Process Mining and Artificial Intelligence for the Design, Sustainability, and Improvement of Collaborative Processes

Martín Rubio[1]

[1]Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Uruguay

Abstract

In recent decades, Process Mining has played a key role in improving organizational processes. Today's growing organizational complexity, with collaborative processes involving multiple actors and interactions, demands new perspectives for Business Process (BP) design, analysis, and improvement. At the same time, there is increasing interest in reducing the environmental impact of human activities, in which processes are naturally involved. BPs sustainability evaluation is key for organizations to be able to improve their operations and minimize its impact. Artificial intelligence (AI) have gained relevance in all areas including process mining, to help and enhanced BPs evaluation and analysis, taking advantage of the large amounts of available process execution data. This doctoral work main objective is to expand the scope of process mining and AI by integrating techniques, practices, and tools for the design, sustainability and improvement of collaborative processes. The aim of this work is to reduce modeling effort, improve the execution and sustainability of collaborative processes, and reduce the gap between models and their actual execution. Validation of results will focus on controlled experiments and experts surveys, as well as case studies in a real context. The proposal will contribute to the field of PM for collaborative processes and sustainability analysis, helping organizations towards evidence-based BPs improvement.

Keywords

Process mining, collaborative processes, sustainability, artificial intelligence

1. Introduction

Over the past decades, process mining [1] has emerged as a discipline that helps organizations analyze their business processes (BPs). These processes reflect increasingly complex technological ecosystems, while organizations show a growing interest in reducing the environmental impact of human activities in which they are embedded. Sustainability [2] is the use of resources without compromising future generations, balancing economic growth, social well-being, and environmental protection. Assessing and improving process sustainability requires measuring their execution through sustainability-related data and measures. This motivates the inclusion of a sustainability dimension in BPs analysis [3], within the devil's quadrangle of time, cost, quality and flexibility [4].

Within collaborative processes [5], that deals with dependencies and interactions among multiple participants within a single overarching process, process design for sustainability and its assessment and improvement have to deal with elements such as data privacy, heterogeneous technologies and data registration. This distributed nature, embedded in diverse technical contexts, presents several challenges for data collection, preparation, discovery and analysis of collaborative processes [6], including the sustainability dimension. In particular, traditional XES event logs and object-centric event logs must deal with data concerning different participants and messages exchanged.

In recent years, artificial intelligence (AI) has permeated all areas, including process mining, enabling the integration of new approaches for design, automation, and data analysis, such as machine learning (ML) for predicting future events based on historical event logs, or content generation using large language models (LLMs), trained on vast amounts of data, to assist with diverse tasks such as process modeling or execution trace analysis [7, 8, 9].

ICPM Doctoral Consortium and Demo Track 2025, October 20-24, 2025, Montevideo, Uruguay

mrubio@fing.edu.uy (M. Rubio)

In this context, this doctoral thesis seeks to extend process mining with artificial intelligence and a sustainability focus, integrating new techniques, practices, methodologies, and tools to design and improve collaborative processes.The integration of different AI techniques, such as process data generation and analysis of traces through generative models, pattern recognition, or agent-based modelling supported by large language models (LLMs) will enhance process mining techniques providing new alternatives and results for sustainability analysis and process design and improvement. This work aims to ease modeling, enhance execution and sustainability in collaborative processes, and narrow the gap between models and execution.

The remainder of this document is organized as follows: Section 2 presents the motivation and research questions. Section 3 reviews related work. Section 4 describes the methodology, and Section 5 outlines the proposal. Finally, Section 6 presents the conclusions.

2. Motivation and Research questions

Process mining provides analytical tools that can be adapted to the proposed context, even in collaborative processes, and within traditional event logs or object-centric event logs. Few proposals address the automation and improvement of BPs with a sustainability focus; a key enabler is collecting and integrating sustainability data into their execution for further analysis and tool support.

Also, the current state of several artificial intelligence applications offers a new set of readily available methods and tools, such as ML, LLMs and Agentic-based approaches (c.f. Section 3, which the integration within process mining techniques and perspectives we aim to asses and provide within the approach. Based on the previous statements, the research question is defined as follows: How can process mining be extended with artificial intelligence methods to capture, analyze, and improve sustainability aspects in collaborative processes?. From the general research question, specific research questions are defined:

- **RQ1.** What process mining-based techniques, practices, methodologies and tools could be useful for the design, sustainability analysis and improvement of collaborative business processes?
- **RQ2.** What process mining approaches exists to deal with business process sustainability and which artificial intelligence methods are integrated, in particular for collaborative processes, and what limitations they present?. The answer to this research question, which will be obtained by a systematic literature review [10, 11], provides the state of the art for the thesis work.
- **RQ3.** How can artificial intelligence methods be integrated within process mining techniques to enhance the design, sustainability, and improvement of collaborative processes?

3. Related work

The application of process mining techniques (discovery, conformance, enhancement, prediction) on business process execution data [5] has primarily focused on orchestration-type processes carried out within a single organization (intra-organizational) and, more recently, on collaborative (inter-organizational) processes [12, 13, 14, 6, 15, 16], mainly in the area of process discovery. The Object-Centric Event Data (OCED) [17] event log, has been introduced as the successor to the traditional XES standard (IEEE 1849-2016) [18] for process log representation, particularly from an object-oriented perspective. This object-oriented log expands the analytical capabilities for process analysis, especially in collaborative processes.

On the other hand, sustainability-oriented processes and approaches such as Green BPM have gained relevance. In [3], a review shows that the main focus has been on measuring and controlling emissions or energy consumption in business processes, primarily addressing the environmental dimension. There are limited approaches that explicitly address specific sustainability measures in business processes [19, 20], particularly those amenable to analysis through process mining [21, 22].

The evaluation of ICTs in [23] aims to manage energy resources efficiently while controlling the impact of the performed activities, and [24] analyzes how supporting software, such as BPMS platforms, artificial

intelligence algorithms [25], and process discovery techniques [26], influences energy consumption. The notion of "Green Data Science" (GDS) was introduced in [27] due to the potential "pollution" that data science itself can cause.

Advances in artificial intelligence have enabled the integration of new automation and data analysis approaches, such as machine learning (ML) for predicting future events [28, 29], or the creation of new content through large language models (LLMs) to assist in various tasks. In particular, LLMs have been used to generate textual descriptions from process data and event logs, to derive process models from textual descriptions, and to leverage textual descriptions of models and event logs for process mining [7, 8, 30], as well as to assess requirements for the use of LLMs and related tools [9].

4. Metodology

To guide the research, the principles of Design Science are followed, [31, 32, 33]. focusing on the creation of artifacts such as: a framework and working methodology, different models (processes, ML, LLMs, and prompt engineering), and algorithms for process mining, data recording and management, sustainability measures, among others. Following this methodology, the research started with the problem definition to identify hypotheses and research questions, followed with a state-of-the-art review on the selected topics, based on the guidelines for systematic literature reviews and mapping studies proposed in [10, 11]. The search protocol was defined focusing on process mining and sustainability first, and then process mining and AI, to find proposals covering the main topics of the work, that could be taken into account for collaborative processes. The proposal is then defined based on the creation of artifacts to solve the problem identified, which will be validated using different approaches (c.f. Subsection 4.1), and with publications in leading regional and international conferences and journals.

4.1. Validation of results

For the validation of the different results, the guidelines in [31, 34] will be followed, considering both the technical validation and the empirical validation of the proposed artifacts. For instance, controlled experiments and surveys will be carried out to validate the adequacy of the proposed solutions with domain experts (e.g., sustainability measures and analysis), to gather their opinions on the usefulness, applicability, and suitability of the proposals in real-world contexts, and with functional prototypes. Furthermore, case studies [34, 35] are planned within the context of real collaborative e-Government processes (AGESIC)[1], which will allow assessing applicability as well as identifying opportunities for improvement and limitations.

5. Proposal

The proposal integrates sustainability into the entire process life cycle, from design and implementation to execution, analysis, and improvement. The focus is on collaborative business processes supported by information systems, particularly BPMS, though it can also extend to other domains. Figure 1 provides an overview, illustrating a BPMS for BPs execution as example.

A key component is a systematic approach for registering sustainability-related data during process execution (i.e. parameters and measures values), drawing, for instance, on the measures described in [23, 19] with corresponding calculation formulae. Conceptual components and tools support this registration by capturing parameter values at execution time. They also distinguish between estimated and measured data, helping users interpret the resulting values appropriately. From process execution, event logs are generated, containing sustainability-related data registered directly during execution i.e. parameters values (and optionally also measures calculated values), as well as adding measures calculation after a log enrichment process, using as basis the registered data i.e. parameters values.

[1] https://www.gub.uy/agencia-gobierno-electronico-sociedad-informacion-conocimiento/

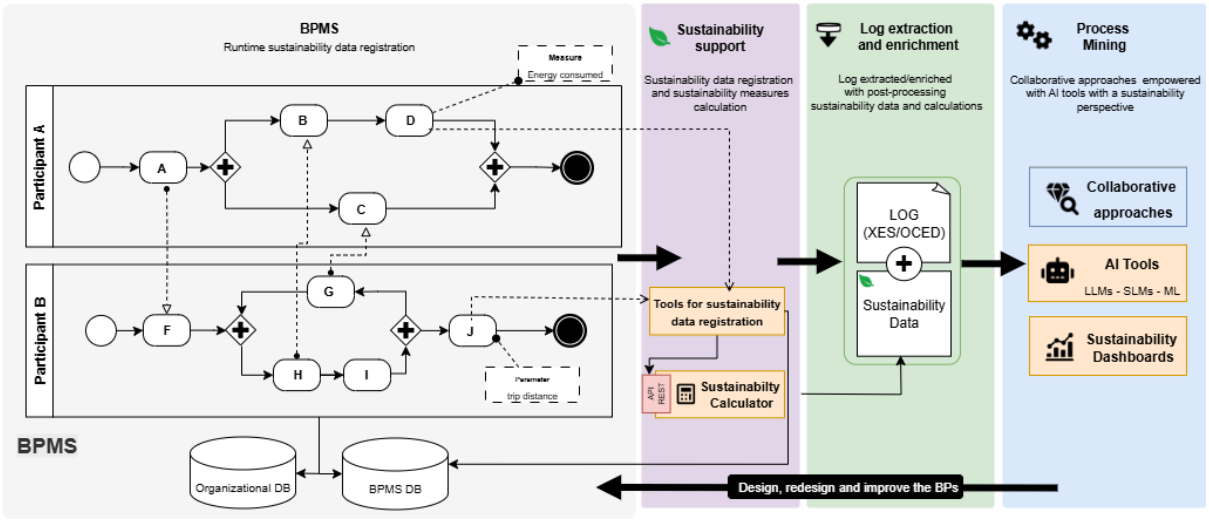

Figure 1: General Overview of the Proposed Approach

These options are described in subsection 5.1, and applies for traditional XES and object-centric event logs.

With the sustainability extended event logs, process mining is performed using traditional approaches, complemented by the new possibilities offered by artificial intelligence tools such as Large Language Models (LLMs) [7, 8, 9] and machine learning in general. The integration of agentics, specific prompt engineering techniques and tools to interact with LLMs such as [36], to provide users with extended support for the analsys of BPs execution with focus on sustainability. A dedicated sustainability dashboard is provided, which presents sustainability measures data and process mining results, and will integrate AI modules, to help users identifying evidence-based improvement opportunities, with sustainability patterns and process improvements guided recommendations.

5.1. Registration of sustainability data

Existing approaches that incorporate sustainability attributes into business processes do not clearly specify how such data can be systematically registered and derived. To fill this gap, this work proposes a systematic approach that integrates concrete sustainability measures based on [23, 19], whose values are either registered directly during execution or calculated in a post-processing stage using sustainability parameters registered during runtime. The focus is on the enviromental dimension, including categories such as energy, emission, material, water, waste and software. We extended them by adding the concrete formulae and parameters needed for calculation, and adding measures related to BPs execution in an information system, such as the language of implementation e.g. Java, Python, etc, the CPU, cloud, among others, which can be estimated based on existing reference data, such as CO2 emissions.

The proposal includes a method to register and calculate sustainability measures, an extension to the XES format to include sustainability data (and translation to object-centric event logs), specific sustainability measures and a measurement taxonomy that distinguish between estimated and actually measured values (e.g. with a phisical devise such as sensors) differentiating approaches by their degree of empiricism, to help users in interpreting the results. Tool support is also provided for sustainibility data registration and microservices-based sustainability measures calculator.

5.2. AI support for process mining tasks

Various artificial intelligence tools can be applied to process mining, as has traditionally been done with machine learning, and more recent studies are focusing on LLMs. Evidence, such as that presented in [37], shows satisfactory performance in a range of process mining tasks using large-scale commercial LLMs. However, it is worth noting that the environmental impact of LLMs can be considerable when

deployed at a global scale [38]. For this reason, alternative approaches will be explored that leverage the capabilities of the technology available for process mining while, from a sustainability perspective, ensuring that the tools themselves do not become part of the problem. Through the concept of agents, multiple specialized process mining tools can be orchestrated in a coordinated manner across different stages of the process mining lifecycle to enhance their tasks. For example, the integration with Model Context Protocol (MCP) [39] servers to allow specific context for LLMs use will be evaluated.

5.3. Object-centric events logs

The integration of data associated with processes and organizational information, particularly sustainability-related data, is key in the context of the proposed work. In previous work the integration of organizational data into process logs [40] was addressed, defining methodologies and tools that enable a joint process and organizational data analysis. This will be translated also into OCED event logs and analsys, leveraging the metamodel and implementation described in [41], with focus on collaborative processes and associated challenges e.g. data privacy and security across organizations, common goals for sustainability, heterogeneous infrastructure, among others. On the other hand, since OCED is still a relatively recent standard, there are few process logs available to support research tasks. In this regard, several strategies are being analyzed to generate synthetic event logs based on OCED such as Conditional Tabular GANs (CTGANs) [42]. Preliminary results indicate that it is possible to generate high-quality synthetic logs using these technologies, which will serve as a valuable input for experimentation with different datasets.

5.4. BPs improvement recommendations

Based on the process mining with IA support analysis, integrating sustainability data and measures, improvement recommendations will be derived to help organizations improve their processes, with focus on collaborative scenarios. By adding the sustainability dimension to the Devil's Quadrangle similarly to [4], sustainability patterns and re-design heuristics will be proposed, that will help organizations to reshape their BPs and/or their organizational/technical support, to provide better results in terms of the BPs objectives, but with less environmental impact.

6. Conclusions

This work addresses the challenges inherent in improving collaborative processes by incorporating an environmental sustainability perspective through process mining, supported by advances in artificial intelligence. These challenges are addressed through the definition of several artifacts which will be validated from different perspectives with controlled experiments and surveys, and real-world cases of collaborative processes within the Uruguayan government, employing data provided by AGESIC.

Suitable techniques, practices, methodologies, improvement recommendations and supporting tools will be integrated and defined for the design and redesign of sustainable processes, taking into account and extending/adapting existing proposals, as well as defining new ones to achieve the stated objectives. The results of this work will be of help to organizations in order to analyze and improve their collaborative BPs with a sustainability focus, minimizing their environmental impact.

7. Acknowledgments

Partially supported by projects: "Minería de procesos con automatización robótica e IA generativa para el diseño y sostenibilidad de procesos colaborativos hiperconectados" funded by Comisión Sectorial de Investigación Científica (CSIC), Universidad de la República, Proy I+D 2024, "22520240100504UD", Uruguay and Academic Graduate program PEDECIBA Informática, Ministry of Education and Culture (MEC) and UdelaR, Uruguay.

8. Declaration on Generative AI

During the preparation of this work, the author used Chat-GPT-5 in order to: Text Translation and Grammar and spelling check. After using these tool, the author reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] W. van der Aalst, Process Mining: Data Science in Action, 2nd.ed., Springer, 2016.

[2] G. Brundtland, Report of the world commision on enviroment and development: our common future, in: United Nations, 1987.

[3] D. Couckuyt, A. V. Looy, A systematic review of green business process management, Business Process Management Journal 26 (2020) 421–446.

[4] H. A. Reijers, Heuristic Workflow Redesign, Springer Berlin Heidelberg, Berlin, Heidelberg, 2003, pp. 207–243. doi:10.1007/3-540-36615-6_6.

[5] W. M. P. van der Aalst, Intra- and inter-organizational process mining: Discovering processes within and between organizations, in: The Practice of Enterprise Modeling, Springer, Cham, 2011, pp. 1–11. doi:10.1007/978-3-642-24849-8_1.

[6] L. Peña, D. Andrade, A. Delgado, D. Calegari, Inter-organizational collaborative bpmn 2.0 business process discovery, Journal of Intelligent Information Systems (2024).

[7] M. Grohs, L. Abb, N. Elsayed, J.-R. Rehse, Large language models can accomplish business process management tasks, in: Business Process Management Workshops, Springer, 2024, pp. 453–465.

[8] H. Kourani, A. Berti, J. Hennrich, W. Kratsch, R. Weidlich, C.-Y. Li, A. Arslan, D. Schuster, W. M. P. van der Aalst, Leveraging large language models for enhanced process model comprehension, CoRR abs/2408.08892 (2024).

[9] A. Berti, H. Kourani, H. Häfke, C.-Y. Li, D. Schuster, Evaluating large language models in process mining: Capabilities, benchmarks, and evaluation strategies, in: Enterprise, Business-Process and Inf.Sys. Modeling, Springer, 2024, pp. 13–21.

[10] B. Kitchenham, Procedures for Performing Systematic Reviews, Technical Report TR/SE-0401, Keele University and Empirical Soft. Eng. Joint Report, Keele, UK, 2004.

[11] B. A. Kitchenham, S. Charters, Guidelines for performing Systematic Literature Reviews in Software Engineering, Technical Report EBSE 2007-001, Keele University and Durham University Joint Report, Keele, UK, 2007.

[12] C. Liu, H. Li, Q. Zeng, T. Lu, C. Li, Cross-organization emergency response process mining: An approach based on petri nets, Math. Problems in Eng. (2020) 1–12. doi:10.1155/2020/8836007.

[13] C. Liu, H. Li, S. Zhang, L. Cheng, Q. Zeng, Cross-department collaborative healthcare process model discovery from event logs, IEEE Transactions on Automation Science and Engineering 20 (2023) 2115–2125. doi:10.1109/TASE.2022.3194312.

[14] J. D. Hernandez, E. Tello, H. M. Marin, U. M. Ramirez, J. A. Mata, Merging event logs for inter-organizational process mining, in: New Perspectives on Enterprise Decision-Making Applying AI Techniques, Springer, 2021, pp. 3–26. doi:10.1007/978-3-030-71115-3_1.

[15] F. Corradini, S. Pettinari, B. Re, L. Rossi, F. Tiezzi, A technique for discovering bpmn collaboration diagrams, Software and System Modeling (2024) 1323–1343. doi:10.1007/s10270-024-01153-5.

[16] J.-V. Benzin, S. Rinderle-Ma, Collaboration miner: Discovering collaboration petri nets, in: Cooperative Information Systems, Springer, 2025, pp. 373–381. doi:10.1007/978-3-031-81375-7_24.

[17] D. Fahland, M. Montali, J. Lebherz, W. M. P. van der Aalst, M. van Asseldonk, P. Blank, L. Bosmans, M. Brenscheidt, C. D. Ciccio, A. Delgado, D. Calegari, J. Peeperkorn, E. Verbeek, L. Vugs, M. T. Wynn, Towards a simple and extensible standard for object-centric event data (OCED) - core model, design space, and lessons learned, CoRR abs/2410.14495 (2024).

[18] X. W. Group, et al., Ieee standard for extensible event stream (xes) for achieving interoperability in event logs and event streams, IEEE Std 1849 (2016) 1–50.

[19] J. Mancebo, F. García, O. Pedreira, M. Á. Moraga, Bpms-game: Tool for business process gamification, Business Process Management Forum, Springer, 2017, pp. 127–140.
[20] M. Medema, B. Popescu, V. Andrikopoulos, D. Karastoyanova, A life cycle and enabling concepts for green business process management., in: Int. Conf. on ICT for Sustainability (ICT4S), 2025.
[21] N. Graves, I. Koren, W. M. van der Aalst, Rethink your processes! a review of process mining for sustainability, in: 2023 International Conference on ICT for Sustainability (ICT4S), 2023, pp. 164–175. doi:10.1109/ICT4S58814.2023.00025.
[22] N. Graves, A. Fritsch, R. Hensen, I. Koren, W. Aalst, Object-centric process mining for semi-automated and multi-perspective sustainability analyses, in: International Conference on ICT for Sustainability (ICT4S), 2025.
[23] A. Hernández González, C. Calero, D. Pérez Parra, J. Mancebo, Approaching green BPM characterisation, Journal of Software: Evolution and Process 31 (2019).
[24] C. Calero, M. Á. Moraga, M. Piattini, Software Sustainability, Springer, 2021, pp. 1–15. doi:10.1007/978-3-030-69970-3_1.
[25] R. Schwartz, J. Dodge, N. A. Smith, O. Etzioni, Green ai, Commun. ACM 63 (2020) 54–63.
[26] A. Delgado, F. García, M. Á. Moraga, D. Calegari, A. Gordillo, L. Peña, Adding the sustainability dimension in process mining discovery algorithms evaluation, in: Business Process Management Forum, Springer, 2023, pp. 163–177.
[27] W. Van Der Aalst, Green data science - using big data in an "environmentally friendly" manner, in: Proceedings of the 18th International Conference on Enterprise Information Systems - Volume 1: ICEIS, SciTePress, 2016, pp. 9–21. doi:10.5220/0006806900010001.
[28] C. Di Francescomarino, C. Ghidini, Predictive Process Monitoring, volume 448, Springer, 2022, pp. 320–346. doi:10.1007/978-3-031-08848-3_10.
[29] 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) 2. doi:10.1007/s44311-024-00002-4.
[30] A. Berti, D. Schuster, W. M. P. van der Aalst, Abstractions, scenarios, and prompt definitions for process mining with llms: A case study, in: Business Process Management Workshops, Springer, 2024, pp. 427–439.
[31] R. J. Wieringa, Design Science Methodology for Inf. Systems and Software Eng., Springer, 2014.
[32] A. R. Hevner, S. T. March, J. Park, S. Ram, Design science in information systems research, MIS Q. 28 (2004) 75–105.
[33] J. Venable, A framework for design science research activities, in: Emerging Trends and Challenges in Information Technology Management, Idea group publishing, 2006, pp. 184–187.
[34] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, Experimentation in Software Engineering, Springer, 2012. doi:10.1007/978-3-642-29044-2.
[35] R. K. Yin, Case study research. Design and methods, 6th edition, Sage Publications, 2017.
[36] H. Kourani, A. Berti, D. Schuster, W. M. van der Aalst, Process modeling with large language models, in: Int. Conf. on Business Process Modeling, Dev. and Support, Springer, 2024, pp. 229–244.
[37] A. Berti, H. Kourani, W. M. van der Aalst, Pm-llm-benchmark: Evaluating large language models on process mining tasks, in: Int. Conference on Process Mining, Springer, 2024, pp. 610–623.
[38] N. Jegham, M. Abdelatti, L. Elmoubarki, A. Hendawi, How hungry is ai? benchmarking energy, water, and carbon footprint of llm inference, arXiv preprint arXiv:2505.09598 (2025).
[39] Anthropic, Introducing the model context protocol, https://www.anthropic.com/news/model-context-protocol, 2024. Accessed: 2025-08-12.
[40] M. Rubio, A. Delgado, L. Tansini, An integrated approach to process and organizational data mining, in: Latin American Computer Conf., IEEE, 2023, pp. 1–10.
[41] D. Calegari, A. Delgado, A model-driven engineering perspective for the object-centric event data (oced) metamodel, in: Int. Conf. on Business Process Management, Springer, 2023, pp. 508–520.
[42] L. Xu, M. Skoularidou, A. Cuesta-Infante, K. Veeramachaneni, Modeling tabular data using conditional gan, Advances in neural information processing systems 32 (2019).