

Customised and Controllable Generative Language Models for Business Applications

Xiaohan Su¹[0009-0003-6159-7453]

University of Technology Sydney, 15 Broadway, Ultimo NSW 2007, Australia
`Xiaohan.Su@student.uts.edu.au`

Large language models (LLMs) have demonstrated remarkable capabilities in natural language understanding and generation. However, their direct application to process mining presents significant challenges. Process mining requires structured reasoning, domain expertise, and sequential logic to analyze and optimize workflows. Process mining is a technique that analyzes event logs to visualize, monitor, and improve business processes by discovering actual workflows, checking compliance, and identifying bottlenecks[1]. Today, it is widely used in business analysis and operational management for data-driven optimization. LLMs, being trained on broad and often unstructured data, lack specialized domain knowledge and fail to accurately model the stepwise execution of business processes. Furthermore, fine-tuning LLMs for enterprise-specific workflows is costly, both in computational resources and maintainability, making it impractical for many organizations.

To address these challenges, graph-based retrieval-augmented generation (Graph RAG) extends this approach by incorporating structured knowledge representations[2]. By organizing workflow information into graph structures, Graph RAG captures hierarchical dependencies, contextual relationships, and organizational structures. However, despite its advantages, Graph RAG does not explicitly encode the sequential order of workflow steps. While it successfully models interdependencies between different organizational units or process components, it lacks a mechanism to enforce process execution constraints. Consequently, Graph RAG remains inadequate for process mining applications where both contextual relationships and execution sequences are essential.

To bridge this gap, we propose GRAG4PM, a novel retrieval-augmented framework specifically designed for process mining. GRAG4PM extends Graph RAG by integrating both hierarchical and sequential dependencies into the retrieval process. This approach ensures that retrieved knowledge aligns with enterprise workflow execution, preserving the logical flow of tasks while maintaining contextual relevance. By leveraging structured workflow representations, GRAG4PM enhances process-aware LLM responses, making them both accurate and interpretable.

In Figure1, GRAG4PM extends Graph RAG by integrating structural relationships with execution order constraints, ensuring that process mining tasks retain both contextual relevance and logical flow. The framework employs a dual-indexing mechanism that combines graph-based retrieval with semantic retrieval. Initially, graph indexing retrieves workflow nodes based on structured dependencies, while vector indexing retrieves semantically similar nodes. Tabell1 displays

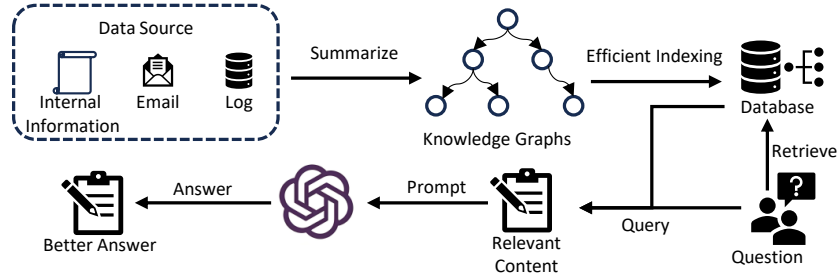


Fig. 1. GRAG4PM Workflow Description

the progress and timeline, future work will explore optimizing hierarchical filtering mechanisms to improve computational efficiency and adapt GRAG4PM for real-time process analysis. Additionally, further research into hybrid retrieval strategies incorporating event log analysis could enhance the framework’s ability to model dynamic workflow adaptations. By addressing these challenges, GRAG4PM aims to bridge the gap between AI-driven natural language processing and structured process mining, fostering more intelligent enterprise workflow automation. This research aims to develop a customizable and controllable

Table 1. Progress and Timeline

Time Period	Progress
Nov 2023 – Nov 2024	Literature review and collection of background information
Nov 2024 – Jan 2025	Completion and submission of a survey paper
Dec 2024 – Feb 2025	Design of the GRAG4PM framework and dataset construction, with two papers submitted
Mar 2025 – Jul 2025	Implementation and adaptation of the framework, including algorithm optimization, with two expected papers
Aug 2025 – Jun 2026	Consolidation of research findings and dissertation writing

language generation model tailored for enterprise process mining. The study is expected to produce three high-quality research papers and a comprehensive model usage report, providing practical guidelines for enterprise adoption. Through these contributions, this research bridges AI-driven language models with structured enterprise workflow optimization.

References

1. Van der Aalst, W.: Process Mining: Overview and Opportunities. *ACM Transactions on Management Information Systems (TMIS)* 3(2), 1–17 (2012).
2. He, X., Tian, Y., Sun, Y., Chawla, N.V., Laurent, T., LeCun, Y., Bresson, X., Hooi, B.: G-Retriever: Retrieval-Augmented Generation for Textual Graph Understanding and Question Answering. *arXiv preprint arXiv:2402.07630* (2024).