An Automating Tendering Performance Interpretation via Process Mining and an LLM-Based Agent

Ferial Hendrata
Department of Industrial and Systems Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
02411960010005@student.its.ac.id

Iwan Vanany
Department of Industrial and Systems Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
vanany@ie.its.ac.id (corresponding author)

Patdono Suwignjo
Department of Industrial and Systems Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
patdono@ie.its.ac.id

Nurhadi Siswanto
Department of Industrial and Systems Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
siswanto@ie.its.ac.id

Received: 18 November 2025 | Revised: 10 January 2026 | Accepted: 23 January 2026

 DOI: https://doi.org/10.48084/etasr.16343

ABSTRACT

Digital transformation in procurement, particularly e-tendering, is generating rich transaction data that can support evidence-based performance evaluation and decision-making. Process mining provides a data-driven methodology for process execution; however, its outputs, including process models and activity performance statistics, frequently pose interpretative challenges for non-expert users and necessitate manual analysis. This challenge can slow decision-making and limit the organizational adoption of process-mining-based performance evaluations. This study presents an end-to-end framework that integrates event-log-based process mining with an LLM-based AI agent to automate the interpretation of descriptive tender performance by translating process-mining evidence into decision-oriented narratives aligned with predefined analytical objectives. In a case study, the agent used a DFG artifact to extract 16 activity labels and their average durations with 100% accuracy, validated against process-mining activity statistics. The agent further produced descriptive narratives that summarize activity performance, identify bottlenecks and dominant delay paths, and prioritize improvement recommendations. An expert review confirmed that the interpretations and suggestions make sense, fit the situation, and can be used for tendering. Overall, the proposed framework operationalizes an evidence-grounded interpretation layer that reduces reliance on manual interpretation, improves the usability of process-mining outputs for procurement decision-makers, and advances automation within the descriptive analytics layer of the augmented Business Process Management (BPM) pyramid.

Keywords-e-tendering; e-procurement; process mining; LLMs; AI agents; process performance

I. INTRODUCTION

The digital transformation of procurement, currently referred to as Procurement 4.0, has accelerated the improvement of e-procurement capabilities, accompanied by the adoption of advanced technologies such as the Internet of Things (IoT) and Artificial Intelligence (AI). This transition signifies a paradigm shift from transactional procurement activities to a more strategic function and increased data utilization [1]. The use of e-procurement is associated with non-financial performance improvements such as transparency, coordination, efficiency, and effectiveness, which strengthen managerial effectiveness [2]. Process efficiency is a key driver of e-procurement success, reflected in shorter cycle times, faster task completion, lower transaction costs, tighter control, and improved coordination and system integration [3-5].

Despite these benefits, organizations may still have limited visibility into how procurement activities unfold in day-to-day execution. In e-tendering, as a form of e-procurement, data transactions are recorded continuously in the underlying systems, creating a solid basis for performance evaluation and evidence-based decision-making. However, conventional performance evaluation approaches, such as strategic, outcome-based, and output-based assessments, mainly provide aggregated, high-level views and remain insufficient for capturing operational realities [3, 5, 6]. Strategic evaluations tend to be normative, outcome-based measures emphasize longer-term impacts without detailing process mechanisms, and output-based indicators focus on end results while overlooking execution behavior. As a consequence, important process dynamics, such as process variability, activity performance, and bottlenecks, can remain hidden and unexplored in the evaluation of tender operations.

To address these limitations, process mining, as a data-driven approach, can extract process-centric knowledge from event logs to discover models, examine conformance, and improve operations, driven by the goals of transparency, performance improvement, and compliance [7]. Several studies have explored the application of process mining in e-procurement, including tendering processes, showing how it can identify process variants [8], diagnose performance bottlenecks [9], and detect compliance issues [10]. However, this technology introduces complexities, especially for users without a deep technical background, because interpreting process model diagrams and process performance metrics often requires specialized skills [11]. Although existing studies have yielded valuable findings and analytical outputs, the interpretation of these results typically depends on manual analysis by experts. This leaves a gap for further development of an end-to-end approach that translates process mining results into actionable insights that can be readily understood and used by non-expert users without manual intervention.

In response to this interpretability challenge, the integration of process mining with AI has received growing attention in the Business Process Management (BPM) domain. Within the augmented BPM pyramid, conversational systems such as Large Language Models (LLMs) can support non-expert workers by transforming process mining findings into more accessible narratives in the descriptive layer [12]. A survey on intelligent cross-organizational process mining positions LLM integration at the performance diagnosis layer, where model outputs and process metrics are interpreted into narratives that support performance diagnosis [13]. Applied studies propose different integration forms, including rule-based interfaces for querying process mining tools through predefined semantic rules, although generalization and scalability remain constrained [14]. Other work combines process mining with domain knowledge and LLM-supported reasoning to strengthen anomaly detection and formulate optimization insights, particularly in financial process environments [15]. In public procurement, LLMs have also been used to extract temporal information from textual documents for event log enrichment, supporting subsequent process analysis and compliance evaluation [16]. These efforts suggest that LLMs can improve accessibility. However, in these settings, LLMs mainly act as supporting components, while decision-oriented interpretation of process mining results still depends on external reasoning rather than being carried out automatically by an analytical model.

Recent advances in LLMs offer a promising direction for translating process mining technical artifacts into natural-language explanations [17]. However, for tender performance analysis, a key gap remains because an LLM-only approach is often insufficient to:

1. Reliably extract quantitative information from multimodal process artifacts,
2. Execute a consistent multi-step analytical workflow (extract-verify-diagnose-summarize),
3. Provide outputs that are systematically validated for accuracy and coverage against expert interpretation.

AI-agent architectures can mitigate these limitations by orchestrating structured tasks, tool usage, and constrained outputs to produce context-specific interpretations aligned with analytical objectives [18, 19]. Based on theoretical perspectives on agentic systems, AI agents can extend the role of LLMs from language generation to goal-oriented operations by enabling interaction, reasoning, and adaptive decision-making for complex workflows [20]. In addition, context-aware smart agents have been proposed to improve user performance in data analysis tasks by generating and integrating contextual information into concise natural-language queries [21], and behavior-driven LLM agents have been applied in dynamic environments to enable adaptive multi-step decision-making [22]. Although these studies are not situated in the BPM domain, they illustrate transferable agent design patterns that can inform process analysis and performance interpretation.

To provide a structured literature review of existing approaches and their limitations, Table I synthesizes key prior studies on process mining for e-procurement and LLM/agent-based approaches relevant to automated interpretation, reporting each study's main focus, methodology/approach, key findings, and identified limitations/gaps. Based on this review, the prior literature can be categorized into four distinct directions: (i) procurement process mining without automated interpretation [8-10], (ii) process mining with LLMs beyond the procurement domain without automated interpretation [14, 15], (iii) LLM-based AI agents without process mining integration [21, 22], and (iv) procurement process mining with LLMs without automated interpretation [16]. This synthesis indicates that a gap remains in tendering performance analysis using process mining, specifically the lack of support for the automated interpretation of process mining outputs into decision-oriented narratives. This study differs from prior work by delivering an end-to-end integration in which an LLM-based agent automatically interprets process-mining outputs into evidence-grounded narratives for decision-oriented tendering performance analysis.

To address this gap, this study pursues two objectives. First, it develops an end-to-end framework that integrates event-log-based process mining with an LLM-based agent to automate tendering performance analysis from evidence generation to narrative reporting. Second, it designs an automated workflow in which the agent interprets process mining outputs and produces controlled narratives aligned with predefined analytical objectives. Accordingly, this study makes two contributions. First, it proposes an end-to-end architecture that bridges process mining outputs and decision-oriented reporting for tendering performance analysis. Second, it defines a controlled analytical workflow implemented by an LLM-based agent to generate evidence-grounded narratives that support managerial decision-making.

II. PROPOSED FRAMEWORK

The proposed framework provides an end-to-end workflow that transforms event logs extracted from the e-tendering database into an interpretable process performance analysis report. As illustrated in Figure 1, the framework is organized into four interrelated stages: (i) Data Preparation and Extraction, (ii) Process Mining, (iii) LLM-Based AI Agent, and (iv) Validation and Evaluation. The first stage converts raw e-tendering records into structured and consistent event logs. The second stage applies process mining tools to derive the process model and relevant performance statistics from the event logs. The third stage introduces an LLM-based AI agent that translates technical outputs into textual insights to facilitate understanding by non-technical stakeholders. Finally, this stage validates and evaluates the generated outputs as a process performance analysis report in terms of numerical accuracy, logical consistency, and contextual relevance.

TABLE I. SUMMARY OF PRIOR STUDIES

Ref.	Main focus	Method/approach & key findings	Identified limitations/gaps
[8]	Process mining for public procurement	PM case study on public procurement dataset; demonstrates discovery/ analysis of procurement process variant	Interpretation of analysis results is manual and not yet supported by automation.
[9]	Process mining for e-tendering performance	PM-based performance analysis on real-world Indonesian construction tender event logs; identifies performance issues/bottlenecks from event data.	Interpretation of analysis results is manual and not yet supported by automation.
[10]	Compliance checking in procurement	Heuristics/PM-based compliance checking on extracted procurement logs; detects compliance deviations.	Interpretation of analysis results is manual and not yet supported by automation.
[14]	Natural Language (NL) access to process mining	NL querying interface for PM tools; improves accessibility for querying PM results.	NL does not address the interpretation of process mining output.
[15]	PM + LLM support for anomaly detection	Domain-knowledge–enhanced PM with LLM-supported reasoning for anomaly detection/ explanations in financial processes.	LLMs are used as additional reasoning, not as an end-to-end automation LLM-based agent.
[16]	LLM/NLP for event log enrichment in European public procurement analysis	NLP/LLM extracts temporal info from documents to enrich event logs for subsequent PM/compliance analysis.	LLMs are used for improving event log extraction, not for interpreting the results of the analysis.
[21]	LLM-based agent for data analysis	Context-aware LLM agent supports data analysis (e.g., generating context-aware queries); improves user efficiency.	LLM-based agent without process mining integration
[22]	LLM agent behavior control	Behavior-driven LLM agent with personality/behavior control in dynamic environments.	LLM-based agent without process mining integration
This study	LLM-based agent interpretation of PM outputs in e-tendering	Multimodal, LLM-based agent workflow interprets DFG artifacts and produces validated diagnostic narratives for descriptive process analytics.	Addresses the end-to-end automation gap by combining process mining with LLM-based AI agents to facilitate automated descriptive performance analysis

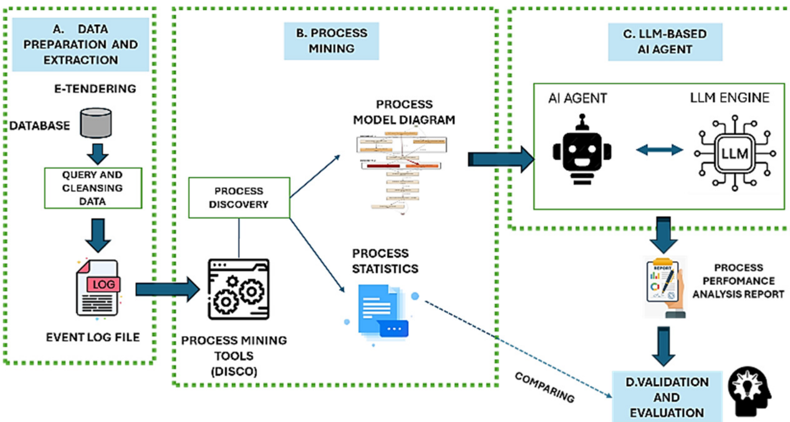

Fig. 1. Process mining and LLM-based agent integration framework

A. Data Preparation and Extraction

The framework begins with the preparation of event data extracted from an e-tendering database. Ideally, each tendering case consists of sequentially executed activities, ranging from early stages, such as announcements, to final stages, such as contract signing. At the very least, each recorded activity must have the tendering code, activity name, start time, and end time. Data preparation involves querying the database followed by data cleaning to ensure accuracy, eliminate duplicates, and guarantee consistency. The cleaned data are then transformed into event log records and exported into standard formats, such as .xls or .csv, to enable subsequent processing in process mining tools.

B. Process Mining: Discovery and Performance Analysis

After importing event log data, process mining tools such as Disco and ProM perform process discovery. In practice, the main output of process discovery is a process model reconstructed from event log data. Process model diagrams, such as DFG, display activity nodes and their flows, along with information about the frequency or average (mean) duration of each activity diagram box. In addition, process mining tools provide descriptive statistics that offer numerical summaries in tabular form at the activity level, frequency, and average duration. Both artifacts represent analytical outputs from process mining algorithms and can be used to assess process performance based on activity execution times, serving as a foundation for analyses such as bottleneck identification. In this study, descriptive statistics produced by process mining are also used as reference values to validate the LLM-based AI agent's reading of the process model diagram.

C. LLM-Based AI Agent

The third stage introduces the core innovation of the framework: an LLM-based AI agent that bridges technical outputs and managerial interpretation. In this stage, the process model diagram produced by process mining in DFG format is organized by the AI agent to be suitable as multimodal input for the LLM. The agent also prepares predefined structured prompts aligned with the analytical objectives before submitting the inputs to the LLM. The LLM then returns interpretation results that can be compiled into different report formats, such as HTML or DOC. Through these functions, the LLM-based AI agent operates as an analytical model that supports automated interpretation.

LLMs can process both textual and non-textual inputs, such as diagrams, tables, and charts, through multimodal capabilities. In [23], multimodal LLMs combine visual inputs with textual context so that process flow graphs, including the DFG format, can be interpreted as sequential structures, parallel branches, or loops and subsequently translated into narrative explanations suitable for business process analysis.

The AI agent plays a central role in organizing process model inputs as data references and combining them with structured prompts before submitting them to the LLM. The structured prompt strategy follows a systematic composition of prompt components, including role definition, task instructions, data context, constraints, and output format or style, which are arranged to support consistent instruction, following [24] as explained in Table II. This prompt structure is specifically designed to guide the analysis of activity-level performance, with a particular focus on duration-based performance indicators derived from process mining outputs. This approach allows LLMs to interpret instructions consistently based on data obtained from process mining results; therefore, it is expected to produce accurate and relevant narrative interpretations.

TABLE II. LLM STRUCTURED PROMPT

Prompt component	Definition
Profile/Role	Who or what the model is acting as.
Context	Background information and context that the model needs to refer to.
Directive	The core intent of the prompt, often in the form of an instruction or a question.
Constraints	Restrictions on what the model must adhere to when generating a response.
Output format/style	The type, format, or style of the output.

D. Evaluation and Validation

In the final stage, two complementary steps are performed: validation and evaluation of the process performance analysis report output from the LLM-based agent. Validation assesses the accuracy of the numerical output generated by the agent by comparing the average duration of activity extracted from the diagram with the reference results obtained from the process mining calculations. Evaluation is performed through experts who assess the quality of the resulting interpretations, with a focus on logical accuracy, clarity, relevance, and depth. This evaluation involves interviews with manager-level procurement professionals who are experienced users of the e-procurement system. The combination of quantitative validation and qualitative evaluation assesses whether the analytical model produced factually accurate outputs and generated meaningful, reliable, and useful interpretations for analyzing tender process performance.

III. RESULTS

A. Case Study

The proposed model was implemented by taking data from a construction e-tendering system for plantation companies in Indonesia. Company X manages, processes, and markets plantation commodities such as palm oil, rubber, sugarcane, tea, coffee, cocoa, and tobacco. The agricultural land area owned by the case study company is around 1.2 million hectares, with a total planted agricultural land area of around 820,000 hectares.

Companies manage their tendering package using e-tendering systems. Tenderers, including both the tendering committee and vendors, benefit from the e-tendering system in several ways. The system facilitates and accelerates the procurement workflow, from the initial announcement to contract signing. In construction tendering, the use of a post-qualification approach enables bidder qualifications to be verified after bid evaluation. In addition, bid documents and related information are submitted electronically, eliminating the need to manage large volumes of physical files because all submissions are handled digitally. The e-tendering system automatically verifies the validity of bidder data, thereby minimizing human error and negligence. The administrative burden of the tendering committee is reduced because the workflow is integrated and all data is recorded digitally. Overall, this e-tendering system has simplified the duties of the tendering committee from the beginning to the end of the construction procurement process. The tendering process involves 16 types of activities, ranging from announcement to contract signing, as illustrated in Figure 2.

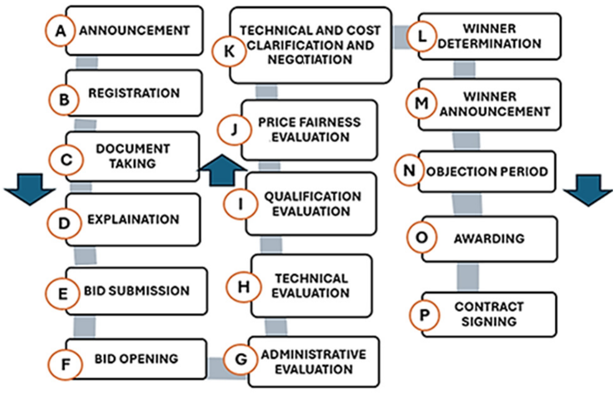

Fig. 2. Tendering activity.

This section presents the results based on the analysis stages carried out using the process mining approach with the help of Disco's academic license software, chosen for its ability to automatically process event logs to produce process visualizations and activity performance measurements. Data were extracted from the e-procurement system, converted into a compatible format (.xls) for import into Disco, and then analyzed using n8n as an LLM-based AI agent after processing by Disco.

B. Data Extraction

The study began with data extraction from the company's e-procurement system, which records all tendering process activities in the form of event logs. This activity data was then converted into an Excel file format (xls) to make it compatible with Disco's process mining software. This log data covers a total of 814 construction tendering packages from company X over a three-year period, with 8,456 recorded activity events. Table III contains an example of tendering activity data extracted from e-procurement.

TABLE III. EVENT LOG EXAMPLE

Tender id	tender_activity	start_time	end_time
53131	Announcement	02/02/2018 08:00	06/02/2018 08:00
53131	Registration	02/02/2018 08:00	06/02/2018 08:00
53131	Download document	02/02/2018 08:00	06/02/2018 08:00
53131	Explanation	06/02/2018 08:00	06/02/2018 09:00
53131	Bid Submission	06/02/2018 09:00	09/02/2018 12:00
53131	Bid Opening	09/02/2018 12:00	13/02/2018 17:00
53131	Administrative Evaluation	09/02/2018 12:00	13/02/2018 17:00

Next, Disco reads the file and configures it to map the data columns from the Excel format to the data elements required by this tool, as shown in Table IV. Each row in the imported file represents one activity or event in the tendering process, with four data columns: tender identifier, activity name, start time, and end time of the activity. Once the attribute mapping is correctly configured, the data import process can be started to convert the data in the extracted file into the Disco process data set.

TABLE IV. DATA CONFIGURATION MAPPING

Extracted file columns	Disco setting attribute
tender_id	Case id
tender_activity	Activity
start_time	Timestamp
end_time	Timestamp

C. Process Mining Result

In this e-tender case, process discovery produced two outputs for further analysis using LLM. The first is a process model in the form of a DFG in Figure 3, which will be used for analyzing process activities with a multimodal LLM. In addition to illustrating the activity flow, the process model also includes the average duration of each activity in the activity diagram boxes and color degradation within the boxes, where darker colors represent longer average duration times.

The second output, Table V, contains the activity average durations that will be used to verify the results of the activity analysis conducted with the multimodal LLM. Meanwhile, in the DFG, the average duration values in the activity node label are already rounded.

TABLE V. THE ACTIVITY AVERAGE DURATION

Code	Activity	Average duration (Days)	Rounding (as in a DFG)
A	Announcement	3.97	4 days
B	Registration	3.97	4 days
C	Download Document	3.97	4 days
D	Explanation	0.047	67.9 mins
E	Bid Submission	3.79	3.8 days
F	Bid Opening	3.14	3.1 days
G	Administrative Evaluation	3.77	3.8 days
H	Technical Evaluation	3.77	3.8 days
I	Price Fairness Evaluation	3.77	3.8 days
J	Qualification Evaluation	3.79	3.8 days
K	Technical And Cost Clarification And Negotiation	0.85	20.4 hours
L	Winner Determination	0.57	13.6 hours
M	Winner Announcement	1.04	25 hours
N	Objection Period	3.96	4 days
O	Winner Designation	1.82	43.7 hours
P	Contract Signing	16	16 days

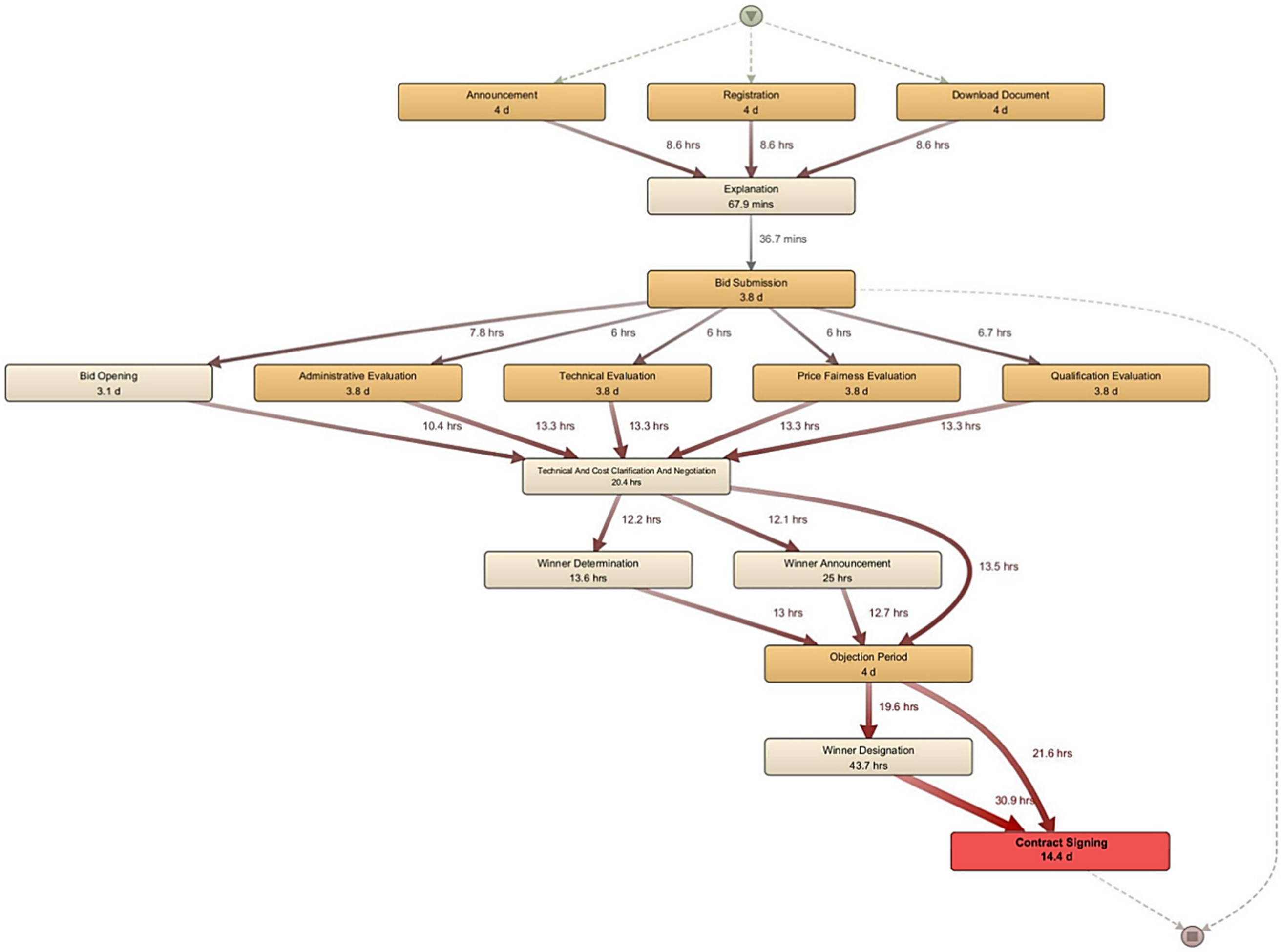

Fig. 3. DFG process model.

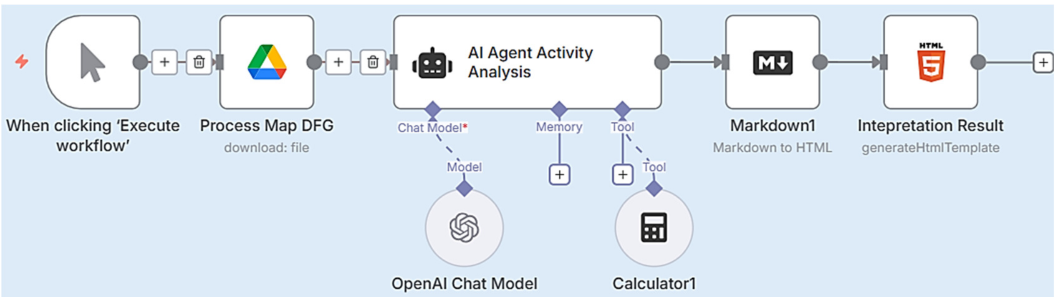

Fig. 4. AI agent visual design with n8n.

D. Analysis with LLM-Based AI Agent

To implement the LLM-based AI agent as an analytical model, this study used n8n as an orchestration layer, which is a low-code workflow automation platform that coordinates multi-step tasks, including reading process mining artifacts (e.g., DFG), composing structured prompts, invoking LLM APIs, and formatting the final analytical reports. Figure 4 shows the visual design of the analytical model on the n8n platform. This figure illustrates the structure of the n8n visual design that automatically converts process mining data into analytical model reports. The process model diagram input is managed by the Google Drive component as a file repository service in this analytical model. The AI agent acts as a model workflow manager, starting from reading input, processing it, and generating output. The AI agent is supported by several tools, including the LLM component, which is the ChatGPT 4.1 API, and a calculator component that assists the LLM in performing numerical calculations. The next components are responsible for formatting interpretation results: the Markdown component, which handles simple report formats, and the HTML component, which displays report results in a browser.

The analysis begins with the reading of the DFG process model, from the AI agent, to be sent and further processed in the LLM. The results of the interpretation analysis from the LLM will be received again by the AI agent and sent to the Markdown component, where they will be compiled into an activity analysis report. This report is then sent to the HTML component to be displayed in a browser. This analytical model setup can automatically analyze the performance of activities in the LLM-based tendering process, starting from processing inputs to creating organized and informative reports. The pre-defined prompt structure in the AI agent heavily influences the generation of a good report and sends it to the LLM. The predefined prompt structure adheres to the following pattern structure: <Role> <Context> <Directive> <Constraint> <Output format>. Tables VI and VII show the predefined prompts and the results of the activity performance analysis. Meanwhile, for the activity analysis result, data validation was conducted on the item Activity Duration Summary by comparing the average duration value with process statistics in Table V. On the other hand, expert validation was conducted for the items Activity Performance Summary Analysis, Bottleneck Interpretation, and Improvement Suggestions. Table VIII shows the results of the validation of the analytical model interpretation for activity analysis.

IV. DISCUSSION AND IMPLICATIONS

A. *Discussion of Results and Limitations*

Process mining does not provide an evaluative performance analysis directly. Instead, it outputs a process model and descriptive statistics that require further interpretation to generate process insights. In the proposed framework, process mining is utilized as an evidence generator, providing visual and numerical outputs. This output, a DFG process model, is then interpreted by an LLM-based AI agent, which transforms it into a structured narrative aligned with the performance analysis objectives.

The DFG is a visual artifact that contains activity labels and the average duration of each activity that is input to the agent. Using predefined structured prompts, the agent guided the LLM to accurately extract performance-related data and generate narrative descriptions. In this configuration, the extraction results achieved 100% accuracy based on manual cross-checking of process mining statistics (Table V). This demonstrates that the LLM's multimodal interpretation is reliable when supported by a well-defined workflow and precise prompts.

TABLE VI. PREDEFINED PROMPT

Predefined Prompt
<Role> You are a Process Mining Analyst specializing in interpreting business process models in Directly-Follows Graph (DFG) format. Your role is to analyze activity flows, detect bottlenecks, and generate a concise narrative describing activity performance within an e-tendering process. You act as a procurement expert that converts process-mining data into management insights. <Context> The input is a process model in DFG format derived from event log analysis of e-tendering activities. Each node in the DFG represents an activity box, and each edge represents a transition between activities. The activity box includes average duration values (in days or hours or minutes) for each activity indicating how long each step takes on average. The process typically consists of sequential activities such as registration, document upload, qualification, evaluation, negotiation, and contract award. You must interpret the DFG to identify where time delays accumulate and how they impact the overall process performance. <Directive> 1. Identify the activity average duration from DFG. 2. Provide a summary of activity performance analysis based on average activity durations. Classify all activities into three categories based on their average duration. 3. Provide an interpretation of the bottleneck analysis and activity flow. 4. Provide improvement suggestions. <Constraint> • Use only information visible or provided in the DFG (activity names, transitions, durations). • Do not assume external process knowledge beyond the tendering context. • Keep the explanation concise, factual, and aligned with process mining terminology. • Avoid speculative assumptions that are not supported by data. <Output> Produce the entire response in pure Markdown, structured as an analytical report for management titled "Activity Performance Analysis". Do not include any text outside the Markdown content (no prefaces, no follow-ups, no code fences). Use headings, tables and lists to highlight performance. Add short remarks explaining performance. 1. Activity Duration Tables 2. Summary of Activity Performance Analysis. Follow the formatting and content guidelines below: a. Content Format - Write the analysis in narrative paragraph form, with bullet points. Each paragraph must clearly describe one of the following categories: Long Duration Activities, Moderate Duration, Short Duration Activities- Within each paragraph: Mention the activity names and their average durations (in days or hours or minutes), - Explain their significance or impact on process flow and performance. Highlight bottlenecks using bold text for emphasis. b. Tone and Style - Write as if you are an Expert in Business Process Performance Analysis. - Avoid raw data tables or lists—focus on interpretive narrative.

TABLE VII. ACTIVITY PERFORMANCE ANALYSIS

1. Activity Duration Overview

Activity	Average Duration
Announcement	4 days
Registration	4 days
Download Document	4 days
Explanation	67.9 minutes
Bid Submission	3.8 days
Bid Opening	3.1 days
Administrative Evaluation	3.8 days
Technical Evaluation	3.8 days
Price Fairness Evaluation	3.8 days
Qualification Evaluation	3.8 days
Technical And Cost Clarification And Negotiation	20.4 hours
Winner Determination	13.6 hours
Winner Announcement	25 hours
Objection Period	4 days
Winner Designation	43.7 hours
Contract Signing	16 days

2. Summary of Activity Performance Analysis

- **Long Duration Activities:** The "Contract Signing" activity stands out with an average duration of 16 days, representing the longest single step and a critical component extending the process timeline. Other activities with notably long durations include "Announcement," "Registration," and "Download Document," each averaging 4 days, as well as the "Objection Period" taking 4 days. These extended durations form a foundation of delay early and late in the process cycle, with the contract signing phase acting as a substantial bottleneck that significantly impacts the total process throughput.

- **Moderate Duration Activities:** Several evaluation-related tasks—"Administrative Evaluation," "Technical Evaluation," "Price Fairness Evaluation," and "Qualification Evaluation"—each last approximately 3.8 days, indicating moderate but consistent processing time within the assessment phase. Similarly, "Bid Submission" (3.8 days) and "Bid Opening" (3.1 days) also fall into this category. The "Technical And Cost Clarification And Negotiation" activity averages about 20.4 hours, acting as a moderate but important transitional step that consolidates evaluation results. Together, these mid-range activities create a balanced workload, but they must align efficiently to prevent cascade delays.

- **Short Duration Activities:** The initial "Explanation" activity at around 68 minutes, along with "Winner Determination" (13.6 hours), "Winner Announcement" (25 hours), and "Winner Designation" (43.7 hours), are relatively shorter in duration compared to other stages. These activities illustrate focused decision and communication phases that, while shorter, are crucial for timely progression but do not appear to be major delays. They function efficiently within the broader timeline but should be maintained carefully to avoid slipping into bottlenecks.

3. Bottleneck and Activity Flow Interpretation

The process reveals clear bottlenecks primarily at the "Contract Signing" phase, which registers the longest duration by a wide margin, indicating a significant delay in concluding the procurement cycle after all evaluations and objections. The objection period also contributes to a cyclical feedback into contract signing, increasing delays. Furthermore, the evaluation phase composed of multiple parallel activities takes substantial time, slowing down progress to the negotiation stage. The flow suggests that even after bid submission and evaluation, the process could be expedited if bottlenecks in contract finalization and objection handling are addressed.

4. Improvement Suggestions

- Streamline and automate contract signing procedures to reduce the disproportionately high 16-day delay, potentially adopting digital signing solutions to speed up this final step.

- Optimize the objection period by setting stricter time caps or enabling parallel processing to prevent additional queuing.

- Enhance evaluation efficiency by cross-training evaluators or applying partial automation to reduce 3.8-day average durations across multiple evaluation steps.

- Monitor and potentially reduce waiting times during Announcement, Registration, and Download Document periods, possibly through better pre-communication and batch processing.

- Consider integrating some clarification and negotiation activities more tightly with winner determination to further reduce the overall elapsed time without compromising decision quality.

- This targeted approach, focusing on known bottlenecks and moderate delays, will greatly enhance the tendering process throughput and reduce time-to-contract.

TABLE VIII. VALIDATION AND EVALUATION

No	Interpretation item	Validation and Evaluation result
1	**Activity Duration Overview**	**Data Validation:** The reading of activity names and average activity durations from the process model image is 100% accurate after being compared to the manually calculated average activity duration data
2	**Summary of Activity Performance Analysis**	**Expert Evaluation:** The interpretation narrative of the average activity duration analysis based on the long, medium, and short activity duration groups produced is logical. In this case, the Contract Signing activity takes the longest time, but in reality, during the tendering process, the time-consuming activity is not the same for every company.
3	**Bottleneck and Activity Flow Interpretation**	**Expert Evaluation:** The results of the bottleneck analysis show logical reasoning and are consistent with the reality of the tendering process. The Contract Signing stage is accurately identified as a major bottleneck due to its administrative nature and the involvement of multiple parties, particularly in construction tendering. The Objection Period and Winner Designation are also understandable causes of delay because they are formal and subject to regulatory timelines. Parallel activities detected by the system, such as evaluation activities, are combined, where necessary, into one, thus facilitating monitoring of the tendering process. Overall, the reasoning provided is credible and accurately reflects the conditions based on the data.
4	**Improvement suggestions**	**Expert Evaluation:** The resulting process improvement recommendations not only identify the main issue (Contract Signing bottleneck) but also generate realistic and contextual solutions. The recommendations are implementable, ranging from contract automation, objection period review, and evaluation stage optimization to performance monitoring digitalization. Thus, the results of this LLM analytical model interpretation can be considered technically and strategically relevant and suitable as a basis for improvement considerations to enhance the efficiency of the e-tendering process.

The agent can generate report narratives that are consistent with the analysis objectives specified in the predefined prompts. In the summary of Activity Performance Analysis, the agent categorizes the average duration of activities into three classes (long, medium, and short) and assigns each one to the appropriate class based on its average duration. For example, Contract Signing (16 days) was classified as a long-duration activity and identified as the primary bottleneck, whereas Explanation (68 minutes) was classified as a short-duration activity. Expert reviewers confirmed that duration grouping and ordering are logical, while noting that dominant sources of delay can vary across organizations and operational settings.

Further interpretation about bottlenecks and activity flow revealed that Contract Signing consistently contributed to overall delays. Additional delays were observed in the objection period and during the evaluation stage. Parallel activity patterns were also identified within the evaluation process, covering administrative, technical, price fairness, and qualification checks. Based on these insights, the agent generated recommendations prioritizing improvements in Contract Signing, followed by objection handling and streamlining evaluation steps (approximately 3.8 days each). Experts assessed these recommendations as feasible and contextually relevant.

However, certain limitations must be acknowledged. First, the quality of the interpretation depends on the clarity of the discovered process model and the underlying event logs. Noisy logs, overlapping activities, or highly variable process paths can reduce both model accuracy and interpretability. A further limitation is the reliance on DFG, which may not fully capture concurrency and complex control-flow patterns, thereby constraining interpretation in highly variant processes. Second, although the LLM showed consistency under controlled prompting, it remains susceptible to prompt sensitivity and overconfidence; therefore, human oversight is necessary, especially for high-impact decisions. Third, the evaluation relied on expert judgment, which introduces subjectivity and limits reproducibility. Lastly, this study relies on a single organizational case, which limits its generalizability, necessitating a wider validation across diverse procurement contexts and data environments.

B. Theoretical Implications

From a theoretical standpoint, the augmented BPM pyramid situates process mining within the Descriptive Process Analytics layer [12]. Process mining outputs, such as DFG process models and activity statistics, serve as performance evidence that typically requires significant human interpretation before they can be transformed into decision-relevant insights. This study enhances this layer by formalizing the interpretive step as an explicit analytical function, operationalized through a controlled LLM-based agent. This approach improves the accessibility and consistency of descriptive insights without implying a transition to predictive or prescriptive autonomy. Methodologically, this study offers an evidence-grounded approach for integrating multimodal LLMs with process artifacts. It treats the DFG as a visual analytics artifact that encompasses both process structure and embedded activity performance information and employs prompt-constrained extraction to produce evidence-based performance narratives. Finally, the proposed architecture contributes to the discourse on AI-driven descriptive analytics by outlining a reusable integration pattern: from evidence generation (process mining) to evidence consumption (agent interpretation) and validation-supported narrative synthesis. This pattern illustrates how agentic interpretation can be introduced while ensuring that outputs remain anchored to auditable process evidence.

C. Practical and Managerial Implications

From a practical and managerial standpoint, the framework democratizes process mining insights by transforming DFG process models based on performance evidence into structured narratives that procurement stakeholders can directly act upon, thus diminishing reliance on process mining specialists for routine interpretation. This reduces time-to-insight by accelerating the transition from process mining outputs to decision-relevant findings based on performance evidence, including bottleneck identification and delay pattern summarization and observed process flow. Improving tendering process execution and reducing delay drivers can contribute to broader supply chain performance outcomes, as prior evidence links effective SCM practices, innovation, and TQM to competitive advantage and firm performance [25]. The workflow also supports data-driven process audits by providing a repeatable procedure to surface bottlenecks, document evidence-linked observations, and generate improvement suggestions grounded in observed durations and process flow rather than subjective judgment. Finally, this study offers an implementable template for deploying similar AI-agent systems, combining a reusable prompt structure with LLMs (e.g., ChatGPT) and an orchestration workflow (e.g., n8n) that can be adapted for procurement monitoring, compliance verification, and recurring performance review across BPM contexts. These implications apply to contexts where event logs and DFG-based process models are available and where stakeholders primarily require descriptive reporting rather than predictive or prescriptive automation.

V. CONCLUSION

This study addresses a practical gap in tendering performance analysis, namely, the limited support for automated interpretation of process-mining outputs for procurement stakeholders without process-mining expertise. The proposed end-to-end architecture integrates event-log-based process mining with an LLM-based agent and a controlled workflow that converts process mining evidence into decision-oriented narratives aligned with predefined analytical objectives. In the evaluated case, the agent extracted 16 activity labels and their average durations from the DFG with 100% accuracy when cross-checked against process-mining statistics, and expert review judged the resulting interpretations and improvement priorities to be logical, contextually relevant, and feasible for tendering use. Overall, the first objective is supported by the quantitative validation of the end-to-end architecture, and the second objective is supported by expert confirmation of the controlled, objective-aligned narratives for tendering use. These findings indicate that interpretation is a key barrier to practical adoption and that an evidence-grounded agent layer can reduce this burden while remaining within descriptive analytics. The results are bounded by log and process model quality, single-case validation, and the need for oversight given prompt sensitivity and overconfidence, as well as subjectivity in expert-based evaluation.

Future work should test the framework across non-construction organizations and different procurement regulations to assess generalizability. In addition, future work should reduce reliance on DFG-only representations by evaluating the interpretation workflow on richer process models, such as Petri nets, BPMN, or process trees, and by incorporating additional performance and conformance evidence beyond directly follow-up relations to improve interpretability in complex and highly variant procurement processes. Third, this approach should be extended to predictive process mining, while keeping the output traceable to process evidence. Fourth, multi-agent coordination should be explored for cross-process optimization across procurement value chains. Finally, implementation readiness should be strengthened by reducing hallucinations through evidence-grounded generation with automatic consistency checks and by reinforcing privacy and compliance through data minimization, pseudonymization of tendering and vendor identifiers, controlled access, and audit trails.

DATA AVAILABILITY

The dataset used in this study is available upon request from the corresponding author, subject to confidentiality restrictions and approval from the data owner.

REFERENCES

[1] N. Jahani, A. Sepehri, H. R. Vandchali, and E. B. Tirkolaee, "Application of Industry 4.0 in the Procurement Processes of Supply Chains: A Systematic Literature Review," *Sustainability*, vol. 13, no. 14, July 2021, Art. no. 7520, https://doi.org/10.3390/su13147520.

[2] N. Kumar and K. K. Ganguly, "Non-financial e-procurement performance measures: Their interdependence and impact on production cost," *International Journal of Productivity and Performance Management*, vol. 70, no. 1, pp. 41–64, Mar. 2020, https://doi.org/10.1108/IJPPM-07-2019-0353.

[3] T. Ramayah, M. H. Roy, K. B. Li, M. Jantan, I. Zbib, and Z. U. Ahmed, "Type of procurement and operational performance: comparing e-procurement and offline purchasing," *International Journal of Services and Operations Management*, vol. 3, no. 3, 2007, Art. no. 279, https://doi.org/10.1504/IJSOM.2007.013093.

[4] J. Hallikas, M. Immonen, and S. Brax, "Digitalizing procurement: the impact of data analytics on supply chain performance," *Supply Chain Management: An International Journal*, vol. 26, no. 5, pp. 629–646, July 2021, https://doi.org/10.1108/SCM-05-2020-0201.

[5] T. S. H. Teo and K. Lai, "Usage and Performance Impact of Electronic Procurement," *Journal of Business Logistics*, vol. 30, no. 2, pp. 125–139, Sept. 2009, https://doi.org/10.1002/j.2158-1592.2009.tb00115.x.

[6] W. H. Hung, C. P. Lin, Y. M. Tai, C. F. Ho, and J. J. Jou, "Exploring the impact of Web-based e-procurement on performance: organisational, interorganisational, and systems perspectives," *International Journal of Logistics Research and Applications*, vol. 17, no. 3, pp. 200–215, May 2014, https://doi.org/10.1080/13675567.2013.837431.

[7] W. Van Der Aalst, *Process Mining*. Springer, 2016.

[8] R. Nai, E. Sulis, R. Meo, F. Gorgerino, G. M. Racca, and L. Genga, "Process Mining on a Public Procurement Dataset: A Case Study," in *Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, vol. 2133, R. Meo and F. Silvestri, Eds. Springer Nature Switzerland, 2025, pp. 477–492.

[9] F. Hendrata, I. Vanany, P. Suwignjo, and N. Siswanto, "Process-Based Performance Analysis of Construction Tendering in Indonesia Using Real-World Event Logs," in *2025 International Electronics Symposium (IES)*, Aug. 2025, pp. 499–504, https://doi.org/10.1109/IES67184.2025.11162072.

[10] M. J. Sangil, "Heuristics-Based Process Mining on Extracted Philippine Public Procurement Event Logs," in *2020 7th International Conference on Behavioural and Social Computing (BESC)*, Nov. 2020, pp. 1–4, https://doi.org/10.1109/BESC51023.2020.9348306.

[11] N. Martin *et al.*, "Opportunities and Challenges for Process Mining in Organizations: Results of a Delphi Study," *Business & Information Systems Engineering*, vol. 63, no. 5, pp. 511–527, Oct. 2021, https://doi.org/10.1007/s12599-021-00720-0.

[12] D. Chapela-Campa and M. Dumas, "From process mining to augmented process execution," *Software and Systems Modeling*, vol. 22, no. 6, pp. 1977–1986, Dec. 2023, https://doi.org/10.1007/s10270-023-01132-2.

[13] Y. Yang, Z. Wu, Y. Chu, Z. Chen, Z. Xu, and Q. Wen, "Intelligent Cross-Organizational Process Mining: A Survey and New Perspectives." arXiv, 2024, https://doi.org/10.48550/ARXIV.2407.11280.

[14] L. Barbieri, E. Madeira, K. Stroeh, and W. Van Der Aalst, "A natural language querying interface for process mining," *Journal of Intelligent*

Information Systems, vol. 61, no. 1, pp. 113–142, Aug. 2023, https://doi.org/10.1007/s10844-022-00759-9.

[15] Y. Li, Z. Ni, and B. Xiao, "Domain Knowledge-Enhanced Process Mining for Anomaly Detection in Commercial Bank Business Processes," *Systems*, vol. 13, no. 7, July 2025, Art. no. 545, https://doi.org/10.3390/systems13070545.

[16] R. Nai, E. Sulis, D. Audrito, V. M. S. Trifiletti, R. Meo, and L. Genga, "Leveraging process mining and event log enrichment in European public procurement analysis: a case study," *Computer Law & Security Review*, vol. 57, July 2025, Art. no. 106144, https://doi.org/10.1016/j.clsr.2025.106144.

[17] A. Berti, D. Schuster, and W. M. P. van der Aalst, "Abstractions, Scenarios, and Prompt Definitions for Process Mining with LLMs: A Case Study." arXiv, 2023, https://doi.org/10.48550/ARXIV.2307.02194.

[18] P. Zhao, Z. Jin, and N. Cheng, "An In-depth Survey of Large Language Model-based Artificial Intelligence Agents." arXiv, Sept. 23, 2023, https://doi.org/10.48550/arXiv.2309.14365.

[19] X. Li, S. Wang, S. Zeng, Y. Wu, and Y. Yang, "A survey on LLM-based multi-agent systems: workflow, infrastructure, and challenges," *Vicinagearth*, vol. 1, no. 1, Oct. 2024, Art. no. 9, https://doi.org/10.1007/s44336-024-00009-2.

[20] L. Hughes *et al.*, "AI Agents and Agentic Systems: A Multi-Expert Analysis," *Journal of Computer Information Systems*, vol. 65, no. 4, pp. 489–517, July 2025, https://doi.org/10.1080/08874417.2025.2483832.

[21] A. K. Jadoon, C. Yu, and Y. Shi, "ContextMate: a context-aware smart agent for efficient data analysis," *CCF Transactions on Pervasive Computing and Interaction*, vol. 6, no. 3, pp. 199–227, Sept. 2024, https://doi.org/10.1007/s42486-023-00144-7.

[22] J. T. Tarigan, B. Wijaya, A. C. Salim, and S. M. Hardi, "An LLM-Based Behavior Agent with Natural Language Personality Control: Enabling Trait-Driven NPC Decision-Making through Prompt Engineering," *Engineering, Technology & Applied Science Research*, vol. 15, no. 5, pp. 26827–26832, Oct. 2025, https://doi.org/10.48084/etasr.12631.

[23] S. Yin *et al.*, "A survey on multimodal large language models," *National Science Review*, vol. 11, no. 12, Nov. 2024, Art. no. nwae403, https://doi.org/10.1093/nsr/nwae403.

[24] Y. Mao, J. He, and C. Chen, "From Prompts to Templates: A Systematic Prompt Template Analysis for Real-world LLMapps." arXiv, 2025, https://doi.org/10.48550/ARXIV.2504.02052.

[25] S. Sanaei, M. Manna, H. Bozorgkhou, S. Heidari, and E. Mehregan, "The Role of SCM practices inCompetitive Advantage and Firm Performance: A Mediating Role of Supply ChainInnovation and TQM," *Tehnički glasnik*, vol. 17, no. 4, pp. 516–523, Oct. 2023, https://doi.org/10.31803/tg-20221223200658.