Exploring Counterfactual Explanations with Temporal and Process Contexts in Predictive Process Monitoring (Extended Abstract)

Andrei Buliga[1,2]

[1]Fondazione Bruno Kessler Trento, Italy
[2]Free University of Bozen-Bolzano, Bolzano Italy

Abstract
Recent efforts in Predictive Process Monitoring (PPM), have increasingly relied on advanced Machine Learning (ML) and Deep Learning (DL) models, which are often considered black boxes due to their internal reasoning being difficult, if not impossible to understand. In light of the transparency and explanation requirements introduced by regulatory frameworks such as the GDPR and the EU AI Act, this has made the explainability of automated decisions increasingly important. This thesis addresses this challenge by developing methods not only to generate process-aware explanations, but also to evaluate them systematically and leverage them for purposes beyond individual predictions. By making the decisions of existing PPM models more transparent, these contributions aim to open their black boxes and foster the practical adoption of predictive monitoring systems. Beyond PPM, this thesis also provides process-oriented perspectives that can inform the design and evaluation of explainability methods in other temporal and structured domains.

Keywords
explainable AI, predictive process monitoring, temporal background knowledge

1. Introduction

Predictive Process Monitoring (PPM) is an emerging field within Process Mining that focuses on forecasting the future course of ongoing process instances, enabling organisations to anticipate outcomes and support critical decisions in domains such as finance, healthcare, and public administration. Advances in machine learning have enabled highly accurate predictions; yet, these models remain black boxes whose reasoning is inaccessible to end users. This opacity undermines trust and accountability and raises serious challenges for organisations operating under strict regulatory frameworks, such as the GDPR and the EU AI Act, where individuals are entitled to receive understandable explanations of algorithmic decisions that affect them.

In the vision of AI-Augmented Business Process Management Systems (ABPMS) [1], process-aware explainability is not a peripheral feature but an essential property of next-generation process management tools. Explanations in this context must go beyond generic explanations based on individual input features and explicitly respect the temporal and structural nature of business processes. However, most efforts to introduce explainability into PPM have transferred techniques from the broader Explainable AI domain with limited adaptation to process-specific characteristics [2]. As a result, explanations may overlook the *temporal background knowledge*, i.e., temporal rules and constraints that govern process executions, limiting their meaningfulness and practical usefulness.

While initial efforts in Explainable PPM (XPPM), such as [3, 4, 5, 6, 7], established a promising research direction, research has focused mainly on identifying the most influential factors leading to predictions, i.e., factual explanations, while considerably less attention has been devoted to offering actionable alternatives during execution, i.e., counterfactual explanations [8]. Existing counterfactual generation techniques, such as [7, 6], have also been evaluated predominantly through qualitative

Proceedings of the BPM 2026 Best Dissertation Award, Doctoral Consortium, and Demonstration & Resources Forum (BPM 2026), Toronto, Canada, September 27th to October 2nd, 2026.
abuliga@fbk.eu (A. Buliga)
0000-0002-8179-9146 (A. Buliga)

criteria, with no standardised evaluation framework for counterfactual explanations in Process Mining. Moreover, these techniques do not guarantee that the generated counterfactual explanations comply with the temporal background knowledge of the underlying process.

Therefore, the current thesis aims to address the following research questions:

RQ1. How should process-aware explanations be evaluated?

RQ2. How can process-aware explanations be generated using temporal background knowledge?

RQ3. How can process-aware explanations be leveraged beyond their explanatory role?

In the following, we give some context for the reader in Section 2, before moving to the contributions of the thesis in Section 3.

2. Background and Context

Modern machine learning models can learn complex relationships from historical data and produce highly accurate predictions. However, for many models, particularly deep neural networks and ensemble methods, it is difficult to understand how the input information is transformed into a prediction. These models are commonly described as *black boxes*: users can observe the information provided to the model and the prediction it returns, but not the reasoning connecting the two [9, 10]. In PPM, this entails that a model may predict that an ongoing trace will have an undesired outcome without clearly indicating which elements of the process execution contributed to that prediction.

Explainable Artificial Intelligence (XAI) aims to make such predictions more understandable by providing additional information about the behaviour of the predictive model [10]. Explanations can be distinguished according to their *scope*:*global explanations* describe the general behaviour of a model, for example by identifying which activities or attributes typically influence the behaviour of the model across the entire event log; *local explanations*, instead, focus on one particular prediction and explains why the model produced that result for a specific ongoing trace.

Explanations also differ according to the question they answer. A *factual explanation* describes which attributes of the observed trace contributed to the current prediction. For instance, it may indicate that a loan application was rejected because a particular activity occurred, or the processing time exceeded a certain threshold. A *counterfactual explanation* instead describes what would need to change for the model to produce a different outcome [11]. For example, it may indicate that the prediction would change from rejection to acceptance if an additional review activity were performed or if a different amount was requested for the loan.

Counterfactuals are particularly relevant to PPM as they provide actionable alternatives rather than only identifying influential features. However, alternative executions that change the prediction are not necessarily meaningful from a process perspective. They may suggest an infeasible activity order, or violate process rules. This motivates the need for *process-aware explanations*: explanations that consider not only the behaviour of the predictive model, but also compliance w.r.t the underlying temporal background knowledge.

3. Contributions

This thesis aims to answer to the three questions above, through three main contributions. First, it introduces an evaluation framework tailored to process-aware explanations in PPM, defining principled criteria for assessing their quality and usefulness. Second, it develops novel methods for generating explanations that ensure they remain realistic and consistent with the structure and constraints of real processes. Third, it advances the field by demonstrating how explanations can be leveraged in more refined ways: either focusing on summarised process-aware explanations of predictive models, or on the use of explanations for data generation and what-if scenario analysis, enabling the exploration of explanations beyond their initial application.

RQ1. How should process-aware explanations be evaluated? This direction started from the observation that existing counterfactual explanation techniques were generally assessed through heterogeneous or qualitative criteria, while the broader XAI community still lacked agreement on a standard evaluation methodology [10]. The thesis therefore introduced a process-aware evaluation framework combining general counterfactual desiderata, such as validity, proximity, plausibility, sparsity, and diversity, with dimensions specific to process data, including temporal conformance. The framework also considers the influence of the generation strategy, trace encoding, event-log characteristics, and the point at which an ongoing trace is explained.

Out of the various desiderata, diversity can be particularly important in PPM as the same prediction may be achieved through several alternative process executions, as the same prediction may be reached through several genuinely different feasible alternatives rather than minor variations of a single trace. To assess whether these alternatives remain consistent with the process, we measure temporal conformance using declarative temporal background knowledge, whose flexibility allows multiple valid execution paths, foregoing the need to prescribe a single procedural model [12].

While, the evaluation revealed that no generation strategy is universally preferable, the proposed framework offers a standardized framework for assessing counterfactual explanations in PPM. Nonetheless, when comparing the different types of method,we observed that generative methods favor speed and minimal changes, but often sacrifice plausibility and process alignment, while case-based methods offer more realistic explanations at higher cost, and are subject to the amount of traces that are available for computation. Instead, hybrid methods, such as Genetic Algorithms, offered an intermediate trade-off by combining existing trace examples with the flexibility to explore new alternatives. This finding connects to the next research questions as a starting point.

RQ2. How can process-aware explanations be generated using temporal background knowledge? The evaluation stage showed that counterfactuals generated through process-agnostic methods may change the prediction while failing to respect the behavioural rules of the underlying process. The second research direction therefore investigated how temporal background knowledge could be integrated into counterfactual generation itself.

The proposed methods use Genetic Algorithms as their population-based search can balance multiple counterfactual desiderata and their crossover and mutation operators can be adapted to process semantics. Moreover, they are easily applied to any type of predictive model and naturally support the exploration of diverse candidate solutions.

To this end, the thesis developed novel model-agnostic methods that incorporate temporal background knowledge directly into the search for counterfactual explanations. Two complementary representations of temporal knowledge were considered. The first defines temporal background knowledge through high-level declarative patterns (DECLARE), commonly used in Process Mining [12]. The second contributions works directly with the temporal logic underlying these rules, providing finer control over the generation process. Although they offer different degrees of flexibility and control, both approaches use the constraints to guide candidate construction, producing counterfactuals that guarantee conformance to the process by design rather than repairing or discarding non-conformant candidates after generation.

The results demonstrate that conformance can be guaranteed without compromising the main qualities expected from counterfactual explanations, thus producing good process-aware counterfactual explanations. These results also motivated investigating whether counterfactuals could be leveraged beyond their original explanatory purpose, as explored in RQ3.

RQ3. How can process-aware explanations be leveraged beyond their explanatory role? Once counterfactuals can be evaluated systematically and generated in accordance with process knowledge, the final research direction considers whether their usefulness can extend beyond explaining individual predictions. The thesis explores two such cases. The first contribution uses sets of factual and counterfactual examples to derive local process-pattern explanations. Rather than presenting several alternative traces individually, recurring patterns are used to summarise the conditions associated with maintaining or changing a prediction. This provides trace-specific explanations through process-level abstractions, capturing combinations of activities and context through process patterns that better reflect how process analysts reason about process behaviour.

Second, counterfactuals are used to generate alternative process executions for what-if analysis, allowing analysts to explore how a process could behave under a desired condition. The proposed contribution enables the realistic simulation of what-if scenarios across multiple process perspectives, by leveraging the integration of counterfactual explanation generators with Business Process Simulation.

These contributions show that counterfactuals can be treated not only as explanations presented to users, but also as reusable artefacts supporting broader tasks such as model understanding, process exploration, and scenario generation.

4. Conclusions and Significance

This thesis represents one of the first systematic and rigorous efforts to adapt Explainable AI techniques to the specifics of Predictive Process Monitoring. Three connected directions were explored: *evaluating, generating, and leveraging counterfactual explanations*. First, it introduced a process-aware framework for *evaluating* whether counterfactual explanations are not only valid for the predictive model, but also meaningful with respect to the underlying process. Second, it provided novel methods for *generating* conformant counterfactual explanations by construction through the use of different representations of temporal background knowledge. Finally, it investigated *leveraging* counterfactuals beyond individual explanations, and using them to derive compact process-pattern explanations or to generate alternative event data for what-if analysis.

Together, these contributions position the research as a bridge between the Process Mining and Artificial Intelligence domains: on one side, adapting and extending XAI methods for the realities of business processes; on the other, showing the AI community how process-oriented perspectives can enrich the design and evaluation of XAI methods, especially in temporal domains, such as sequential decision making tasks [13]. More broadly, it also highlights how symbolic temporal knowledge and statistical learning can be combined in a Hybrid AI setting to produce explanations that are both predictive and aligned with domain constraints, with potential relevance for structured and safety-critical applications beyond Process Mining [14, 11, 15].

Future research could extend this work by: (i) investigating hybrid approaches that integrate temporal and multi-perspective process knowledge directly into neural counterfactual generators, avoiding explicit intermediate representations and improving scalability; (ii) developing multi-trace counterfactuals in which alternatives are generated jointly to account for shared resources, temporal dependencies, and process-level effects, rahter than independently generating counterfactual explanations; and (iii) conducting user-centred studies to evaluate how counterfactual explanations influence users' understanding and trust of predictive models, and ability to make informed decisions. Such studies would complement technical evaluation and clarify whether process-aware explanations can foster the practical adoption of PPM. Together, these directions move XPPM closer to the vision of AI-Augmented Business Process Management, in which predictive and explanatory capabilities are integrated into interactive, trustworthy, and process-aware decision-support systems.

Declaration on the Use of AI

During the preparation of this manuscript, the author used an LLM-based AI assistant for proofreading, and language editing to improving the clarity of the ,manuscript. The author takes full responsibility for the content of this publication.

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. URL: https://doi.org/10.1145/3576047. doi:10.1145/3576047.

[2] P. Ceravolo, M. Comuzzi, J. De Weerdt, C. Di Francescomarino, F. M. Maggi, Predictive process monitoring: concepts, challenges, and future research directions, Process Science 1 (2024) 1–22. URL: https://doi.org/10.1007/s44311-024-00002-4.

[3] W. Rizzi, C. D. Francescomarino, F. M. Maggi, Explainability in predictive process monitoring: When understanding helps improving, in: D. Fahland, C. Ghidini, J. Becker, M. Dumas (Eds.), Business Process Management Forum - BPM Forum 2020, Seville, Spain, September 13-18, 2020, Proceedings, volume 392 of *LNBIS*, Springer, 2020, pp. 141–158. doi:10.1007/978-3-030-58638-6_9.

[4] A. Padella, M. de Leoni, O. Dogan, R. Galanti, Explainable process prescriptive analytics, in: 2022 4th International Conference on Process Mining (ICPM), 2022, pp. 16–23. doi:10.1109/ICPM57379.2022.9980535.

[5] N. Mehdiyev, P. Fettke, Local post-hoc explanations for predictive process monitoring in manufacturing, in: F. Rowe, R. E. Amrani, M. Limayem, S. Matook, C. Rosenkranz, E. A. Whitley, A. E. Quammah (Eds.), 29th European Conference on Information Systems, ECIS 2021, Marrakech, Morocco, 2020, 2021. URL: https://aisel.aisnet.org/ecis2021_rp/35.

[6] C. Hsieh, C. Moreira, C. Ouyang, Dice4el: Interpreting process predictions using a milestone-aware counterfactual approach, in: C. Di Ciccio, C. Di Francescomarino, P. Soffer (Eds.), 3rd International Conference on Process Mining, ICPM 2021, Eindhoven, The Netherlands, October 31 - Nov. 4, 2021, IEEE, , 2021, pp. 88–95. doi:10.1109/ICPM53251.2021.9576881.

[7] T.-H. Huang, A. Metzger, K. Pohl, Counterfactual explanations for predictive business process monitoring, in: M. Themistocleous, M. Papadaki (Eds.), Information Systems, Springer International Publishing, Cham, 2022, pp. 399–413. URL: https://doi.org/10.1007/978-3-030-95947-0_28.

[8] 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, ECIS Research-in-Progress Papers (2021).

[9] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, ACM Comput. Surv. 51 (2019) 93:1–93:42. doi:10.1145/3236009.

[10] A. B. Arrieta, N. D. Rodríguez, J. D. Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, F. Herrera, Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion 58 (2020) 82–115. doi:10.1016/J.INFFUS.2019.12.012.

[11] S. Verma, V. Boonsanong, M. Hoang, K. Hines, J. Dickerson, C. Shah, Counterfactual explanations and algorithmic recourses for machine learning: A review, ACM Comput. Surv. 56 (2024). doi:10.1145/3677119.

[12] C. Di Ciccio, M. Montali, Declarative Process Specifications: Reasoning, Discovery, Monitoring, Springer International Publishing, Cham, 2022, pp. 108–152. doi:10.1007/978-3-031-08848-3_4.

[13] J. a. Fonseca, A. Bell, C. Abrate, F. Bonchi, J. Stoyanovich, Setting the right expectations: Algorithmic recourse over time, in: Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization, EAAMO '23, Association for Computing Machinery, New York, NY, USA, 2023. URL: https://doi.org/10.1145/3617694.3623251. doi:10.1145/3617694.3623251.

[14] A. d'Avila Garcez, L. C. Lamb, Neurosymbolic AI: the 3rd wave, Artif. Intell. Rev. 56 (2023) 12387–12406. URL: https://doi.org/10.1007/s10462-023-10448-w. doi:10.1007/S10462-023-10448-W.

[15] S. Teso, Ö. Alkan, W. Stammer, E. Daly, Leveraging explanations in interactive machine learning: An overview, Frontiers in Artificial Intelligence Volume 6 - 2023 (2023). doi:10.3389/frai.2023.1066049.