Annals of Process Engineering and Management

www.apem.reapress.com

Ann. Proc. Eng. Manag. Vol. 2, No. 2 (2025) 91–100.

Paper Type: Original Article

Digital Transformation in Behshahr Industrial Development Holding Corporation: Leveraging the BPLLM Framework for Process Optimization and Business Growth

Mohammad Bagherian[1,*], Behnam Bagherian[2], Nasim Nahavandi[1]

[1] Department of Industrial Engineering, Faculty of Engineering, Tarbiat Modares University, Tehran, Iran; m.bagherian00@gmail.com; Nasim_nahavandi@yahoo.com.

[2] Department of Railway Transportation Engineering, Iran University of Science and Technology, Tehran, Iran; bagherian@rail.iust.ac.ir.

Citation:

Received: 19 August 2024 *Revised: 06 January 2025* *Accepted: 14 March 2025*	Bagherian, M., Bagherian, B., & Nahavandi, N. (2025). Digital transformation in Behshahr Industrial Development Holding Corporation: Leveraging the BPLLM framework for process optimization and business growth. *Annals of process engineering and management*, 2(2), 91-100.

Abstract

Process-aware Decision Support Systems (DSSs) have been traditionally enhanced by incorporating Artificial Intelligence (AI) functionalities to facilitate accelerated and informed decision-making. To advance its digital transformation and capitalize on AI, Behshahr Industrial Development Holding Corp. can tap into the innovative Business Process Large Language Model (BPLLM) framework to optimize business decisions and processes. The framework boosts flexibility, autonomy, and accuracy in supporting process-aware decision-making by blending advanced Natural Language Processing (NLP) and Business Process Management (BPM) capabilities. Key characteristics of BPLLM include process-specific informational Retrieval-Augmented Generation (RAG), fine-tuning models to meet organizational needs, and process-aware chunking. Evaluating this framework in various scenarios has revealed the high ability of BPLLM to identify activities and sequence flows and offer precise and relevant answers. By leveraging this framework, Behshahr Industrial Development Holding Corp can enhance supply chain management, production efficiency, and market and customer analysis, facilitating sustainable and intelligent growth.

Keywords: Business process large language model framework, Large language model, Business process management, Decision support system, Digital trans formation.

1 | Introduction

Artificial Intelligence (AI)-Augmented Business Process Management Systems (ABPMSs) represent a new generation of human-centric information systems characterized by great flexibility, autonomy, and extensive conversational and self-improving capabilities [1].

Corresponding Author: m.bagherian00@gmail.com

https://doi.org/10.48314/apem.v2i2.32

In this context, traditional process-aware Decision Support Systems (DSSs) have been empowered with AI to ensure quick and quality decisions by understanding and explaining the factors affecting decision-making [2].

Given the recognized potential of Large Language Models (LLMs) [3], [4] to assist human decisions in business contexts [1], this topic has not yet been widely examined in scientific sources, and to the best of our knowledge, no empirical validation has been conducted on the effectiveness of LLMs in supporting process-aware decisions. In light of these considerations, this study introduces a novel approach to BP analysis and description that adopts LLMs to implement a process-aware DSS. To enhance the conversational abilities of LLMs in answering BP-related tasks, we propose to adopt a process-aware Retrieval-Augmented Generation (RAG) framework in conjunction with fine-tuning.

The overall system, known as Business Process Large Language Model (BPLLM) and fine-tuned on a specific process model, helps users comprehend and execute various process-related tasks through natural language. In the proposed framework, a LLaMA 2 [5], [6] is mixed with a variant of RAG [7] tailored to deal with the specific aspects of the structural and behavioral representation of BPs.

This research explores the following research questions aimed at evaluating the performance of BPLLM within its components and its different settings:

- *RQ1: How can adopting BPLLMs impact process-aware decisions in Behshahr Industrial Development Holding Corp?*
- *RQ2: How does fine-tuning BPLLMs with business-specific data of Behshahr Industrial Development Holding Corp. affect their performance?*
- *RQ3: How can we adopt BPLLMs to optimize the supply chain and reduce Behshahr Industrial Development Holding Corp. expenses?*
- *RQ4: What is the role of process chunking and information retrieval within the BPLLM framework in boosting the productivity of manufacturing processes in the holding?*
- *RQ5: What is the effect of selecting process representation strategies on the efficiency of BPLLMs in industrial contexts like Behshahr Industrial Development Holding Corp?*

2 | Business Process Models

A business process model represents the BP structure, encompassing the activities to be executed and the constraints governing their sequence. It also encapsulates criteria indicating the initiation and termination of the process, along with details concerning participants, IT systems, and data [8]. Business Process Modeling Notation (BPMN) defines a process model that includes a set of graphical constructs, as the one reported in *Fig. 1*, which can be classified into three categories: 1) events, 2) activities, and 3) gateways.

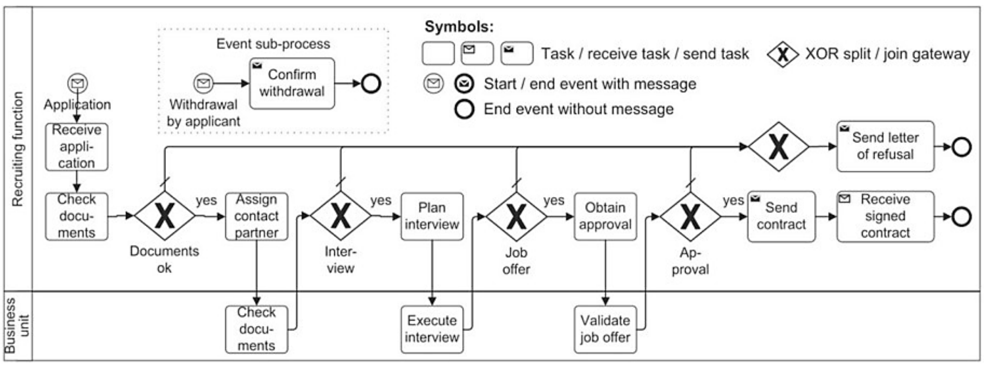

Fig. 1. A process model by Lohrmann and Reichert [9] specified in business process modeling notation.

3 | Research Method

Business process large language model

This section presents the BPLLM framework. *Fig. 2* illustrates the operational steps utilized by the BPLLM pipeline for answering BP-related queries.

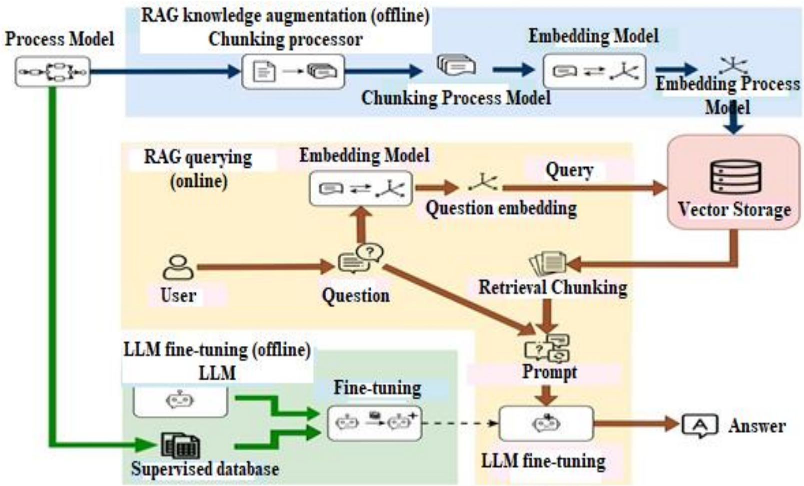

Fig. 2. Overall architecture of the business process large language model pipeline framework.

3.1 | Retrieval-Augmented Generation Knowledge Augmentation

To advance digital transformation, Behshahr Industrial Development Holding Corp. has adopted BPLLMs as key tools to optimize organizational processes.

Business process modeling notation-specific chunking

During this process, the BPMN file is notably scanned to isolate self-contained semantic chunks defined by the content within BPMN tags. These chunks are then added to the final list to be stored in the vector database [10]. *Fig. 3* epicts its execution on a BPMN file.

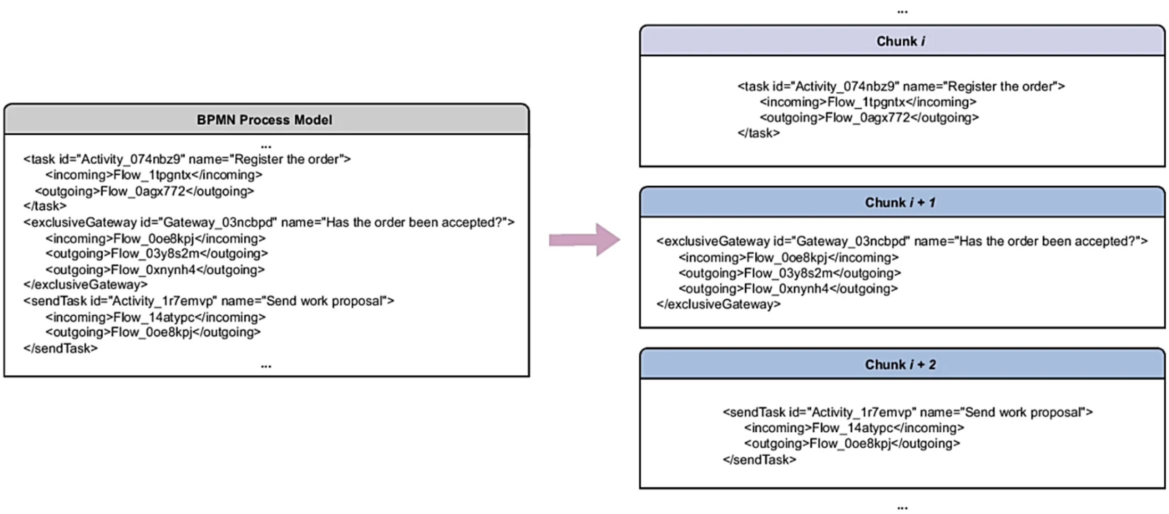

Fig. 3. An example of the application of the business process modeling notation-specific chunking technique.

Various embedding models suitable for semantic representation were used, which obtained the best results with the all-MiniLM-L6-v22 model.

3.2 | Retrieval-Augmented Generation Querying

This phase includes the Retrieval and Answering steps as follows:

Retrieval

In this step, the relevant chunks of the process model are retrieved, which are useful for generating accurate and contextually relevant answers to the user queries.

Answering

The answering step requires two main components: An LLM and its corresponding tokenizer. Initially, the prompt is crafted by combining the user query with the contextually relevant chunks retrieved in the previous phase (*Fig. 4*).

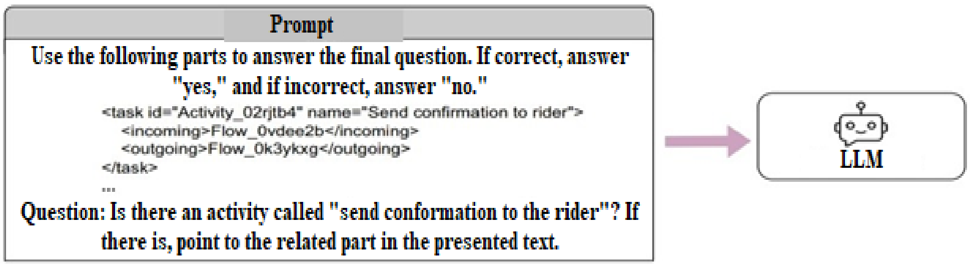

Fig. 4. An example of the prompt passed into the large language model.

4 | Empirical Validation

This section reviews the proposed experiments to answer the research questions in Section 1.

4.1 | Adopted Process Models

This research considers three distinct BPs. The first process is analyzed using textual Data Flow Graph (DFG) descriptions of activities and sequence flows, along with the corresponding BPMN process model. In contrast, the remaining two processes are evaluated solely through their BPMN models.

4.2 | Experimental Setting

The proposed BPLLM system is designed based on generative models and gives feedback to the user in natural language.

4.2.1 | Quantitative evaluation

This section proposes several experiments to answer the research questions reported in Section 1. The overall idea is to evaluate the BPLLM performance from different aspects (RAG, chunking, embedding, and fine-tuning). All the evaluations are carried out using the goods delivery process described in the previous section. The goods delivery process is represented using the DFG in natural language (Hereinafter, we use the term "Natural Text" for brevity). *Table 1* displays the natural language description of the DFG that identifies the process model fed into the LLM.

The queries utilized in these experiments should be answered to recognize structural information and behavioral aspects within the model. *Fig. 5* presents various checks sperformed on an excerpt of the goods delivery reference process.

Table 1. Extracts of the data flow graph in the natural language provided to the large language model.

Structural Process Information (Activities and Events)	Behavioral Process Information (Sequence Flows)
The goods delivery process involves the following activities: Customer order arrived Check payment Register the order ….. Select a rider and create a work proposal Pay the rider Receive customer satisfaction Order completed	The goods delivery process involves the following transitions: It starts with "Customer Order Arrived." It ends with "Order Completed." From "Create an Account," it proceeds with "Check Credit Card and Payment." ….. From "Retrieve Customer Information," it proceeds with "Check Payment." From "Select Rider and Create Work Proposal," it proceeds with "Send Work Proposal." From "Estimated Arrival Time," it proceeds with "Send the Customer the Waiting Time."

The effectiveness of the LLM and the BPLLM framework in describing various process-related aspects was estimated based on accuracy. Accuracy is the percentage of exact predictions made by the LLM in addressing such queries out of the total number of expected answers. The framework answers aligned with positive expected responses are designated True Positives (TP), while those matched with negative expected responses are designated True Negatives (TN). In contrast, False Positives (FP) occur when the framework yields positive responses when negative responses are expected. However, False Negatives (FN) arise when the framework produces negative responses when positive responses are expected. Accordingly, accuracy is defined as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}. \tag{1}$$

Something happened on our end. Here's the error code, in case you need it: 0x87E10BC6.

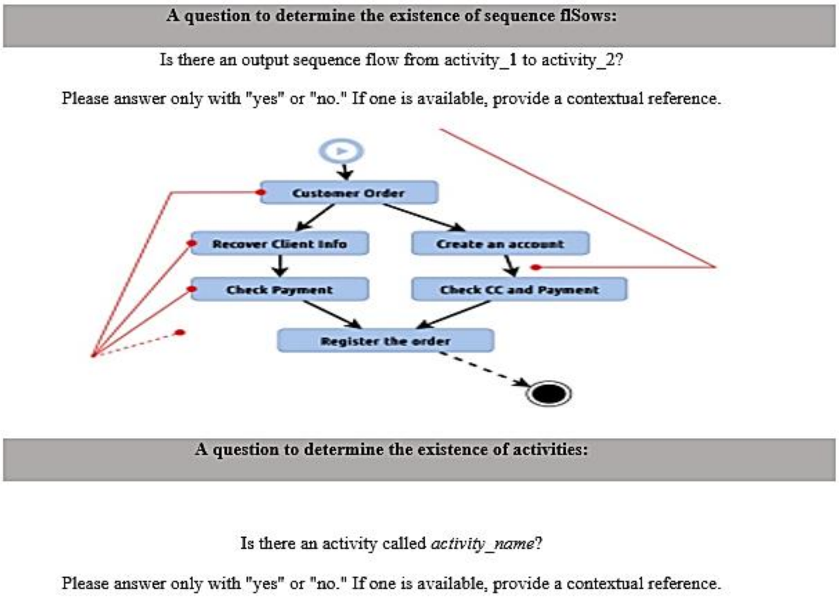

Fig. 5. Prompts for structural and behavioral evaluation.

4.2.2 | Qualitative evaluation

By preserving the optimal BPLLM configuration, an additional qualitative assessment of the pipeline was conducted, and more nuanced inquiries were posed about the whole BPMN process model.

4.3 | Evaluation Results

This section analyzes the BPLLM framework evaluation results under various experimental conditions based on the RQs given in Section 1.

Model's accuracy results: *Table 2* compares the accuracy of a pure LLM and the BPLLM framework for the food delivery process, offered in two forms: Natural text and BPMN.

Observations:

I. Significant enhancement in the BPLLM performance: Using BPLLM substantially increased the LLM's accuracy, which can be due to RAG's ability to retrieve relevant information and natural language representation, which assist in driving more accurate decisions.

II. Irrelevant graphical information in BPMN was removed: Due to LLMs' limited context windows, irrelevant graphical information in the BPMN file was removed, which directly impacted data processing and model accuracy.

III. Pure LLM illusion problem: In some cases, the pure LLM provided inaccurate or irrelevant information about the process model, highlighting its weakness in precisely understanding process models.

Chunking accuracy results:

Table 3: Accuracy for different process model representations (Natural text and BPMN) and distinct chunking techniques (No chunking, fixed-size, recursive, and BPMN-specific).

Table 2. Comparison of evaluation results for the pure large language model and retrieval-augmented generation framework.

Methodology	Process 1	Language Model	Accuracy
Pure LLM	None	Llama 2 13B	42.78%
RAG	Natural text	Llama 2 13B	65.48%
RAG	BPMN	Llama 2 13B	60.57%

The best result is shown in bold.

Table 3. Evaluation results after chunking and correcting prompts.

Process Model Representation	Chucking	Prompts	Language Model	Accuracy (%)
Natural text	No chucking	Uncorrected	Llama 2 13B	65.48
Natural text	No chucking	Corrected	Llama 2 13B	67.27
Natural text	Fixed-size	Uncorrected	Llama 2 13B	54.76
Natural text	Fixed-size	Corrected	Llama 2 13B	60.71
Natural text	Recursive	Uncorrected	Llama 2 13B	55.95
Natural text	Recursive	Corrected	Llama 2 13B	61.90
BPMN	No chucking	Uncorrected	Llama 2 13B	60.57
BPMN	No chucking	Corrected	Llama 2 13B	64.86
BPMN	Fixed-size	Uncorrected	Llama 2 13B	60.78
BPMN	Fixed-size	Corrected	Llama 2 13B	73.53
BPMN	Recursive	Uncorrected	Llama 2 13B	67.65
BPMN	Recursive	Corrected	Llama 2 13B	73.53
BPMN	BPMN-specific	Uncorrected	Llama 2 13B	68.63
BPMN	BPMN-specific	Corrected	Llama 2 13B	76.47
BPMN	BPMN-specific	Corrected	Llama 2 7B	68.47
BPMN	BPMN-specific	Corrected	Llama 3 8B Instruct	89.78
BPMN	BPMN-specific	Corrected	Llama 3.1 8B Instruct	90.38*
BPMN	BPMN-specific	Corrected	Mistral Instruct 7B v0.2	88.85
BPMN	BPMN-specific	Corrected	Mistral Instruct 7B v0.3	90.18
BPMN	BPMN-specific	Corrected	Mixtral 8x7B Instruct v0.1	79.81

*The best result

These findings generated t-SNE charts for the most effective combinations, as depicted in *Fig. 6*.

Table 5 reports the outcomes of the experiments aimed at evaluating the influence of fine-tuning on the BPLLM performance. The table presents the BPLLM accuracy for different Parameter-Efficient Fine-Tuning (PEFT) quantizations (int4 and int8).

The comparison of *Table 5* and the best result reported in *Table 4* indicates that fine-tuning increased accuracy.

Table 4. Evaluation results for different process representations and embedding models.

Process Model Representation	Embedding Model	Language Model	Accuracy (%)
Natural text	All-MiniLM-L6-v2	Llama 3.1 8B Instruct	87.50
Natural text	Paraphrase-xlm-r-multilingual-v1	Llama 3.1 8B Instruct	85.58
Natural text	Bert-finetuned-BPMN	Llama 3.1 8B Instruct	87.88
BPMN	All-MiniLM-L6-v2	Llama 3.1 8B Instruct	90.38*
BPMN	Paraphrase-xlm-r-multilingual-v1	Llama 3.1 8B Instruct	90.20
BPMN	Bert-finetuned-BPMN	Llama 3.1 8B Instruct	69.23

*The best result

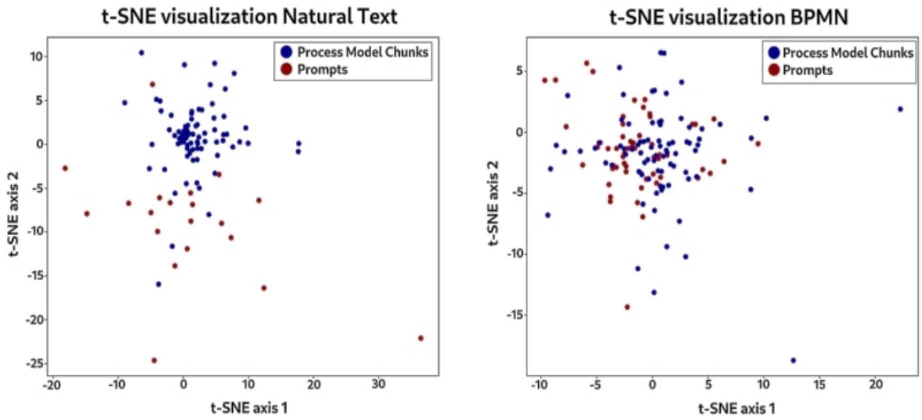

Fig. 6. T-SNE charts depicting the embeddings of the prompts and chunks derived from the natural text (Left) and the business process modeling notation (Right) process model. The embeddings are computed using bert-finetuned-business process modeling notation (Left) and all-MiniLM-L6-v2 (Right).

Continuing our analysis of the results, we examined the potential effects of enlarging the knowledge base within the vector database by incorporating multiple process models into the BPLLM.

The impact of integrating two distinct BPs (i.e., goods delivery and reimbursement process models) into the knowledge base was initially examined. The findings indicated a slight reduction in the performance of the BPLLM pipeline. However, when two similar BP models (e.g., goods delivery and E-commerce) were incorporated, a marked decline was observed in accuracy for both the base BPLLM and the pipeline using the fine-tuned LLM.

Table 5. Evaluation results for parameter-efficient fine-tuning quantization and enlarged knowledge base.

Process Models	Fine-Tuning	PEFT Quantization	Accuracy (%)
Goods delivery	✓	Int4	92.31*
Goods delivery	✓	Int8	90.58
Goods delivery, reimbursement		-	82.16
Goods delivery, reimbursement	✓	Int4	84.62
Goods delivery, reimbursement	✓	Int8	81.87
Goods delivery, e-commerce		-	80.26
Goods delivery, e-commerce	✓	Int4	80.53
Goods delivery, e-commerce	✓	Int8	79.85

*The best result

Does the model contain duplicate transitions?	**Which activities can never co-occur?**
No, there is no duplicate transition in the proposed BPMN. Each sequence flow has a unique ID that connects two distinct elements, ensuring there are no duplicate transitions in the model.	Activities that can never co-occur: "Create a user account" (Activity_1uxpywg) and " retrieve customer information" (Activity_0zlhwln) "Check payment" (Activity_0mt2w5k) and "Check credit card and payment" (Activity_1x5mf5x). These two activity pairs are separated by exclusive gateways, meaning only a single activity can be executed simultaneously.
Can you suggest the shortest path toward reaching the result after the "register order" activity?	Which activities follow the "customer order" event after the exclusive gateway?
The shortest path toward reaching the outcome after the "register order" activity: 1) Activity_074nbz9 "Register order" 2) Gateway_1lf96zg 3) Activity_05qnapc "Select rider and create work proposal" 4) Activity_1r7emvp "Send work proposal" 5) Gateway_03ncbpd "Is the order accepted?" 6) Flow_03y8s2m 7) Activity_00fcmsc "Rider accepted the order" 8) Gateway_0yayjl1 9) Activity_02rjtb4 "Send confirmation to the rider" 10) Activity_1nwzzzm "Send the customer the waiting time" However, two parallel paths from Gateway_0yayjl1 exist, with the other one as: 11) Activity_1623ajo "Estimate the arrival time" 12) Activity_1nwzzzm "Send the customer the waiting time" Both paths arrive at an outcome, though the first one is shorter.	Activities occurring following the "customer order" event and after the "has the customer registered" exclusive gateway include: • If the response to the gateway is "yes", "retrieve customer information" • If the response to the gateway is "no, "create a user account. These activities are connected to the "customer order" event through sequence flows Flow_08ptwir and Flow_0npkvg0. **Can you describe the meaning of the path between "register customer order" and "rider accepted the work proposal"?** The path between "register customer order" and "rider accepted the work proposal" is: Activities occurring after the "register customer order" event and linked to the "Is customer registered?" exclusive gateway are: If the response to the gateway is "yes," the "retrieve customer information" activity is performed. If the response [to the gateway] is "no," the "create a user account" activity is performed. These activities are connected to the "register customer order" event through "Flow_08ptwir" and "Flow_0npkvg0" sequence flows.

Fig. 7. Excerpts of qualitative evaluation results.

These findings underscore the importance of careful management of the size and composition of the knowledge base within BPLLM, particularly when multiple process models with varying degrees of similarity are incorporated.

4.3.1 | Qualitative evaluation discussion

BPLLM's capability to appropriately respond to process-centered inquiries posed by a group of potential users was also evaluated, considering the whole goods delivery BPMN model as input. The complete set of questions and answers can be found in the replication package. *Fig. 7* presents an informative sample for this experimentation. These findings demonstrate the ability of BPLLM to: 1) identify duplicate elements in a process model and ensure process integrity and correctness, 2) navigate and optimize process paths for efficiency analysis, 3) understand exclusive gateways and mutually exclusive activities, 4) trace process flows, and 5) explain sequences of events, implying comprehension of the rationale behind BP steps.

Based on the obtained answers, the selected users posit that the framework can promptly provide satisfactory responses, thereby establishing its reliability for everyday usage.

5 | Conclusion

BPLLM's capabilities and achievements in Behshahr:

I. Increased process analysis capability: The BPLLM framework improves structural and behavioral knowledge of the processes by integrating RAG technologies and LLM fine-tuning. This approach generates accurate and contextually relevant answers to queries related to manufacturing processes, supply chain, and other key operations in the holding.

II. Promoted language interactions: Leveraging LLM fine-tuning, BPLLM reinforces user support capability in comprehending strategic processes and decisions.

III. Process-aware chunking: An approach based on process-aware chunking is proposed, which identifies and selects the most appropriate options for holding by evaluating various embedding models.

IV. Business growth support: BPLLM has become an effective tool for digital transformation and competitiveness in Behshahr Industrial Development Holding Corp., thanks to its ability to optimize processes and increase productivity.

As part of our digital transformation path, the following measures are recommended:

I. Analysis of executive aspects of processes: Reviewing data associated with execution times and activity costs to offer practical and useful information to managers will help boost productivity and mitigate operating expenses in the holding.

II. Evaluation and improvement of embedding models: Using more advanced models to compute the proximity or distance of execution traces in reporting events, providing more accurate viewpoints of process performance.

III. Integration of knowledge graphs: Exploring the possibility of combining knowledge graphs with BPLLM to boost accuracy in the retrieval stage and offer more profound insights.

IV. Utilizing symbolic AI solvers: Combining symbolic AI reasoning capabilities with the BPLLM framework can facilitate more intelligent decisions in complex business processes.

V. Concluding remarks: With its unique abilities in process analysis and optimization, the BPLLM framework plays a vital role in digital transformation in Behshahr Industrial Development Holding Corp. By emphasizing productivity enhancement, cost reduction, and sustainable growth support, this approach can pave the way for intelligent development and further competitiveness of the holding in both domestic and international markets.

References

[1] Dumas, M., Fournier, F., Limonad, L., Marrella, A., Montali, M., Rehse, J. R., … ., & Weber, I. (2023). AI-augmented business process management systems: A research manifesto. *Association for computing machinery transactions on management information systems*, *14*(1), 1–19. https://doi.org/10.1145/3576047

[2] Agarwal, P., Gao, B., Huo, S., Reddy, P., Dechu, S., Obeidi, Y., … ., & Carbajales, S. (2022). A process-aware decision support system for business processes. *Proceedings of the 28th ACM sigkdd conference on knowledge discovery and data mining* (pp. 2673–2681). Association for computing machinery. https://doi.org/10.1145/3534678.3539088

[3] Estrada-Torres, B., Del-Río-Ortega, A., & Resinas, M. (2024). Mapping the landscape: Exploring large language model applications in business process management. *Enterprise, business-process and information systems modeling* (pp. 22–31). Cham: Springer, Cham. https://doi.org/10.1007/978-3-031-61007-3_3

[4] Fahland, D., Fournier, F., Limonad, L., Skarbovsky, I., & Swevels, A. J. E. (2024). *How well can large language models explain business processes?* https://doi.org/10.48550/arXiv.2401.12846

[5] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., … ., & Lample, G. (2023). *LLaMA: Open and efficient foundation language models.* https://doi.org/10.48550/ARXIV.2302.13971

[6] Touvron, H., Martin, L., Stone, K., Albert, P., Almahairi, A., Babaei, Y., … ., & Scialom, T. (2023). *Llama 2: Open foundation and fine-tuned chat models*. https://doi.org/10.48550/arXiv.2307.09288

[7] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., … ., & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in neural information processing systems*, *2020-Decem*, 9459–9474. http://dx.doi.org/10.48550/arXiv.2005.11401

[8] Dumas, M., La Rosa, M., Mendling, J., Reijers, H. A. (2018). *Fundamentals of business process management*. Springer. https://doi.org/10.1007/978-3-662-56509-4

[9] Lohrmann, M., & Reichert, M. (2016). Effective application of process improvement patterns to business processes. *Software & systems modeling*, *15*(2), 353–375. https://doi.org/10.1007/s10270-014-0443-z

[10] Agostinelli, S., De Luzi, F., Canito, U., Ferraro, J., Marrella, A., & Mecella, M. (2022). A data-centric approach to design resilient-aware process models in BPMN. *Business process management forum* (pp. 38–54). Springer, Cham. http://dx.doi.org/10.1007/978-3-031-16171-1_3