engineers. This hybrid process accelerates model development while preserving alignment with decision needs.

This paper applies a structured approach that combines process modeling, conceptual and analytical data modeling, and requirements analysis, supported by prompt-based interactions with an LLM. It is applied to a real-world industrial case known for high variability, technological diversity, and constant adaptive challenges. The results show how this multi-driven approach supports targeted and effective decision-making by aligning business objectives with data.

This paper is structured as follows. Section 2 analyzes related works. Section 3 describes the use case where the proposed approach will be applied. Section 4 details the steps of the proposed approach. Section 5 shows an instantiation of the use case. Section 6 presents the conclusions and future work.

2. Related Work

The increasing complexity of modern industrial environments poses significant challenges for effective decision-making. Frequent changes in operational settings, safety constraints, and regulatory requirements demand iterative cycles of requirements engineering, process modeling, and data alignment. In such settings, aligning the semantics of available data with user information needs is essential but difficult to automate - particularly in the conceptual design of Data Warehouses (DWs) or domain models [4, 5].

Recent studies have explored the use of LLMs to assist engineers in automating or accelerating parts of this process. Nouri et al. [6] developed a prompt-based LLM prototype to support the specification of safety requirements in autonomous driving systems. Their approach significantly reduced the time required for hazard analysis and risk assessment (HARA) but highlighted limitations in domain expertise, hallucination risks, and difficulties in interpreting non-textual input. Zhao et al. [7] proposed LlmRe, a method for zero-shot relation extraction from unstructured text, using in-context learning and a three-stage decomposition strategy. Their framework demonstrated good adaptability across domains, which helped reduce the dependence on labeled data, a valuable trait for industrial scenarios with heterogeneous information sources.

Domain modeling has been another area of interest. Yang et al. [5] introduced an iterative, multi-step LLM-based framework to extract modeling elements and identify higher-level patterns. Their results showed improvements in F1-scores for class and relation identification, though issues with abstraction and complex pattern detection remain. Similarly, Chen et al. [8] compared several prompt engineering strategies and found that, while LLMs achieved high precision in model generation, they often missed key elements and struggled with modeling best practices.

Other works have focused on supporting conceptual modeling tasks. Cámara et al. [9] assessed the ability of ChatGPT to assist in generating UML class diagrams and OCL constraints. They reported that models generated were often syntactically valid but semantically inconsistent and required significant human refinement. Rizzi [4] explored LLM-based multidimensional model refinement in DW design, achieving improved results with carefully designed instructions but still requiring manual validation, especially for tasks involving shared hierarchies and functional dependencies.

Overall, while LLMs have shown potential in helping with requirements engineering and model creation, studies consistently indicate that human involvement is necessary to ensure that the meanings are correct and fit with specific industry standards. This supports the development of hybrid, human-in-the-loop approaches — as explored in this paper — where LLMs accelerate the initial modeling effort, but outputs are validated and refined collaboratively.

Unlike previous works, which often focus on isolated modeling tasks, our approach integrates both data-driven and requirements-driven perspectives in a unified pipeline that explicitly targets the alignment of user requirements with available industrial data, supported by LLMs.

3. Industrial Use Case: Colep Packaging

Colep Packaging[1], a company of the RAR Group[2], is the Iberian leader in general line packaging and an important European supplier of aerosols. With production in Portugal, Spain, Poland, and Mexico, and more than 700 employees, it operates with a vertical integration model, developing metal and plastic solutions for several markets [10]. With over five decades of experience, Colep Packaging operates in an industrial environment marked by high operational complexity, both at the technological and managerial levels. The main raw material used in the production of metal packaging is tinplate, a low-carbon steel sheet coated with tin, selected according to strict technical specifications (thickness, strength, surface treatment). This material arrives at the factory in coils and undergoes a structured production process involving multiple stages, from printing (lithography) and cutting to printing, assembly, quality control, and final dispatch.

The core production flow is represented in Fig. 1, using a BPMN (Business Process Model and Notation) diagram. This visual model details each stage of the process and explicitly highlights key decision points and sub-processes. For example, depending on the type of product or volume, different printing technologies (offset or digital) may be selected, each with its own implications in terms of setup time, cost, and flexibility. The diagram helps clarify the operational logic while also making visible the possible paths and conditional branches that increase the complexity of execution and planning. It also makes it easier to identify which parts of the process are more sensitive to variability and where analytical support is more critical, since BPMN explicitly models alternative paths and decision points (e.g., the choice between offset and digital printing), highlighting where execution may diverge.

The company combines offset and digital printing technologies, each with distinct operational trade-offs. Offset printing is more suitable for large production volumes and ensures high color consistency, while digital printing, although slower and more limited in print quality, allows for quick adjustments, minimal setup time, and greater flexibility in small or customized batches. Managing this hybrid setup requires the coordination of materials, design files, and production schedules, increasing the need for adaptable and responsive planning systems.

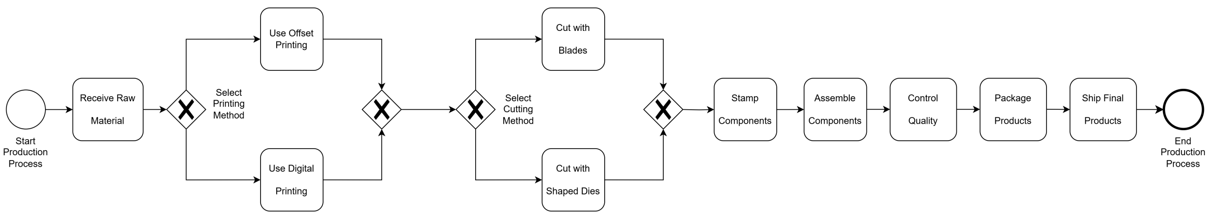

Figure 1: BPMN Representation of the Production Process.

Beyond the technological setup, the complexity of the use case becomes more evident when considering the dynamic nature of the industrial scenario. The company faces constant changes that affect production and decision-making processes, such as the integration of new equipment, frequent reconfiguration of the shop floor layout, supply chain disruptions, coordination between multiple work centers, adaptation to new labor models (e.g., shift systems or four-day weeks), and compliance with sustainability and energy efficiency targets. These factors create a high-variability environment that challenges traditional planning approaches and requires flexible, data-driven systems capable of integrating information from multiple sources.

[1]https://colep-pk.com/pt-pt
[2]https://www.rar.com/pt/

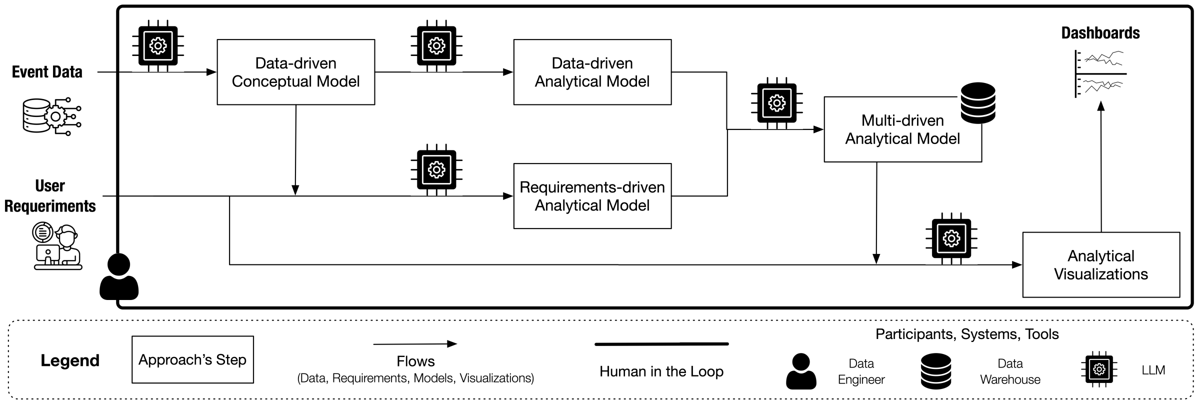

Figure 2: Approach for Multi-driven Analytical Modeling Supported by LLMs.

To support decision-making under these conditions, Colep Packaging relies on digital tools such as Manufacturing Execution Systems (MES) and industrial simulation models, which allow for real-time monitoring, predictive analysis, and optimization of production flows. However, even with such systems in place, the integration between business objectives and operational processes remains a challenge. The variability and heterogeneity of data sources, combined with evolving user needs, highlight the need for approaches that can simultaneously consider what the data reveal (data-driven) and what the decision-makers require (requirements-driven). This context of high complexity justifies the relevance of applying a multi-driven decision-making approach, as proposed in this paper.

4. Multi-driven and LLM-assisted Analytical Modeling

To address the challenges of aligning business goals with complex data environments, this paper applies a structured, LLM-supported approach that integrates process and data modeling and requirements analysis. This section presents the approach in two stages: the first describes the overall approach that combines data and user requirements to build a multi-driven analytical model; the second details the structured prompts used to guide each interaction with the LLM.

4.1. Proposed Approach

Although different implementations can be followed, this paper adopts the analytical data model as a DW system [11] with fact and dimension tables.

As mentioned, two analytical data models can be identified, one derived from the data and the other from the requirements, both considering the identified conceptual data model of the domain. In the several tasks, an LLM supports this approach with increased efficiency (the models can be obtained in a short period of time) and handling the complexity that is usual in complex industrial settings, with a high diversity of data sources and data sets. These models are later integrated by the LLM. Along all the steps, the data engineer validates the proposals and makes the necessary suggestions/corrections towards the model to be implemented. Following this two-fold path ensures that the data available for analysis guides the identification of the analytical data model, but this model is then refined considering the user's needs. In case of a lack of data for specific user requirements, this approach ensures that the data engineer and the organization are aware of it, taking proper actions to collect such data in the future.

With the multi-driven analytical data model, the LLM supports the identification of useful visualizations. To be possible, this work considers that the user requirements are specified in iStar [12], as this allows the specification of the strategic, decision, and information goals [13]. Strategic Goals (SGs) are related to the main objectives of the business process that are being enhanced, representing a desired change from a current situation to a future one. Decision Goals (DGs) represent decisions that use information to provide benefits for the organization, operationalizing the SGs into actions by answering the question, *"How can a strategic goal be achieved?"*. Decisions can be explained in terms of objectives or tasks. Information Goals (IGs) are linked to the question *"How can decision goals be achieved in terms of information required?"*. IGs outline the data that must be gathered, usually through analysis. As a result, they can be described in terms of goals or in terms of the analysis process. With the IGs, and as suggested in the work of Lavalle et al. [14], the analytical dashboards can be organized by the SG and the supporting DGs and IGs. This proposal of analytical visualization can be refined by the data engineer, validating the dashboards that are made available for the users.

4.2. Prompt Strategy

Besides the data and user requirements as input, another key component of the approach is the prompt strategy to interact with the LLM. To obtain useful results and avoid hallucination [15], the prompts must detail the input of each step and the expected output. For the input, this work considers:

- **Data**: include the activities of the business processes and the different events associated with those activities. Activities usually relate different entities of the domain, such as *Customer* and *Order* for *place order*, with a set of time-series events. To avoid mixing entities, activities, and events, the prompt must clarify the several concepts present in the input data.
- **User requirements**: include the SGs to be addressed and the way they are supported by the DGs and IGs. These concepts should be clarified in the prompt, which must also highlight that the IGs commonly point to business or process indicators associated with the main entities of the domain or with the supporting activities, respectively.
- **Visualization tasks**: useful visualizations must align the analytical task, the type of data, and the cardinality of the data, among other characteristics. Pointing to or adopting a specific taxonomy or classification for visualization tasks enhances the resulting visualizations.

For data, only the metadata needs to be made available (ensuring confidentiality, since no sensitive values are shared). This is important given privacy and confidentiality concerns. This metadata includes the available tables and the corresponding attributes. The proposed prompts are next detailed.

4.2.1. From the Event Data to the Data-driven Conceptual Model

Here, the metadata of the available data is used as input to obtain the data-driven conceptual model. The prompt is as follows:

> Consider a dataset with the following tables and the associated attributes. With this metadata, infer the several relationships between the entities of the domain, and propose an entity-relationship diagram that represents the conceptual model of the domain. This conceptual model must include the entities and their attributes, as well as the relationships among these entities. The conceptual model can only include entities, attributes, and relationships inferred from the available metadata. Also, consider that the entities of the domain have events that are occurrences of the activities that affect entities over time. These events or activities cannot be considered entities of the domain. The tables and their attributes are: *«to be defined»*.

4.2.2. From the Data-driven Conceptual Model to the Data-driven Analytical Model

The conceptual data model obtained in the previous step is used as input to infer an analytical data model. The prompt is as follows:

> Consider the conceptual data model obtained in the previous prompt. Assume we want to create a data warehouse system for the analysis of that data. Adopt a constellation schema if multiple fact tables are needed, integrating the several fact tables through shared dimensions. For each star schema of the constellation, describe the fact table, the dimension tables, and the attributes of these tables. For each fact table, list the measurable indicators (metrics).

4.2.3. From the User Requirements and the Data-driven Conceptual Model to the Requirements-driven Analytical Model

Here the conceptual model guided by the data and the user requirements is used to derive an analytical model with fact and dimension tables. The prompt is as follows:

> Consider user requirements specified using the iStar extension for Data Warehouses. In it, Strategic Goals (SG), Decision Goals (DG), and Information Goals (IGs) are used. SGs are related to the main objectives of the business process that are being enhanced, representing a desired change from a current situation to a future one. DGs represent decisions that use information to provide benefits for the organization, operationalizing the SGs into actions by answering the question, "How can a strategic goal be achieved?" IGs guide the information needed to achieve a DG by responding to the question, "How can decision goals be achieved in terms of information required?" IGs outline the data that must be gathered, usually through analysis. As a result, they can be described in terms of goals or in terms of the analysis process. The requirements are: *«to be defined»* Based on the conceptual data model obtained from the data and the user requirements mentioned above, propose an analytical data model suitable for a data warehouse system that addresses the entities of the domain. Identify one or more fact tables that meet the IGs, and define the dimension tables needed to provide the relevant contextual information. Use a constellation schema if needed. For each fact table, list the measurable indicators (metrics), and for each dimension, list the attributes required to support filtering or grouping. Clearly indicate how each IG is supported by the analytical data model.

4.2.4. From the Data- and Requirements-driven Analytical Models to the Multi-driven Analytical Model

Now the LLM must merge the two previous analytical models, maintaining the alignment between available data and requirements. The prompt is as follows:

> Given the two analytical data models — one derived from the available data and the other from user requirements in the context of the entities of the domain - integrate both into a unified, multi-driven analytical model. Use a constellation schema if multiple fact tables are needed. When dimensions or facts overlap, unify them; when they diverge, include both with appropriate relationships. Highlight where the models complement each other or where gaps in data exist.

4.2.5. From the Multi-driven Analytical Model to the Analytical Visualizations

Based on the final model, the LLM should propose useful visualizations for users based on the IGs. The prompt is as follows:

> Based on the multi-driven analytical model and the IGs, suggest analytical dashboards or visualizations that can support decision-making. For each visualization, specify: i) the SG and DG it supports; ii) the data it uses (fact and dimension tables); iii) the type of chart (e.g., time series, bar chart, heatmap); and iv) the recommended aggregation and filtering dimensions. Whenever possible, organize the visualizations by SG, showing how each IG contributes to it.

5. Instantiation to the Colep Industrial Use Case

This section presents the step-by-step application of the multi-driven decision-making approach. The LLM used in the instantiation was the ChatGPT language model[3]. All prompts were detailed in the previous section. For reproducibility, the used prompts and obtained results are available at [16].

5.1. Data-driven Conceptual Model

The first step identified a conceptual data model based on the structure of the available industrial dataset (Fig. 3). Only metadata with the tables names and their attributes was made available. The resulting model identifies key entities within the production domain, such as Customer, Order, Material, Operation, and WorkCenter, and outlines the relationships between them. For instance, it shows that customers place orders, which in turn triggers the production process. Each order uses a specific set of materials required for the production of the ordered items. Orders include a sequence of operations that are performed at a particular work center (physical or logical area of the industrial setting, such as

[3]ChatGPT-4, released by OpenAI in May 2025, available at https://chat.openai.com

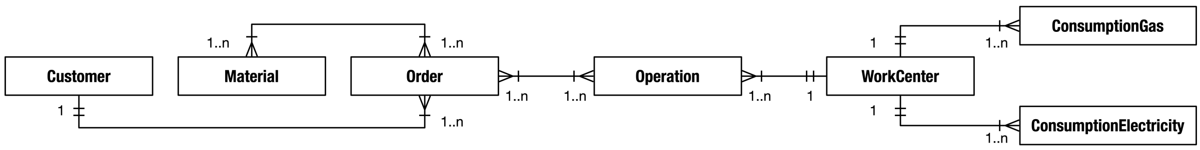

Figure 3: Data-driven Conceptual Data Model.

lithography or cutting, where a specific type of operation takes place). The model also incorporates energy consumption data, linking each work center to its associated gas and electricity usage through the entities ConsumptionGas and ConsumptionElectricity.

It is important to note that this conceptual model does not represent process sequences or execution flow; rather, it describes the static relationships among the core business entities relevant to the domain. This conceptual structure provides a consistent and realistic representation of the production environment and serves as a solid foundation for analytical modeling in the following steps. This model (Fig. 3) was already validated by the data engineer, who disregarded a redundant relationship between orders and work centers identified by the LLM, as operations already relate these two entities.

5.2. Data-driven Analytical Model

After the conceptual data model, the LLM suggested three main fact tables: one focused on order analysis (Fact_Orders), another on operation execution (Fact_Production), and a third on energy consumption (Fact_EnergyConsumption). These fact tables are connected to dimension tables, such as Date (Dim_Date), Time (Dim_Time), Customer (Dim_Customer), Work Center (Dim_WorkCenter), and Operation (Dim_Operation). This model is discussed in subsection 5.4.

5.3. Requirements-driven Analytical Model

Considering the concepts of the domain and the user requirements specified in iStar (Figure 4), the LLM suggested a structured mapping between the IGs and the existing analytical data model. It should be noted that the iStar model was designed by the authors, based on requirements elicited in collaboration with the industrial partner. Each IG was linked to the relevant fact and dimension tables, confirming that the previously defined constellation schema could support all analytical needs derived from the SGs and DGs.

The result confirmed the need for a constellation schema, but centered on two fact tables: Fact_Production, capturing detailed records of operational execution (e.g., duration, idle time, and energy per operation); and Fact_EnergyConsumption, summarizing energy usage per time period and work center. These are linked to dimensions such as WorkCenter, Operation, Material, Date, and Time, depending on the case, enabling flexible filtering and aggregation.

It is important to highlight that: i) a traceability matrix was proposed by the LLM, identifying the metrics and dimensions required to support each IG. This ensured that the analytical model covers the full set of user requirements defined in iStar and clarified how each element of the schema contributes to addressing specific decision-making needs; ii) no fact table for orders was suggested by the LLM, which makes sense as the user requirements are focused on operational concerns, such as reducing shelf time and energy consumption, and not on more business-oriented objectives, such as increasing the number of orders over time.

5.4. Multi-driven Analytical Model

Considering both analytical data models, Figure 5 depicts the integrated data model, highlighting the differences between them when applicable.

As Figure 5 shows, most of the fact and dimension tables were suggested by both data- and requirements-driven perspectives (tables with headers represented in a white color). The requirements-

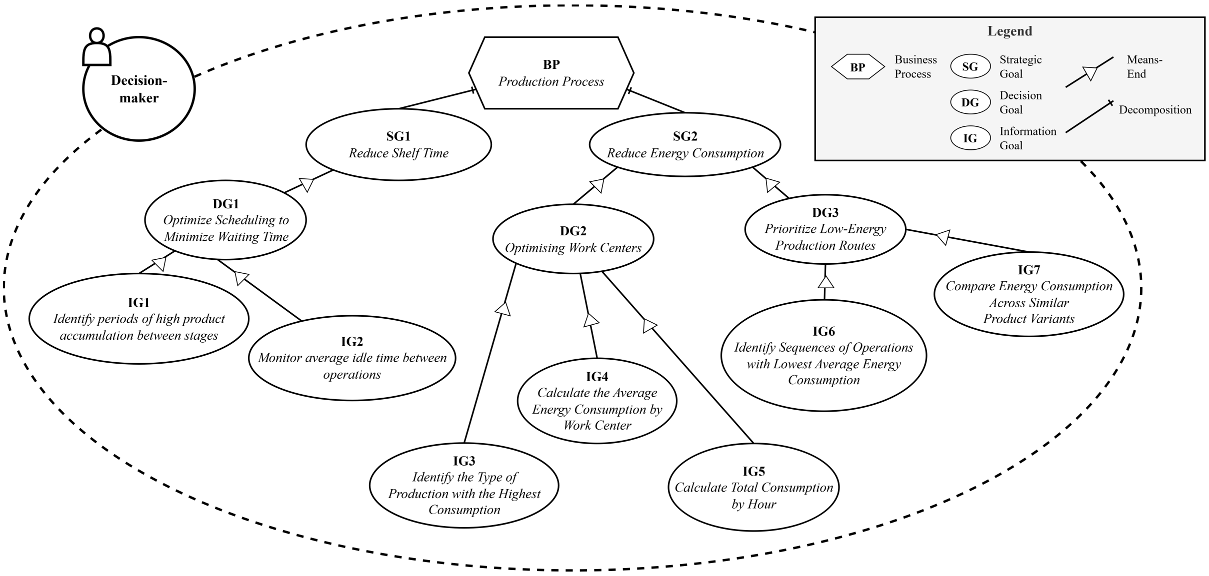

Figure 4: User Requirements Specified in iStar.

driven perspective suggested mainly additional metrics to include in the fact tables, addressing the different IGs and computing the indicators for answering the user's requirements (these indicators are represented in a light-grey color). The data-driven perspective, with the tables suggested only by this perspective represented with the header in a dark-grey color, is related to information present in data that is not needed to answer this specific set of user requirements. However, these can be useful for further analysis, as mentioned previously in subsection 5.3, as the use case only included a limited set of user requirements with operational concerns. With organizational decision-making in mind, the integrated analytical model provides a comprehensive analysis of different business and process indicators, making available the proposal of a DW that is not devised for current informational needs but prepared to answer to future analyses.

5.5. Analytical Visualizations

The final step of the approach consists in translating the multi-driven analytical model into visualizations that directly support decision-making. From the dashboards suggested by the LLM (with a semi-automatic alignment process, where the LLM proposes and a data engineer validates), we selected the one associated with the DG "Optimize Work Centers" (Figure 6) of the SG "Reduce Energy Consumption," allowing the analysis of IG3, IG4, and IG5.

First, a DW system was physically instantiated in the industrial setting, following the multi-driven analytical model. Afterwards, a dashboard that includes the visualization of energy consumption by work center and time period was implemented, addressing IG4 and IG5. For IG3, a key limitation was found. Although the production type exists in the dataset, it contains only a single value - offset. This lack of variability prevents any meaningful comparison between different production types, which is precisely what the goal requires. As a result, while IG3 is formally represented in the dashboard, the findings fall short. As illustrated in Figure 6, the bar showing the "offset" production type visually highlights this limitation. Meaningful analyses would require collecting energy consumption data across multiple distinct production types. Pointing these findings is important for future data collection tasks, suggesting the definition of a data policy regarding current and future information needs.

Despite this limitation, the dashboard reveals valuable insights regarding IG4 and IG5. For example, as shown in the average consumption chart, work centers CS08 and CS13 stand out with the highest average electricity consumption. While in some cases high consumption aligns with longer operation durations, the dashboard also reveals exceptions - such as certain work centers with long duration but comparatively lower energy consumption - highlighting the importance of analyzing both dimensions

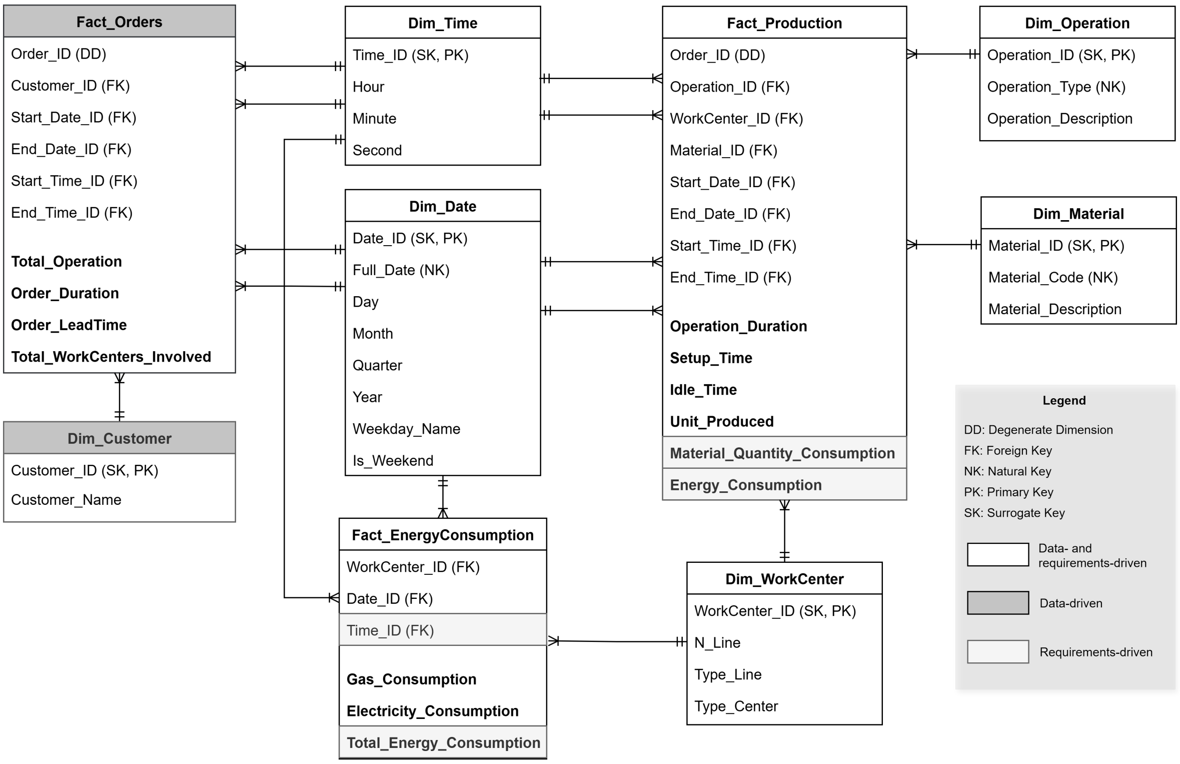

Figure 5: Multi-Driven Analytical Model.

jointly. Additionally, the heatmap related to IG5 reveals clear temporal patterns, with consumption peaking at 9:00am on Wednesdays and 8:00am on Thursdays. Such patterns may reflect production scheduling practices and highlight potential areas for optimization, such as load balancing or energy cost reduction strategies during peak hours. Therefore, the dashboard not only operationalizes the analytical objectives but also provides concrete evidence to guide energy-aware decision-making.

Moreover, one of the distinctive advantages of the multi-driven approach is that it not only reveals what can be done, but also what cannot. A clear example is DG3, "Prioritize Low-Energy Production Routes". While highly relevant from both operational and environmental perspectives, this goal could not be addressed analytically due to the absence of data on routing between work centers. IG6 would require detailed logs of inter-work center transitions, including operation sequences, timestamps, and energy consumption per route - information that is not currently captured in the available datasets. Nonetheless, this requirement was anticipated in the design of the analytical model, as suggested by the LLM, and the model was structured to accommodate such data in future iterations. As a result, while the data analysis is not possible yet, the model is already structured to incorporate it as soon as the necessary data becomes available. Still, the implemented dashboard shows the practical value of the multi-driven approach by transforming analytical objectives into visual tools that inform decision-making and expose data gaps that matter.

6. Conclusions and Future Work

This paper proposed a multi-driven approach to industrial decision-making that combines data-driven and requirements-driven perspectives, supported by LLMs. By leveraging LLMs to analyze metadata and user requirements through prompt-based interactions, we demonstrated how it is possible to accelerate the generation of conceptual and analytical models while maintaining alignment with strategic business objectives. The approach was applied to a real-world industrial case at Colep Packaging, an environment marked by operational complexity and variability. The results showed that, although validation and refinement by a data engineer are still necessary, LLMs can anticipate a significant portion of the

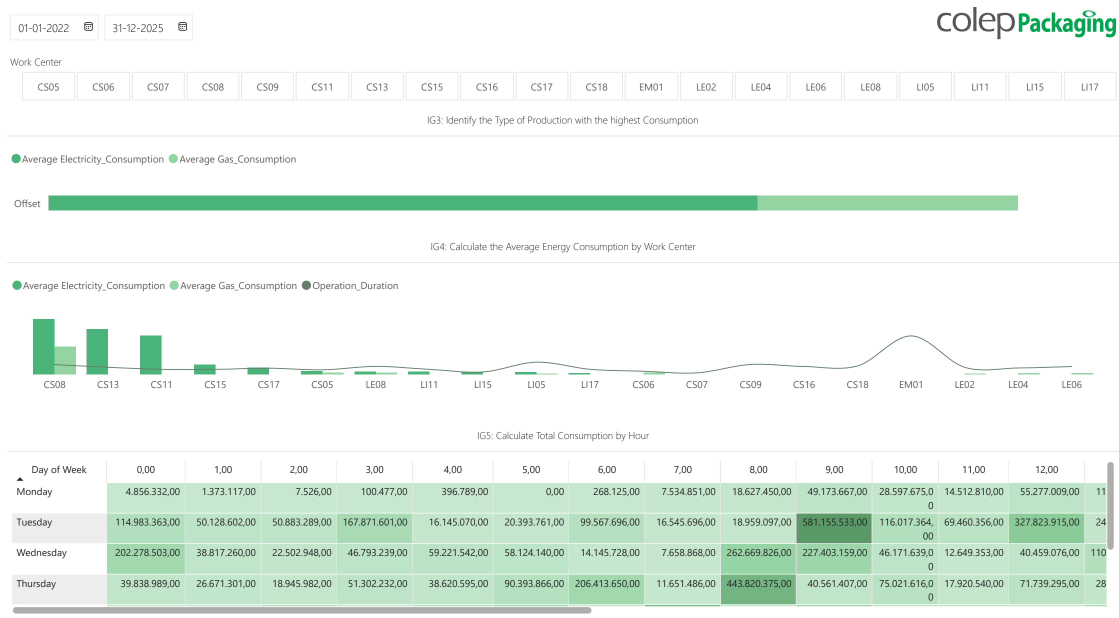

Figure 6: Dashboard of SG Reduce Energy Consumption, IG3, IG4 and IG5.

modeling work, reducing the initial manual effort and improving the alignment between available data and actual decision-maker needs. The approach proved especially valuable in revealing not only what can be analyzed with existing data but also the analytical limitations caused by information gaps.

In the context of the ongoing digital transition, the integration of LLMs into industrial modeling workflows exemplifies how advanced AI technologies can enhance data utilization and accelerate informed decision-making processes. Moreover, by enabling more efficient and targeted analyses, this approach supports green transition objectives by optimizing resource use, identifying inefficiencies, and guiding sustainable operational strategies. Thus, the proposed approach contributes to both digital innovation and environmental responsibility within industrial ecosystems.

Beyond these results, the case study revealed key learning points: LLMs accelerate model creation but require expert validation; prompt design is critical to avoid ambiguities; metadata alone can be enough preserving data confidentiality; and, the multi-driven approach both aligns data with requirements and exposes analytical gaps.

As future work, we could explore the use of alternative LLMs to assess their efficiency and effectiveness across the different modeling stages. Furthermore, the approach needs to be applied to different industrial use cases, addressing validity concerns since the current evaluation relies on a single case and a specific LLM version. Broadening the scope will strengthen the generalizability and reliability of the approach.

Acknowledgments

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Unit Project Scope UID/00319/Centro ALGORITMI (ALGORITMI/UM) and the European Union under the Next Generation EU, through a grant of the Portuguese Republic’s Recovery and Resilience Plan (RRP) Partnership Agreement, within the scope of the project PRODUTECH R3 – "Agenda Mobilizadora da Fileira das Tecnologias de Produção para a Reindustrialização". This paper uses icons made available by www.flaticon.com.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Grammarly for sentence polishing and rephrasing, and ChatGPT 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] R. Khusnutdinov, T. Wang, K. Müller, Data-driven decision-making in industry 5.0: Real-time analytics for smart manufacturing systems, Journal of Manufacturing Systems 79 (2024) 101243. doi:10.1016/j.jmsy.2024.101243.

[2] K. Peffers, T. Tuunanen, M. A. Rothenberger, S. Chatterjee, A design science research methodology for information systems research, Journal of Management Information Systems 24 (2007) 45–77. doi:10.2753/MIS0742-1222240302.

[3] Z. Li, A. Kumar, M. Santos, J. Trujillo, Llm-analytica: Large language models for autonomous industrial data understanding and dashboard generation, Expert Systems with Applications 239 (2024) 121348. doi:10.1016/j.eswa.2024.121348.

[4] S. Rizzi, Using ChatGPT to Refine Draft Conceptual Schemata in Supply-Driven Design of Multidimensional Cubes, in: 27th International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data (DOLAP'2025), co-located with EDBT/ICDT 2025, 2025.

[5] Y. Yang, B. Chen, K. Chen, G. Mussbacher, D. Varró, Multi-step iterative automated domain modeling with large language models, in: Proceedings of the ACM/IEEE 27th International Conference on Model Driven Engineering Languages and Systems, 2024, pp. 587–595.

[6] 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.

[7] 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. doi:doi.org/10.1145/3650400.3650478.

[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. doi:doi.org/10.1109/MODELS58315.2023.00037.

[9] 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. doi:doi.org/10.1007/s10270-023-01105-5.

[10] Colep Packaging, A nossa história, 2024. URL: https://colep-pk.com/pt-pt/quem-somos-colep/a-nossa-historia/.

[11] R. Kimball, M. Ross, The data warehouse toolkit: the definitive guide to dimensional modeling, 3rd ed., John Wiley & Sons, 2013.

[12] F. Dalpiaz, X. Franch, J. Horkoff, istar 2.0 language guide, CoRR abs/1605.07767 (2016). doi:https://doi.org/10.48550/arXiv.1605.07767. arXiv:1605.07767.

[13] 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://www.sciencedirect.com/science/article/pii/S0164121213002446. doi:https://doi.org/10.1016/j.jss.2013.10.011.

[14] A. Lavalle, A. Maté, M. Y. Santos, P. Guimarães, J. Trujillo, A. Santos, A methodology for the systematic design of storytelling dashboards applied to industry 4.0, Data & Knowledge Engineering 156 (2025) 102410. URL: https://www.sciencedirect.com/science/article/pii/S0169023X25000059. doi:https://doi.org/10.1016/j.datak.2025.102410.

[15] Z. Sun, Q. Wang, H. Wang, X. Zhang, J. Xu, Detection and mitigation of hallucination in large reasoning models: A mechanistic perspective (2025). doi:10.48550/arXiv.2505.12886.
[16] C. Vieira, V. Sousa, P. Guimarães, D. Rodrigues, A. Vieira, M. Y. Santos, Multi-driven and llm-assisted analytical modeling, 2025. URL: https://doi.org/10.5281/zenodo.15784597. doi:10.5281/zenodo.15784597.