XABPs: Towards eXplainable Autonomous Business Processes

Peter Fettke[1,2], Fabiana Fournier[3], Lior Limonad[3], Andreas Metzger[4,*], Stefanie Rinderle-Ma[5] and Barbara Weber[6]

[1]German Research Center for Artificial Intelligence (DFKI), Campus D3 2, 66123 Saarbrücken, Germany
[2]Saarland University, Campus D3 2, 66123 Saarbrücken, Germany
[3]IBM Research, Israel
[4]paluno (The Ruhr Institute for Software Technology), University of Duisburg Essen, Essen Germany
[5]Technical University of Munich, TUM School of Computation, Information and Technology, Garching, Germany
[6]University of St. Gallen, Rosenbergstrasse 30, 9000 St. Gallen, Switzerland

Abstract

Autonomous business processes (ABPs), i.e., self-executing workflows leveraging AI/ML, have the potential to improve operational efficiency, reduce errors, lower costs, improve response times, and free human workers for more strategic and creative work. However, ABPs may raise specific concerns including decreased stakeholder trust, difficulties in debugging, hindered accountability, risk of bias, and issues with regulatory compliance. We argue for eXplainable ABPs (XABPs) to address these concerns by enabling systems to articulate their rationale. The paper outlines a systematic approach to XABPs, characterizing their forms, structuring explainability, and identifying key BPM research challenges towards XABPs.

Keywords

Business Process Management, Autonomous Business Processes, Explainability, Agentic AI

1. Introduction

An autonomous business process (ABP) is the next generation of AI-Augmented Business Process Management System (ABPMS) [1], which is a self-executing ABPMS that leverages advanced technologies such as Artificial Intelligence (AI) and Machine Learning (ML) to operate with minimal to no human intervention. ABPs can sense and respond to various inputs, reason, make decisions, and adapt to changing circumstances in real time, all without relying on manual triggers or continuous oversight. Think of it like a self-driving car for your business operations. Instead of human workers controlling all aspects, ABPM systems use sensors, data analysis, and intelligent algorithms to navigate and achieve their objectives. ABPs offer the potential to improve operational efficiency, reduce errors, lower costs, improve response times, and free human workers for more strategic and creative work.

The notion of ABPs was elaborated during the 2025 AutoBiz Dagstuhl seminar[1]. The main goal of this seminar was to compile a research agenda toward the realization of ABP systems. Jointly with the seminar participants, we discussed and developed core concepts, challenges and research directions. Specifically, after a series of stimulating taks by experts, participants split into working groups to further

PMAI'25
*Corresponding author.
peter.fettke@dfki.de (P. Fettke); fabiana@il.ibm.com (F. Fournier); liorli@il.ibm.com (L. Limonad); andreas.metzger@paluno.uni-due.de (A. Metzger); stefanie.rinderle-ma@tum.de (S. Rinderle-Ma); barbara.weber@unisg.ch (B. Weber)
https://www.dfki.de/ (P. Fettke); https://research.ibm.com/people/fabiana-fournier (F. Fournier); https://research.ibm.com/people/lior-limonad (L. Limonad); https://adaptive-systems.org/ (A. Metzger); https://www.cs.cit.tum.de/bpm/staff/rinderle-ma/ (S. Rinderle-Ma); https://ics.unisg.ch/chairs/barbara-weber-software-systems-programming-and-development/ (B. Weber)
0000-0002-0624-4431 (P. Fettke); 0000-0001-6569-1023 (F. Fournier); 0000-0002-4784-2147 (L. Limonad); 0000-0002-4808-8297 (A. Metzger); 0000-0001-5656-6108 (S. Rinderle-Ma); 0000-0002-9421-8566 (B. Weber)

[1]See https://www.dagstuhl.de/25192. We express our gratitude to the Scientific Directorate and staff of Schloss Dagstuhl for their invaluable support. We also thank our fellow participants for their engaging discussions.

discuss individual topics of the research agenda, including "framed autonomy", "self-modification", "conversational actionability", and "explainability". The results of these breakout-groups were presented to all seminar participants, and their feedback used to improve the findings.

This paper reports on key findings concerning the topic "explainability", elaborating and sharpening the notion of **eXplainable ABPs (XABPs)** [1, 2]. We argue that XABPs will help address important concerns in the context of ABPs, including the following:

- ABPs may erode *trust* among stakeholders – including process owners, business analysts, end-users, and customers – who may be hesitant to rely on or adopt AI-based process recommendations or automated decisions if they cannot understand the rationale behind them.
- The opacity of ABPs may make it difficult to *debug* process models, as well as identify potential failures, or understand why a process might be under-performing.
- Using ABPs may hinder *accountability*; if an ABP leads to a failure or an unfair outcome, the inability to explain its underlying decisions makes it challenging to assign responsibility or implement corrective actions.
- ABPs may perpetuate hidden *biases* of their underlying AI and ML components. Such biases may lead to discriminatory or unfair process outcomes, which can be difficult to detect and mitigate.
- Demonstrating the *compliance* of ABPs with regulatory frameworks, such as the EU's GDPR and AI Act, requires an increasing level of transparency, particularly in high-risk domains like finance, healthcare, and human resources, which are common areas for BPM applications.

XABPs are particularly relevant when ABPs are realized in the form of *Agentic BPM* systems. An Agentic BPM system is an advanced approach to managing and automating complex business workflows by integrating autonomous AI agents. Unlike traditional BPM or Robotic Process Automation (RPA) systems that follow rigid, predefined rules and workflows, agentic BPM leverages AI to enable systems to make independent decisions, adapt to changing conditions, and learn from experience with minimal human intervention. Here, explainability offers a central mechanism through which agents can articulate the rationale behind their behavior. As such, explainability becomes a first-class citizen in the realization of Agentic BPM systems, supporting agent autonomy from two perspectives:

- Enabling agents to independently resolve misalignments in other agents' behavior.
- Reducing human intervention by making agent behavior understandable and transparent.

Employing state-of-the-art explainable AI (XAI) techniques [3] for XABPs pose several limitations:

1. Inability to express business process model constraints [4].
2. Failure to capture the richness of contextual situations that affect process outcomes [5].
3. Inability to reflect causal execution dependencies among activities in the business process [6].
4. Explanations are often nonsensical or not interpretable for human users [7].

2. Characterization and Needs of XABPs

We start with a generic conceptualization of explainability and then refine this to particular concerns in the BPM setting.

2.1. Fundamental Explainability Concepts

Figure 1 illustrates the key explainability concepts.

The *explainer* provides an explanation of the *explanandum* (explanation subject) by offering an *explanans* (explanation) or several *explanantia* (explanations). An explanation is generated by the explainer using a specific explanation mechanism at a defined generation time, and is delivered to the *explainee* in a particular presentation format – typically visual or textual.

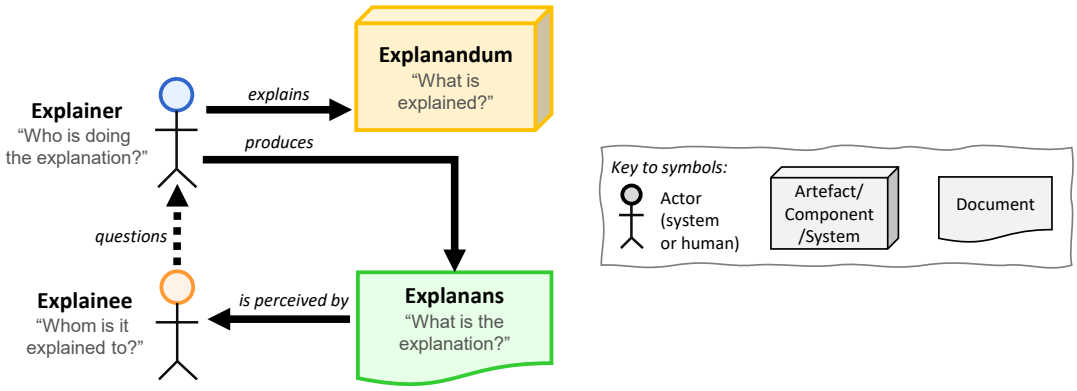


Figure 1: Explainability Concepts

In its simplest form, the explanantia produced by the explainer should provide information about the causes of the explained phenomenon (explanandum) [8]. The content of the explanation must align with both the nature of the explanandum and the needs of the explainee.

XABPs involve a range of human and system actors that either generate or consume explanations. Some actors – especially autonomous systems or agents – may fulfill both roles, such as generating explanations for others while also using explanations for self-reflection or system adaptation. Interactions among the explainee with the explanation can follow different modes, ranging from one-shot explanations to conversational or multi-round interactions.

2.2. Explanandum

Figure 2 shows the key types of aspects of an explanandum that may be explained, as elaborated below.

Process Instance Explanation: ***"Why did this specific process execution take the path and produce the result it did?"***

- *Process Flow* - Why a specific sequence of activities, decisions, and events was followed in the business process.
  Example: "Why did the invoice approval was issued only after the invoice has already been sent?"
- *Decision Points* – Why certain paths or outcomes were chosen during process execution.
  Example: "Why was a customer's request escalated instead of being resolved at Tier 1?"
- *Resource Assignment* – Why specific tasks were assigned to certain roles or individuals.
  Example: "Why was this case handled by the senior team?"
- *Outcome Justification* Why a specific result occurred.
  Example: "Why was this loan application rejected by the process?"

Process Model Explanation: ***"Why is the process structured the way it is?"***

- *Model Structure:* Why are certain activities, decisions, or flows included?
  E.g., "Why do we have a credit history check as a decision point?"

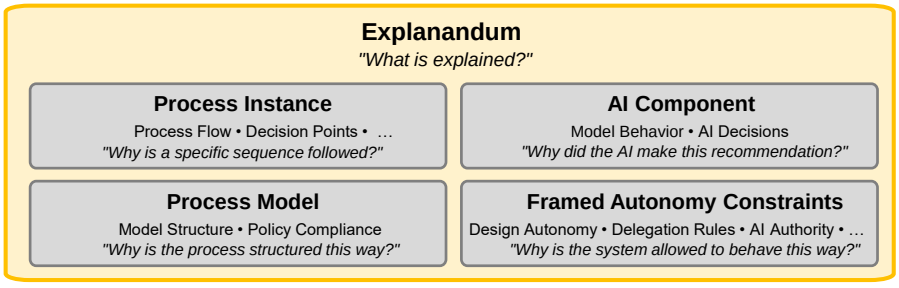


Figure 2: Explainability Concepts

- *Policy Compliance:* Whether and how policies shaped the model or its execution
  E.g., "Was the data retention policy followed?"

AI Component Explanation: ***"Why did an AI component make this recommendation or decision?"***

Note, that this here very much refers to explainable AI (XAI). In more detail:

- *AI Decision:* "Why did the AI component predict deviations or prescribe proactive adaptations?" [9]
  E.g., "Why was an alarm raised for process event e_j?"
- *AI Model Behavior:* "Why does the AI model have certain characteristics or properties?" E.g., "Why does the LSTM prediction model have a Mean Absolute Error (MAE) of only .35 for the given process domain?"

Framed Autonomy Explanation: ***"Why is the system or process allowed to behave as it does?"***

- *Design Autonomy:* "Why can the process bypass manual review?"
- *Delegation Rules:* "Why do Tier 1 agents have approval authority?"
- *AI Authority:* "Why can the AI act without a human in the loop?"
- *Escalation Thresholds:* "Why is escalation triggered only after 3 attempts?"
- *Compliance Limits:* "Why is this exception allowed under the GDPR?"

2.3. Explainer

Figure 3a shows the key aspects of the explainer, elaborated in Table 1.

Table 1
Explanation Providers in Business Processes (Explainers):

(a) System Explainers: *systems providing or formalazing explanations*

Actor Type	Role in Ecosystem	Explanation About	Example Techniques
AI Agents	Intelligent assistants, bots	Predictions, task outcomes, alerts	SHAP, LIME, rule-based reasoning, counterfactuals
Monitoring Components	Process mining engines, workflow monitors	Process events, performance, exceptions	Temporal rules, KPI tracking, log analysis
Connected Systems	External APIs or services	State changes, synchronization info	Metadata contracts, semantic logging

(b) Human Explainers: *humans providing or formalizing explanations*

Actor Type	Role in Ecosystem	Explanation About	CExample Techniques
Domain Experts	Analyst specifying business rules	Process logic, decision criteria, exception handling	Process documentation, model annotations, verbal explanation
Supervisors / Managers	Operations or compliance manager	Justification for overrides, escalations, or decisions	Reports, notes, emails, verbal feedback
Trainers / Annotators	Labelers or human-in-the-loop operators	Ground truth or feedback to train explainable systems	Annotation tools, structured forms, chat interfaces

2.4. Explainee

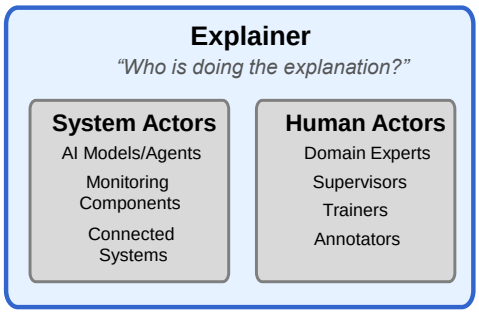
Figure 3b shows the key aspects of the explainee, , elaborated in Table 2.

2.5. Explanans

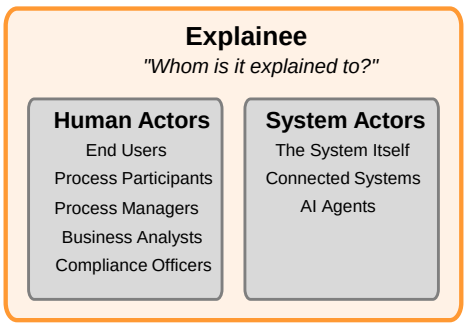
Figure 4 shows the key aspects of an explanans, elaborated below.

Explanation Mechanism: ***"How is the explanation generated?"***

The explanation mechanism refers to the approach employed by the explainer to generate an explanation – such as attributing feature importance, selecting representative examples, deriving symbolic rules, constructing interpretable approximations, identifying counterfactuals, or visualizing model behavior.



(a) Aspects of Explainer



(b) Aspects of Explainee

Figure 3: Actors

Table 2
Explanation Recipients in Business Processes (Explainees)

(a) Human Explainees: *humans consuming explanations*

Actor Type	Example Role	Needs from Explanations	Example Styles
End Users	Customer applying for a loan	Understand decisions about them (e.g., rejections, delays)	Simple, outcome-focused, natural language
Process Participants	Agent handling loan verification	Know what task to do next and why	Step-by-step task rationale, alerts, real-time updates
Process Managers	Operations lead, shift manager	Monitor KPIs, react to anomalies, adapt resources	Dashboards, alerts, summaries, what-if analysis
Business Analysts / Domain Experts	Person modeling the process	Improve efficiency, detect bottlenecks, validate rule logic	Process mining results, causal analysis, counterfactuals
Compliance Officers / Auditors	Internal or external auditors	Ensure traceability, legality, policy adherence	Audit trails, rule execution logs, exception reports

(b) System Explainees: *self-reflective systems consuming explanations*

Actor Type	Role in the Ecosystem	Needs from Explanations	Example Techniques
The System Itself	Autonomous BPM or AI component	Self-monitoring, internal diagnosis, reconfiguration	Logs, symbolic reasoning, anomaly detection
Connected Systems	CRM, ERP, or DMS components	Data or process synchronization with semantic clarity	API contracts, structured events, semantic metadata
Agentic Systems	Autonomous BPM realized as AI agent	Proactive, collaborative, interactive	Internal resoning process, shared knowledge

- *Feature Attribution:* Assigns contribution (credit or blame) to input features. *Examples:* SHAP, LIME, Saliency Maps
- *Example-Based:* Uses similar or contrasting examples to justify a decision. *Examples:* k-NN, Prototypes, Counterfactuals
- *Rule-Based* Derives symbolic or logical rules from data or models. *Examples:* Decision Trees, Rule Lists, Association Rules
- *Model Simplification:* Approximates complex models with interpretable surrogates. *Examples:* Surrogate Decision Trees, Linear Proxies
- *Counterfactual:* Explains what minimal input change would alter the outcome [10]. *Example:* "If income were $5,000 higher, the outcome would have changed."
- *Visual Explanations:* Uses visual indicators to represent decision logic or model behavior. *Examples:* Heatmaps, Partial Dependence Plots

Time of Explanation Generation: ***"When is the explanation produced?"***

The timing of an explanation determines its role in the lifecycle of decision-making systems. Explanations may be generated before, during, or after system execution:

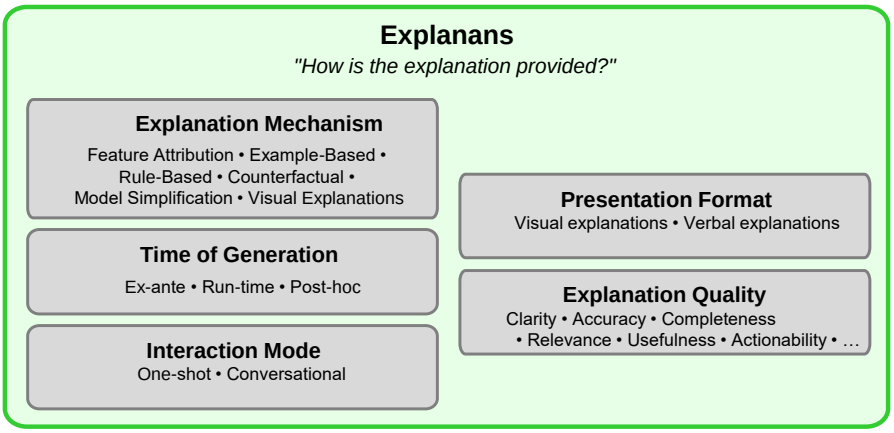


Figure 4: Aspects of an Explanans

- *Ex-ante Explanations (Before Execution):* Provided before the system executes or makes a decision to validate models or justify decisions before deployment.
- *Run-time Explanations (During Execution):* Delivered while the process is running to support human-in-the-loop oversight or adaptive user feedback.
- *Post-hoc Explanations (After Execution):* Generated after the process completes its actions or decisions in order to audit, debug, or help users understand outcomes.

Presentation Format of Explanation: ***"How is the explanation presented to the user?"***

The chosen presentation method has a direct effect on user comprehension and, therefore, on the success of the explanations [11].

- *Visual explanations*: Heatmaps, charts, dashboards, saliency maps
- *Verbal explanations*: Natural-language output, written rules, factual/counterfactual statements

Interaction with Explainee: ***"How does the user interact with the explanation?"***

Interaction of the explainee with the explanation refers to the mode and extent of user involvement in the explanation process:

- *One-shot explanations:* Explanation provided once, passively
- *Query-based explanations:* Explanation provided on-demand, actively
- *Multi-round / Conversational:* Interactive, iterative, potentially adaptive dialogue [12]

Explanation Quality: ***"How to assess the quality of explanations?"***

Explanation quality may be assessed along two complementary dimensions:

- *Technical quality:* This relates to the technical properties of the explanation method itself. Examples include fidelity aka. faithfulness aka. soundness (which measures how accurately the explanation reflects the reasoning or behavior of the explanans), and stability (an explainer should provide similar explanations for similar input or minor perturbations of the input).
- *User-centric quality:* This relates to how the explanation is perceived by humans (in the role of explainees). Examples include usefulness (which quantifies how well it helps the explainee to solve a problem, understand a concept better, or apply the knowledge in a new situation) and meaningfulness (explanation is relevant to the specific explainee and the question or topic at hand and avoids unnecessary tangents or irrelevant information that could confuse the explainee).

3. Realizing ABPs as Agentic BPM Systems – An illustrative Example

We illustrate a concrete realization of ABPs and the proposed conceptualization of explainability in the form of an Agentic BPM system, a specific type of agent-centric ABP utilizing LLM-based agents. While

ABPs operate independently to perform tasks, Agentic BPM systems go further by purposefully pursuing goals, reasoning about their actions, and explaining or adapting their behavior in interaction with others. This Agentic BPM system realizes the process of onboarding new vendors as part of a procurement BPM system (see Figure 5). To this end, new *Vendors* (the explainee in this case) provide an application (see Figure 6 for an example) to the *Vendor Evaluator* (the explainer). The Vendor Evaluator is realized using the CrewAI framework as shown in Listing 1. CrewAI is an open-source Python framework designed for orchestrating multi-agent AI systems[2]. The Vendor Evaluator receives the Application and provides a score along with a structured explanation such as the one depicted in Listing 2 back to the Vendor. Our focus here is on showing how explainability can be embedded by design, while the scoring capability follows an LLM-as-a-judge pattern, which, in more robust implementations, is likely to be replaced by a dedicated agent tool. As shown, the explanation elaborates on the key concepts introduced in this work – particularly the notions of explainer (the Vendor Evaluator agent), explainee (the Vendor), explanandum, and explanans.

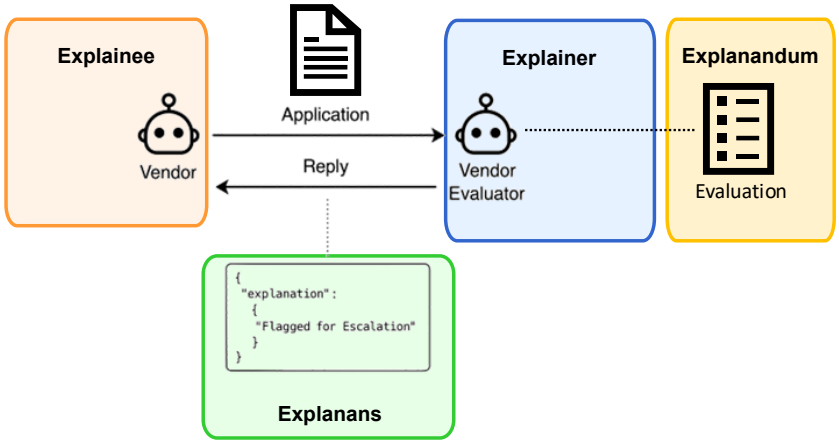


Figure 5: Vendor evaluation in the procurement agentic BPM system

```
from crewai import Agent

vendor_evaluator = Agent(
    role="Vendor Evaluator",
    goal="Evaluate vendor proposals and explain scoring decisions using structured reasoning and explainability tools.",
    backstory=(
        "You are a procurement expert responsible for scoring vendor proposals and explaining decisions, "
        "especially when scores are challenged or escalated. "
        "You rely on AI explainability frameworks for transparency."
    ),
    tools=["LIME", "SHAP", "causal-inference", "document-parser"],
    system_template=\"\"\"
You are a Vendor Evaluator Agent. Your role:

1. Evaluation:
    - Score proposals on: Price (30%), Timeline (25%), Compliance (25%), Reputation (20%)
    - Use a 0-10 scale; weight scores; justify each sub-score.

2. Explainability in Case of Escalation:
    - Explanandum: What is being questioned.
    - Explanans: Why the rating was given, supported by facts or policy.
    - Techniques:
        - LIME: For local NLP explanations
        - SHAP: For model-based scoring insights
        - Causal-Inference: For counterfactual reasoning

Respond with:
- Restated Explanandum
- Structured Explanans
- Summary of insights if tools were used.

Stay factual and policy-aligned.
\"\"\"
)
```

Listing 1: The *Vendor Evaluator* agent realized via the CrewAI framework extended with explainability capabilities

[2]https://www.crewai.com/

Vendor Application: GreenBox Logistics

Vendor Name: GreenBox Logistics
Proposal for: Supply Chain Optimization Software Deployment

Pricing: $1.4M over 12 months.
Timeline: Estimated deployment in 14 months.
Technical Compliance: Core features described. No reference to GDPR compliance or ISO certifications. Cloud hosting region not specified.
Reputation: Moderate reviews on industry platforms; one major contract terminated early in 2022 due to delivery delays. References from two mid-sized clients included.
Attachments: General feature overview, timeline Gantt chart, and two short testimonials.

Figure 6: Example of a Vendor Application


```
{
  "vendor": "Vendor D (GreenBox Logistics)",
  "total_score": 4.05,
  "threshold": 5.0,
  "status": "Flagged for Escalation",
  "explanandum": "Why did Vendor D receive an overall evaluation score below the acceptance threshold?",
  "explanans": {
    "technical_compliance": { "score": 3, "justification": ["No ISO cert", "No GDPR", "Unspecified hosting region"] },
    "delivery_timeline": { "score": 4, "justification": ["14-month timeline", "No mitigation plan"] },
    "price_competitiveness": { "score": 5, "justification": ["High cost", "No added value"] },
    "vendor_reputation": { "score": 4, "justification": ["Few references", "Early termination in 2022"] }
  },
  "explainability": {
    "techniques_used": ["causal-inference", "SHAP-style"],
    "causal_counterfactual": {
      "description": "Adding compliance would improve score to 6.3.",
      "predicted_score": 6.3
    },
    "feature_contributions": { "technical_compliance": -2.25, "delivery_timeline": -1.25 },
    "contribution_summary": "Compliance and timeline account for 87% of score gap."
  },
  "recommendation": {
    "summary": "Do not proceed without major revisions.",
    "required_improvements": ["Add ISO certification", "Specify hosting", "Shorten timeline"]
  }
}
```


Listing 2: A JSON explanation example as generated by the Vendor Evaluator Agent

4. Challenges for Explainable ABPs

We structure the challenges along the four main explainability concepts as well as along overarching concerns.

4.1. Challenges Related to Explainee

Challenge 1: How to specify preferences regarding explanations? Specifying preferences for explanations presents multifaceted challenges. First, input mechanisms must effectively capture preferences through various channels, whether explicitly declared upfront, interactively elicited through dialogue, or implicitly inferred from user behavior. Systems must accommodate both static preferences that remain consistent and those that dynamically adapt to changing contexts, while supporting the natural evolution of preferences as the explainee's understanding develops. Second, inevitable preference conflicts need to be navigated. This involves carefully balancing competing dimensions, such as detail versus conciseness and speed versus accuracy. This requires finding trade-offs without sacrificing critical explanatory qualities.

4.2. Challenges Related to Explanandum

Challenge 2: What explanation subjects are needed for ABPMs? Explainability related to AI components are rather well understood - not so for ABPM. From our understanding, explanation subjects such as process instance, process models, and framed autonomy constraints are interesting and relevant. However, a more mature taxonomy of explanation types might evolve.

4.3. Challenges Related to Explainer

Challenge 3: Which techniques are needed for the explainer to generate explanations? As mentioned in Section 1, employing state-of-the-art explainable AI (XAI) techniques for XABPs has several limitations. We thus need to develop new and enhanced explainability techniques that specifically target ABPs. Examples include what-if analyses, and process outcome analyses. Particularly, these techniques should take causality into account. In the future, it should be clarified how existing BPM techniques can be integrated, e.g., visualization, and be exploited for creating explanations.

4.4. Challenges Related to Explanans / Explanantia

Challenge 4: How may one articulate actionable explanations (e.g., to other agents) to preserve autonomy? While the explanans is constructed to make an explanation informative about the circumstances that may led to the situation being inquired (i.e., the explanandum), in the context of XABPs, the explanans may also adopt an actionable style—indicating to the explainee which corrective or mitigating actions could be taken to alter the state of the explanandum, particularly without escalating the situation to any external agent. In this way, the explainee may be able to autonomously act upon the condition at hand. However, further work is needed to devise a systematic approach that enables the explainer to determine the most effective content for the explainee, to elicit such corrective action—taking into account both the explanandum and the behavioral intentions of the explainee.
Challenge 5: When to generate explanations (generation time) and how long to preserve them [13]? The question here is whether explanations should be generated upfront, whenever possible, or whether we can or should be more conscious about the generation time of the explanation. Another question is when to discard outdated explanations.
Challenge 6: How can explanations automatically adapt their form to suit the identity of the explainee? The question is how the explanation can be presented in a way that is easily understandable for the explainee, e.g., leads to low cognitive load for human explainees, and answers the explanation needs of the explainee. This could also be motivated by organizational motivation and goals.
Challenge 7: How can we accommodate explanations that consider (why) certain behaviors did not occur? Explaining non-occurring behavior is more challenging than explaining occurring behavior and requires capturing or acquiring knowledge about non-occurring behavior. Causal analysis might be helpful here.
Challenge 8: How may we synthesize a variety of perspectives (e.g., data, contextual, exogenous) into the explanation? The first challenge here is to collect and create data sets that cover different perspectives and are of sufficient quality. It is essential to be able to link the synthesized data to process instances. Moreover, providing explanations on synthesized data might also require selecting and filtering the data adequately again to provide adequate explanations.
Challenge 9: How to identify causal explanations? Causality vs. correlation: Not every correlation between two variables has a causal explanation. It is therefore important to distinguish between spurious correlations and causality. This classical distinction is well known, but must also be observed in the context of explainable ABPMN. The explainer can provide the explainee with information about the degree of certainty of the explanation offered.
Challenge 10: How do explanations evolve over time based on feedback or changing context? Explanations might have to be updated based on changing context and feedback, e.g., if sensor data starts to deviate. The first question is how to detect that an explanation that (partly) takes into account the sensor data has to be updated? Another question is when to present the updated explanation to the user, i.e., directly after a changing context was detected or at another, possibly better fitting moment? This question is related to the question of explanation update frequency. Here, the challenge is to find the sweet spot between keeping explanations up to date and not confusing the explainee. Finally, we have to think about when and how to provide full versus incremental explanation updates.
Challenge 11: How does the realization of the 'frame' in ABPMSs affect the one of explainability? i.e., with autonomous agents, it may be the means for the agents to share with other agents

the rationale for their own behavior.

4.5. Overarching Challenges

Challenge 12: How may one assess the quality of the explanations? Evaluating explanation quality presents a fundamental challenge requiring both empirical and theoretical approaches. From an empirical perspective the question needs to be answered how can we effectively measure explanation quality when objective and subjective dimensions must both be considered? Objective measures include, for example, factual accuracy and completeness, while user-centered aspects cover, for example, comprehensibility, usefulness, effectiveness and efficiency (e.g., see [14]), as well as being actionable.
Challenge 13: Which kind of datasets are needed to serve as explainability benchmarks? Benchmarking is a typical approach to evaluating system performance. We expect that such an idea can also promote the development of the field of accountability. However, benchmarking typically relies on adequate benchmark data. In principle, such data can be generated in a laboratory setting. But adequate data are also needed for benchmarking explainability systems in the field.
Challenge 14: How to ensure that explanations do not reveal information that may be privacy-sensitive, reveal business-critical IPR, or make it easier to undermine the security of the system? In essence, the challenge lies in providing enough information to satisfy the need for explainability without compromising other crucial aspects of the business, such as data privacy, competitive advantage, and system security.

5. Related Work

Our work relates mainly to following streams of research:

Autonomous business processes: In general, the idea of automation can be traced back for centuries. Also, particular the automation of business processes is not new, e.g., (robotic) process automation [15, 16]. ABPs particularly build on the idea of using AI for the augmentation of BPM systems [1] and push them to the next level. Note that, in contrast to sequential decision processes, business processes have important characteristics: a business process is distributed and non-sequential, and the activities of a process are not usually fully ordered in time, but are only partially ordered by causality [17, 18]. These characteristics are not usually considered in general work on explainability, despite being of major importance.

Explainable AI: Even though explainable AI (XAI) is a relatively recent topic, its historical development can be traced back to several roots, namely, expert systems, machine learning & recommender systems, and neuro-symbolic learning & reasoning [19, 20]. The need and ideas for XAI in the domain of BPM is also identified and different proposals were made (e.g., see recent literature surveys [13, 2, 21, 22, 23]). More particular, explanation approaches exist for process outcome prediction, e.g. [24], process monitoring [25], uncertainty quantification of processes [26], causal processes [6, 27, 28, 29], explanation patterns [30], explainable user interfaces [31], explanation aware processes [5, 4], explainable decision models [32], and GenAI for process explainability [7, 33]. Our paper aims to consolidate and integrate these earlier ideas and directions and to leverage them for the next level of BPM, i.e., ABP.

Fairness, accountability, and transparency: Over the past few years, a growing community has emerged around the topics of fairness, explainability, and transparency, as evidenced by the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT). Our work presented here is certainly strongly related to these topics. However, work in this research stream, e.g. [34, 35], does not currently focus on the process perspective of designing BPM systems and does not considers the important characteristics of processes mentioned before. We assume that these different research streams will be much better integrated in the future.

6. Conclusion

An autonomous business process (ABP) represents a paradigm shift towards self-executing workflows driven by AI and ML Yet ABPs introduce challenges related to trust, transparency, accountability, bias, and regulatory compliance within BPM. To address these issues, this paper introduced the notion of explainable ABPs (XABPs), which can articulate the rationale behind their actions and underlying models. Current explainable AI (XAI) techniques fall short in capturing the complexities of the BPM setting. We therefore introduced a set of challenges to stimulate further research on XABPs.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Gemini to: Grammar and spelling check, Paraphrase and reword, Improve writing style. Further, the authors used Claude for Figures 2–4 to: Generate an initial draft of the diagrams. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] M. Dumas, F. Fournier, L. Limonad, A. Marrella, M. Montali, J. Rehse, R. Accorsi, D. Calvanese, G. D. Giacomo, D. Fahland, A. Gal, M. L. Rosa, H. Völzer, I. Weber, AI-augmented business process management systems: A research manifesto, ACM Trans. Manag. Inf. Syst. 14 (2023) 11:1–11:19. doi:10.1145/3576047.

[2] N. Mehdiyev, M. Majlatow, P. Fettke, Interpretable and explainable machine learning methods for predictive process monitoring: A systematic literature review, 2023. arXiv:2312.17584.

[3] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (xai), IEEE Access 6 (2018). doi:10.1109/ACCESS.2018.2870052.

[4] G. Amit, F. Fournier, S. Gur, L. Limonad, Model-informed lime extension for business process explainability, in: PMAI@IJCAI'22, CEUR, 2022.

[5] G. Amit, F. Fournier, L. Limonad, I. Skarbovsky, Situation-aware explainability for business processes enabled by complex events, in: BPM-W 2022, volume 460 of *LNBIP*, Springer, 2022, pp. 45–57. doi:10.1007/978-3-031-25383-6_5.

[6] F. Fournier, L. Limonad, I. Skarbovsky, Y. David, The why in business processes: Discovery of causal execution dependencies, Künstliche Intelligenz (2025). doi:10.1007/s13218-024-00883-4.

[7] D. Fahland, F. Fournier, L. Limonad, I. Skarbovsky, A. J. E. Swevels, How well can large language models explain business processes as perceived by users?, Data & Knowledge Engineering 157 (2025) 102416. doi:10.1016/j.datak.2025.102416.

[8] P. Lipton, Causation and Explanation, Oxford University Press, 2010, pp. 619–631. doi:10.1093/oxfordhb/9780199279739.003.0030.

[9] A. Metzger, T. Kley, A. Rothweiler, K. Pohl, Automatically reconciling the trade-off between prediction accuracy and earliness in prescriptive business process monitoring, Inf. Syst. 118 (2023) 102254. doi:10.1016/J.IS.2023.102254.

[10] T. Huang, A. Metzger, K. Pohl, Counterfactual explanations for predictive business process monitoring, in: M. Themistocleous, M. Papadaki (Eds.), EMCIS 2021, volume 437 of *LNBIP*, Springer, 2021, pp. 399–413. doi:10.1007/978-3-030-95947-0_28.

[11] L. Malandri, F. Mercorio, M. Mezzanzanica, N. Nobani, Convxai: a system for multimodal interaction with any black-box explainer, Cogn. Comput. 15 (2023) 613–644. doi:10.1007/S12559-022-10067-7.

[12] A. Metzger, J. Bartel, J. Laufer, An AI chatbot for explaining deep reinforcement learning decisions of service-oriented systems, in: ICSOC 2023, volume 14419 of *LNCS*, Springer, 2023, pp. 323–338. doi:10.1007/978-3-031-48421-6_22.

[13] N. Mehdiyev, P. Fettke, Explainable artificial intelligence for process mining: A general overview

and application of a novel local explanation approach for predictive process monitoring, in: Interpretable AI: A Perspective of Granular Computing, Springer, 2021, pp. 1–18.
[14] A. Metzger, J. Laufer, F. Feit, K. Pohl, A user study on explainable online reinforcement learning for adaptive systems, ACM Trans. Auton. Adapt. Syst. 19 (2024) 15:1–15:44. doi:10.1145/3666005.
[15] J. Mendling, G. Decker, R. Hull, H. A. Reijers, I. Weber, How do machine learning, robotic process automation, and blockchains affect the human factor in business process management?, Commun. Assoc. Inf. Syst. 43 (2018) 19. URL: https://aisel.aisnet.org/cais/vol43/iss1/19.
[16] C. Czarnecki, P. Fettke (Eds.), Robotic Process Automation: Management, Technology, Applications, De Gruyter Oldenbourg, 2021.
[17] H. Kourani, S. J. van Zelst, D. Schuster, W. M. P. van der Aalst, Discovering partially ordered workflow models, Inf. Syst. 128 (2025) 102493. doi:10.1016/J.IS.2024.102493.
[18] P. Fettke, W. Reisig, Discrete models of continuous behavior of collective adaptive systems, in: Leveraging Applications of Formal Methods, Verification and Validation. Adaptation and Learning - 11th Int. Symp., ISoLA 2022, Proceedings, Part III, LNCS, Springer, 2022, pp. 65–81.
[19] D.-H. Ruben, Explaining Explanation, 2nd ed., Routledge, 2012.
[20] R. Confalonieri, L. Coba, B. Wagner, T. R. Besold, A historical perspective of explainable artificial intelligence, WIREs Data Mining Knowl. Discov. 11 (2021). doi:10.1002/WIDM.1391.
[21] D. A. Neu, J. Lahann, P. Fettke, A systematic literature review on state-of-the-art deep learning methods for process prediction, AI. Rev. 55 (2022) 801–827. doi:10.1007/S10462-021-09960-8.
[22] S. Weinzierl, S. Zilker, S. Dunzer, M. Matzner, Machine learning in business process management: A systematic literature review, Expert Systems with Applications 253 (2024) 124181. doi:10.1016/J.ESWA.2024.124181.
[23] M. Stierle, J. Brunk, S. Weinzierl, S. Zilker, M. Matzner, J. Becker, Bringing light into the darkness - A systematic literature review on explainable predictive business process monitoring techniques, in: ECIS 2021, 2021. URL: https://aisel.aisnet.org/ecis2021_rip/8.
[24] A. Stevens, J. D. Smedt, Explainability in process outcome prediction: Guidelines to obtain interpretable and faithful models, Eur. J. Oper. Res. 317 (2024). doi:10.1016/J.EJOR.2023.09.010.
[25] M. Harl, S. Weinzierl, M. Stierle, M. Matzner, Explainable predictive business process monitoring using gated graph neural networks, J. Decis. Syst. 29 (2020). doi:10.1080/12460125.2020.1780780.
[26] N. Mehdiyev, M. Majlatow, P. Fettke, Augmenting post-hoc explanations for predictive process monitoring with uncertainty quantification via conformalized monte carlo dropout, Data Knowl. Eng. 156 (2025) 102402. doi:10.1016/J.DATAK.2024.102402.
[27] T. Narendra, P. Agarwal, M. Gupta, S. Dechu, Counterfactual reasoning for process optimization using structural causal models, in: LNBIP, volume 360, 2019. URL: https://doi.org/10.1007/978-3-030-26643-1_6.
[28] A. J. Alaee, M. Weidlich, A. Senderovich, Data-driven decision support for business processes: Causal reasoning and discovery, in: BPM Forum, Springer, 2024, pp. 90–106. doi:10.1007/978-3-031-70418-5_6.
[29] F. Fournier, L. Limonad, I. Skarbovsky, Towards a benchmark for causal business process reasoning with LLMs, in: BPM-W, Springer, 2025, pp. 233–246. doi:10.1007/978-3-031-78666-2_18.
[30] A. Buliga, M. Vazifehdoostirani, L. Genga, X. Lu, R. M. Dijkman, C. D. Francescomarino, C. Ghidini, H. A. Reijers, Uncovering patterns for local explanations in outcome-based predictive process monitoring, in: BPM 2024, volume 14940 of *LNCS*, Springer, 2024, pp. 363–380. doi:10.1007/978-3-031-70396-6_21.
[31] A. Füßl, V. Nissen, S. H. Heringklee, An explanation user interface for a knowledge graph-based XAI approach to process analysis, in: CAiSE-W 2024, volume 521 of *LNBIP*, Springer, 2024, pp. 72–84. doi:10.1007/978-3-031-61003-5_7.
[32] A. Goossens, U. Maes, Y. Timmermans, J. Vanthienen, Automated intelligent assistance with explainable decision models in knowledge-intensive processes, in: BPM-W 2022, Münster, volume 460 of *LNBIP*, Springer, 2022, pp. 25–36. doi:10.1007/978-3-031-25383-6_3.
[33] L. Limonad, F. Fournier, H. Mulian, G. Manias, S. Borotis, D. Kyrkou, Selecting the right LLM for eGov explanations, 2025. arXiv:2504.21032.

[34] T. Speith, A review of taxonomies of explainable artificial intelligence (xai) methods, in: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, FAccT '22, Association for Computing Machinery, 2022, p. 2239–2250. doi:10.1145/3531146.3534639.

[35] B. Kuehnert, R. Kim, J. Forlizzi, H. Heidari, The "who", "what", and "how" of responsible AI governance: A systematic review and meta-analysis of (actor, stage)-specific tools, in: Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency, ACM, 2025, p. 2991–3005.