Compliance Checking for Public Administration Processes using Retrieval-Augmented Generation in LLMs: Novel Directions and Challenges

Alessandro Gianola[1,*,†], Chrysoula Zerva[2,†]

[1]INESC-ID/Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[2]Instituto de Telecomunicações/Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

Abstract

Public administration processes operate within complex environments, where they need to comply with evolving legal norms. Nevertheless, their specifications and guidelines documentation are often unstructured and hard to analyze automatically. This position paper proposes a novel hybrid approach that combines Large Language Models, Retrieval-Augmented Generation, and symbolic reasoning to align process specifications with legal guidelines. We outline key research questions and introduce a pipeline for extracting, structuring, and aligning business-related, textual content expressed in natural language with regulatory guidelines, to support automated compliance checking in public administration contexts.

Keywords

BPM, Compliance Checking, LLMs, RAG, Alignments

1. Introduction

Public administration (PA) processes operate within complex and dynamic regulatory ecosystems, where legality and transparency must be continuously upheld [1]. These processes are governed by a multitude of evolving legal requirements, many of which are specified in textual formats, such as regulations or policy documents. Moreover, they are intended to follow structured "happy paths" envisioned by legislators. However, in practice, PA processes frequently deviate from these idealized flows due to the realities of day-to-day operation and evolving societal needs [1].

As a result, a particularly pressing issue in this context is compliance checking [2, 3, 4], which is the task of verifying whether processes, as they are executed concretely, conform to external normative constraints. Unlike in industrial or manufacturing domains where processes are typically well documented and structured [5], PA processes are accompanied by documentation that is often highly unstructured, stored in heterogeneous formats (e.g., PDF specifications, procedural manuals, regulatory texts), and articulated in natural language.

The specific compliance problem in which we are interested in this work concerns the *alignment* of contractual specification documents of PA processes (e.g., document files defining enterprise architecture references, or technological requirements) with legal guidelines that these processes should satisfy, which are often unstructured and expressed in natural language. These documents contain essential business-related information, possibly from concrete process executions, and assume a deep understanding of the underlying processes they constrain.

Traditionally, compliance verification in PA processes has relied on manual audits or basic document review procedures. These approaches are often time-consuming and error-prone, especially given the scale and complexity of the information that must be examined, making automated compliance

Overlay 2025
*Corresponding author.
†These authors contributed equally.
alessandro.gianola@tecnico.ulisboa.pt (A. Gianola); chrysoula.zerva@tecnico.ulisboa.pt (C. Zerva)
https://arsr.inesc-id.pt/~agianola/ (A. Gianola)
0000-0003-4216-5199 (A. Gianola); 0000-0002-4031-9492 (C. Zerva)

[1]https://reform-support.ec.europa.eu/what-we-do/public-administration-and-governance_en

checking a highly desirable goal. Classical methods for formal process analysis from Process Mining (PM) [6], a research area that integrates Business Process Management (BPM) [7] and Data Science, offer useful tools for examining process execution data registered in event logs, and exploit that information to improve and enact processes. PM relies on formal methods and automated reasoning, leveraging symbolic techniques, such as SAT/SMT solving [8, 9, 10], to analyze event logs and extract useful insights. However, these techniques typically rely on structured information contained in logs and well-defined formalisms and struggle to incorporate unstructured textual data. Even when PM methods incorporate other *perspectives*, such as interactions with contextual data [11, 12], the problem of analyzing processes involving unstructured data remains largely open.

This position paper explores an emerging solution space to handle compliance checking for PA processes in the presence of unstructured information. Recent advances in Artificial Intelligence (AI), particularly in Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG) [13], offer new opportunities to address this challenge. Indeed, LLMs have demonstrated remarkable capabilities in extracting, generating, reasoning about, and transforming natural language content, which makes them well suited for tasks involving textual information extraction [14, 15, 16, 17]. Moreover, various approaches attempting to combine natural language capabilities of LLMs with symbolic representations have recently emerged [18, 19, 20], such as the ones based on Chain-of-Thought (CoT;[21, 22]). The use of LLMs has gained momentum as a means to bridge the gap between unstructured textual data and formal representations that enable computational reasoning, such as the one that is required to solve PM tasks. RAG techniques further enhance LLMs by allowing dynamic retrieval of external knowledge, which is in turn provided as additional context to the model, thus enabling more accurate and context-aware outputs [23]. This is particularly useful in compliance scenarios where the analysis depends on retrieving relevant information among several long, complex documents with multiple interrelated legal and business sources.

However, current applications of LLMs and RAG face critical limitations: they often struggle with long contexts [24], lack structured process-awareness, and fail to systematically (i) incorporate formal business process knowledge and (ii) combine formal methods from PM to analyze processes. For PA compliance, this means they cannot yet reliably map between what the process specification documents declare and what regulations mandate. Furthermore, explicitly incorporating contextual knowledge from business process models or insights from event logs into LLM pipelines is still underexplored.

This work advocates for a novel direction: combining LLMs, RAG, and PM-inspired symbolic reasoning to support automated compliance checking for PA processes. We aim to develop an integrated method that *aligns* specification documents, enriched with contextual data from the underlying process, with regulatory guidelines. This novel approach will draw inspiration from the symbolic techniques employed in data-aware declarative process mining and conformance checking [25, 26, 27], where the alignment is with 'normative processes' expressed via declarative rules. Differently from alignments in PM, which compare execution log traces with process models, our challenge lies in aligning specification documents—enriched with business-relevant information (possibly incorporating information on concrete executions)—with regulatory guidelines, which shifts the focus from execution-model alignment to document-guideline alignment. Our method will be significantly enhanced with RAG techniques employed to allow LLMs to process the textual documents, in order to return formal representations that enable reasoning in a symbolic setting.

For this purpose, we propose a hybrid pipeline comprising: (i) a RAG-based LLM-driven method for extracting and structuring knowledge from diverse textual sources, grounded in PA-specific domain context; (ii) a formal symbolic schema to represent structured content extracted from documents; (iii) an alignment algorithm, inspired by the aforementioned procedures employed in declarative PM, capable of systematically comparing enriched structured specifications with formalized guidelines to compute compliance scores and identify discrepancies. The goal is to quantify the degree of alignment between these layers and identify the nature and severity of any discrepancies.

Our methodology is structured around four key research questions that will guide our proposal:

RQ1: How can RAG effectively integrate heterogeneous structured (e.g., databases, forms) and unstructured (e.g., policies, specifications) sources to enhance compliance-checking accuracy?

RQ2: What strategies can optimize retrieval and long-context processing in LLMs to preserve semantic fidelity and maximize interpretability?

RQ3: Can business process models or event logs serve as feedback mechanisms, helping align LLM outputs with real-world compliant behavior?

RQ4: How can the LLM extraction process be effectively integrated with a symbolic reasoner to compute alignments? Is it sufficient to decouple the process into three separate stages—information retrieval, schema population, and symbolic alignments?

This work aligns with a research initiative financed by a Portuguese project jointly with European funds[2] in collaboration with two PA entities: the Agência para a Modernização Administrativa (AMA) and the Instituto de Gestão Financeira e Equipamentos da Justiça (IGFEJ). These partnerships ensure that our research is grounded in real-world scenarios and continuously validated by public sector needs.

In summary, the final goal of this work is to outline the research directions and challenges for building an automatic compliance framework for PA, rooted in LLM-based reasoning. Our claim is that LLMs, enhanced with RAG and symbolic methods, can effectively process unstructured documents and regulatory guidelines, converting them into structured representations that allow automated reasoning.

2. Related works

As explained in the introduction, the objective is to assess the compliance of unstructured specification documents with normative guidelines for PA processes. In this section, we briefly present the works that are related to this research problem, which is intrinsically multidisciplinary. Indeed, the proposed approach spans multiple research areas. Natural Language Processing (NLP) is used to handle unstructured text through LLMs. PM contributes techniques for analyzing processes, particularly conformance checking. Building on PM, the approach also incorporates formal methods, such as SAT [28] and SMT solving [29], which serve as powerful backend reasoning tools to support alignment computations.

PM [30], rooted in BPM [31], integrates data science, AI, and formal methods and utilizes observed execution data from event logs to discover, analyze, and enhance processes. Given the complexity of contemporary processes interacting with relevant data objects, integrating control-flow and data aspects is essential [7, 32, 12]. This has driven multi-perspective PM, which analyzes processes through dimensions beyond control flow such as data, time, and resources. This is crucial for analyzing real processes in PAs, involving data objects shared among various departments.

One of the main tasks in multi-perspective PM is conformance checking [6]. Conformance checking identifies deviations and commonalities between a process model (which can also be expressed in a declarative way) and an event log by computing alignments. The goal is to find an optimal alignment with minimal cost, despite challenges posed by unbounded data (e.g., integers or reals) in model runs. Data-aware conformance checking focuses on AI techniques such as the A* algorithm [11] and leverages industry-strength automated reasoning tools like SMT solvers ([33, 10, 34] to handle unbounded data and compute the distance between observed traces and runs in expressive data-aware process models.

Compliance of process models ensures that the design and execution of business processes adhere to predefined rules, regulations, or legal guidelines [3]. Several approaches have been introduced (e.g.,[35, 36]), also using formal verification techniques [5]. Indeed, assessing compliance can be seen as an instance of more general verification against properties in some logic. While several approaches exist for the verification of multi-perspective processes, significant advancements have emerged through symbolic reasoning, where the most sophisticated ones are based on SMT solving [37, 38, 39, 12]. Other settings based on formal verification to check compliance leverage declarative process languages, such as Dynamic Condition Response (DCR) graphs, which can be used to express both the reference models from laws and the business process models [40]. However, in all the aforementioned approaches, laws are captured by formal symbolic models, and not by unstructured documents as usually in PA processes.

Aligning business documents with guidelines is akin to conformance checking. Indeed, the goal is to compute an 'alignment score' that measures their distance, with a higher score indicating more

[2]OptiGov: https://sciproj.ptcris.pt/176741PRJ

discrepancies. Also, if the guidelines precisely define a process model and vice versa, then establishing conformance of a log with the model is equivalent to assessing compliance with the guidelines (e.g., when expressed as logical rules via a declarative model [26, 27]). However, this is often not the case [4], since generally conformance implies compliance, but not necessarily vice versa. There could be deviations from the 'conforming runs' of the normative model that are still consistent with the laws. In addition, a simple compliance perspective is insufficient. In our scenario, we need to align with the guidelines (which generally do not define a process model), not just a model (or a log), but a document enriched with log information about the process executions. Given the presence of logs, we aim to build on symbolic methods developed in data-aware PM to address this problem. However, because of differences in formulation, these methods cannot be applied off-the-shelf.

To extract and transform unstructured data from textual documents, we will use LLMs. While the potential of NLP and NL Generation (NLG) for BPM has been recognized for over a decade [41, 42], practical implementation in PM tools remains limited. NLG has been used to explain event logs and verbalize BPMN models [41, 43], but without reasoning support. NLP techniques have also been used to extract and compare BPMN models with textual descriptions [44]. Recently, conversational interfaces for PM have been proposed, but LLMs are used as black boxes and reasoning is only partially supported [45]. Our context requires combining business-informed contextual learning with logical reasoning, dealing with the inherent uncertainty or conflict of textual information [46, 47, 48, 49]. To identify discrepancies between documents (still needing formal methods to be validated), we will take inspiration from recent results integrating Retrieval Augmented Generation (RAG) into LLMs, as well as methods that allow for structured reasoning beyond chain of thought (CoT) such as ReAct, and methodologies that allow for reasoning over constraints and formalisms [50, 51, 52, 53].

The alignment procedure will use formal methods to automatically compare symbolic representations of documents and guidelines. We will be inspired by SMT approaches for process-related logics [54, 55].

3. Proposed Methodology

To assess compliance in PA processes, this work proposes a hybrid, multiphase methodology that integrates recent advances in LLMs, RAG, and PM-inspired symbolic reasoning. Our objective is to develop a formal method, inspired by data-aware PM, to mine the relevant process-related information from the log data, and embed it into the specification documents, transform them into a symbolic format, and then use an alignment procedure to compute the discrepancy score between log-informed documents and the 'guidelines schema'.

Motivated by the research questions (RQ1–RQ4) outlined in the introduction, our approach unfolds across four methodological phases, each corresponding to key components of the envisioned pipeline.

Symbolic Formalization of Guidelines and Specifications (Related to RQ4). The first phase focuses on designing a symbolic schema to uniformly represent both contractual specification documents and regulatory guidelines. This schema must be expressive enough to capture business-relevant constraints (e.g., actor responsibilities, data dependencies, technological requirements and specifications) while remaining computationally tractable for formal reasoning.

Unlike traditional models in PM that describe control-flow behaviors or event sequences, this symbolic representation will need to encode higher-level normative statements and contextual dependencies often found in legal and specification texts. Special attention will be given to common structural patterns in PA documentation (e.g., templates, reference architectures, recurring legal phrases) and known ontologies. Given the nature of these documents, which often include not only temporal dependencies but also complex constraints involving multiple actors, data objects, and the relations among them, the schema must be expressive enough to capture both dynamic and structural aspects of compliance. To this end, propositional temporal logic like LTL [56] does not seem to be sufficient, but we aim to employ suitable logical formalisms that combine temporal reasoning (such as temporal logics) with conditions over data (as in first-order logic). In particular, decidable fragments of First-Order Linear Temporal Logic (FO-LTL), such as LTLfMT [54, 57], appear to be promising candidates, as they allow representing temporal obligations while retaining the ability to reason over data attributes and inter-entity dependencies in a computationally tractable way. This formalization lays the groundwork for subsequent alignment and supports modular reasoning on extracted content.

Information Extraction via LLMs and RAG (Related to RQ1, RQ2, RQ3). In the second phase, we develop a knowledge extraction pipeline based on LLMs, enhanced through RAG and CoT-inspired techniques. This component transforms diverse textual inputs—including specifications, technical requirements, and guidelines—into the structured symbolic schema defined in Phase 1. To ensure precision and relevance, we explore in-context learning and domain-specific fine-tuning of LLMs on business-aware data. Retrieval mechanisms are used to dynamically access relevant context (e.g., specifications, regulations, or log-related information) at inference time.

Furthermore, by exploring CoT and explanation-providing methods, as well as uncertainty and factuality-checking techniques, we aim to address challenges related to semantic fidelity (RQ2), ensuring that LLM outputs reflect the true intent of the source material. We also propose to integrate insights from event logs into this phase, effectively grounding the extracted information in real execution traces. This not only enhances the reliability of the LLM outputs (RQ3), but also enables the inclusion of factual process behavior in the alignment logic.

Symbolic Alignment via Formal Methods (Related to RQ4). In the third phase, we develop a formal alignment algorithm that compares the structured representation of specification documents against formalized regulatory schemas. Drawing from data-aware declarative conformance checking in PM, we design symbolic reasoning procedures, based on SAT/SMT solving to assess whether the information extracted from documents conforms to legal norms.

The algorithm computes a compliance score that quantifies the degree of alignment, while also identifying specific discrepancies and their types (e.g., missing constraints, conflicting obligations). This method should align two symbolic abstractions derived from unstructured content.

A central research question here (RQ4) is whether the pipeline can be effectively decoupled into modular steps (information retrieval, schema population, and alignment computation) or whether tighter integration is needed to preserve accuracy and traceability across components.

Validation and Evaluation. The final phase involves testing and validating the methodology using real specification documents and guidelines provided by partner PA entities (AMA and IGFEJ). In order to do so, we will need to (i) measure the precision and recall of LLM-based extraction methods; (ii) evaluate the interpretability and reliability of the computed alignment scores, and (iii) test the system's ability to handle complex, heterogeneous document types. Based on the empirical feedback we obtain on the aforementioned points (involving both automated and human-based evaluation), we plan to iteratively refine the schema, extraction pipeline, and alignment logic.

4. Example Scenarios in Public Administration

We now focus on the usual scenarios provided by our partner institutions, AMA and IGFEJ. AMA is the public institute responsible for promoting and developing administrative modernization in Portugal; its mission includes coordinating programs in regulatory simplification, electronic administration, and the delivery of public services. IGFEJ manages the financial asset and technological resources of the Portuguese Ministry of Justice; it plays a central role in ensuring the proper functioning of the judicial system across the national territory.

A concrete application of the proposed methodology arises in the *enterprise alignment* subprocesses present in the workflows of both institutions. In these subprocesses, related PA entities submit to these two institutions architectural scenarios with technological specifications that must be checked against normative guidelines before approval. On the one hand, at AMA the evaluation concerns transversal modernization and interoperability principles, ensuring accessibility and uniformity across public services. On the other hand, At IGFEJ specifications from justice-sector entities are assessed against sector-specific requirements, such as procurement rules, security standards, or data governance obligations in judicial infrastructures.

Currently, these evaluations rely on manual inspection of heterogeneous documentation, a process that is time-consuming and prone to inconsistencies. By applying our methodology, RAG-enhanced LLMs can extract relevant knowledge from submitted specifications and regulatory texts, while symbolic alignment procedures identify (mis)alignments. This hybrid approach supports AMA and IGFEJ evaluators by providing automatic and explainable compliance analysis.

5. Discussion and Conclusions

Challenges and Limitations. Despite its potential, the proposed methodology faces several challenges and limitations. First, LLMs struggle with long and complex documents, especially when legal and technical content must be interpreted together. Ensuring that critical constraints are accurately extracted without oversimplification or hallucination remains a key concern. Second, the retrieval component in RAG pipelines may fail to fetch relevant context or misalign retrieved knowledge with the user's query, compromising the accuracy of the structured output. Third, while symbolic reasoning offers formal guarantees, it relies on the completeness and correctness of the extracted schema, which may be affected by noise or ambiguity in the source texts. Furthermore, domain adaptation of LLMs to PA-specific language and legal discourse is still underdeveloped, and fine-tuning may be limited by the availability of high-quality labeled data. There are also scalability concerns: aligning large numbers of documents and lengthy regulations using formal methods can be computationally intensive. Finally, the interpretability and explainability of the alignment results must be addressed, as PA stakeholders require transparent justifications for any detected misalignments to support legal and operational decisions.

Final considerations. In summary, this work sets the stage for a novel research trajectory focused on combining LLMs, RAG techniques, and formal methods to address one of the most persistent challenges in public administration: ensuring process compliance in the face of complexity. It outlines a vision of more intelligent and automated public administrations, supported by cutting-edge AI and formal reasoning tools. By providing tools that assist in the automatic checking of compliance against legal norms, this work aspires to support public administrators in ensuring legality and improving accountability.

Acknowledgments

This work was partially supported by the 'OptiGov' project, with ref. n. 2024.07385.IACDC (DOI: 10.54499/2024.07385.IACDC), fully funded by the 'Plano de Recuperação e Resiliência' (PRR) under the investment 'RE-C05-i08 - Ciência Mais Digital' (measure 'RE-C05-i08.m04'), framed within the financing agreement signed between the 'Estrutura de Missão Recuperar Portugal' (EMRP) and Fundação para a Ciência e a Tecnologia, I.P. (FCT) as an intermediary beneficiary. This work was also partly supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia, I.P. (FCT), under projects UID/50021/2025 and UID/PRR/50021/2025.

References

[1] I. Kregel, B. Distel, A. Coners, Business process management culture in public administration and its determinants, Bus. Inf. Syst. Eng. 64 (2022) 201–221. URL: https://doi.org/10.1007/s12599-021-00713-z. doi:10.1007/S12599-021-00713-Z.

[2] G. Governatori, Z. Milosevic, S. Sadiq, Compliance checking between business processes and business contracts, in: International Enterprise Distributed Object Computing Conference, 2006.

[3] S. Sadiq, G. Governatori, Managing regulatory compliance in business processes, in: Handbook on Business Process Management, 2015.

[4] N. van Beest, H. Groefsema, A. Cryer, G. Governatori, S. Colombo Tosatto, H. Burke, Cross-instance regulatory compliance checking of business process event logs, IEEE Trans. Software Eng. 49 (2023).
[5] M. Alberti, F. Chesani, M. Gavanelli, E. Lamma, P. Mello, M. Montali, P. Torroni, Expressing and verifying business contracts with abductive logic programming, Int. J. Electron. Commer. (2008).
[6] W. van der Aalst, Process Mining, Springer, 2016.
[7] M. Dumas, M. La Rosa, J. Mendling, H. Reijers, Fundamentals of Business Process Management, Springer, 2018.
[8] M. Boltenhagen, T. Chatain, J. Carmona, Optimized SAT encoding of conformance checking artefacts, Computing 103 (2021) 29–50. URL: https://doi.org/10.1007/s00607-020-00831-8. doi:10.1007/S00607-020-00831-8.
[9] J. Ojeda, Conformance checking artefacts through weighted partial maxsat, Inf. Syst. 114 (2023) 102168. URL: https://doi.org/10.1016/j.is.2023.102168. doi:10.1016/J.IS.2023.102168.
[10] P. Felli, A. Gianola, M. Montali, A. Rivkin, S. Winkler, Data-aware conformance checking with smt, Inf. Syst. (2023).
[11] F. Mannhardt, M. de Leoni, H. Reijers, W. van der Aalst, Balanced multi-perspective checking of process conformance, Computing (2016).
[12] A. Gianola, Verification of Data-Aware Processes via Satisfiability Modulo Theories, volume 470 of *Lecture Notes in Business Information Processing*, Springer, 2023. URL: https://doi.org/10.1007/978-3-031-42746-6.
[13] P. Lewis, E. Perez, A. Piktus, F. Petroni, V. Karpukhin, N. Goyal, H. Küttler, M. Lewis, W. Yih, T. Rocktäschel, S. Riedel, D. Kiela, Retrieval-augmented generation for knowledge-intensive NLP tasks, in: Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/6b493230205f780e1bc26945df7481e5-Abstract.html.
[14] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei, Language models are few-shot learners, in: Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html.
[15] G. Team, P. Georgiev, V. I. Lei, R. Burnell, L. Bai, A. Gulati, G. Tanzer, D. Vincent, Z. Pan, S. Wang, et al., Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context, arXiv preprint arXiv:2403.05530 (2024).
[16] Y. Kuratov, A. Bulatov, P. Anokhin, I. Rodkin, D. Sorokin, A. Sorokin, M. Burtsev, Babilong: Testing the limits of llms with long context reasoning-in-a-haystack, Advances in Neural Information Processing Systems 37 (2024) 106519–106554.
[17] T. Kojima, S. S. Gu, M. Reid, Y. Matsuo, Y. Iwasawa, Large language models are zero-shot reasoners, Advances in neural information processing systems 35 (2022) 22199–22213.
[18] M. Besta, J. Barth, E. Schreiber, A. Kubicek, A. C. Catarino, R. Gerstenberger, P. Nyczyk, P. Iff, Y. Li, S. Houliston, T. Sternal, M. Copik, G. Kwasniewski, J. Müller, L. Flis, H. Eberhard, H. Niewiadomski, T. Hoefler, Reasoning language models: A blueprint, CoRR abs/2501.11223 (2025). URL: https://doi.org/10.48550/arXiv.2501.11223.
[19] DeepSeek-AI, D. Guo, D. Yang, H. Zhang, J. Song, R. Zhang, R. Xu, Q. Zhu, S. Ma, P. Wang, X. Bi, X. Zhang, X. Yu, Y. Wu, Z. F. Wu, Z. Gou, Z. Shao, Z. Li, Z. Gao, A. Liu, B. Xue, B. Wang, B. Wu, B. Feng, C. Lu, C. Zhao, C. Deng, C. Zhang, C. Ruan, D. Dai, D. Chen, D. Ji, E. Li, F. Lin, F. Dai, F. Luo, G. Hao, G. Chen, G. Li, H. Zhang, H. Bao, H. Xu, H. Wang, H. Ding, H. Xin, H. Gao, H. Qu, H. Li, J. Guo, J. Li, J. Wang, J. Chen, J. Yuan, J. Qiu, J. Li, J. L. Cai, J. Ni, J. Liang, J. Chen, K. Dong, K. Hu, K. Gao, K. Guan, K. Huang, K. Yu, L. Wang, L. Zhang, L. Zhao, L. Wang, L. Zhang, L. Xu, L. Xia, M. Zhang, M. Zhang, M. Tang, M. Li, M. Wang, M. Li, N. Tian, P. Huang, P. Zhang, Q. Wang, Q. Chen, Q. Du, R. Ge, R. Zhang, R. Pan, R. Wang, R. J. Chen, R. L. Jin, R. Chen, S. Lu,

S. Zhou, S. Chen, S. Ye, S. Wang, S. Yu, S. Zhou, S. Pan, S. S. Li, DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning, CoRR abs/2501.12948 (2025). URL: https://doi.org/10.48550/arXiv.2501.12948.
[20] W. Tang, V. Belle, Ltlbench: Towards benchmarks for evaluating temporal logic reasoning in large language models, CoRR abs/2407.05434 (2024). URL: https://doi.org/10.48550/arXiv.2407.05434.
[21] G. Feng, B. Zhang, Y. Gu, H. Ye, D. He, L. Wang, Towards revealing the mystery behind chain of thought: A theoretical perspective, in: A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine (Eds.), Advances in Neural Information Processing Systems, volume 36, Curran Associates, Inc., 2023, pp. 70757–70798. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/dfc310e81992d2e4cedc09ac47eff13e-Paper-Conference.pdf.
[22] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al., Chain-of-thought prompting elicits reasoning in large language models, Advances in neural information processing systems 35 (2022) 24824–24837.
[23] G. Izacard, P. Lewis, M. Lomeli, L. Hosseini, F. Petroni, T. Schick, J. Dwivedi-Yu, A. Joulin, S. Riedel, E. Grave, Few-shot learning with retrieval augmented language models, CoRR abs/2208.03299 (2022). URL: https://doi.org/10.48550/arXiv.2208.03299.
[24] B. Jin, J. Yoon, J. Han, S. Ö. Arik, Long-context LLMs meet RAG: Overcoming challenges for long inputs in rag, in: The Thirteenth International Conference on Learning Representations, 2024.
[25] F. Maggi, M. Dumas, L. García-Bañuelos, M. Montali, Discovering data-aware declarative process models from event logs, in: Proc. BPM, 2013.
[26] F. Maggi, A. Marrella, F. Patrizi, V. Skydanienko, Data-aware declarative process mining with SAT, ACM Trans. Intell. Syst. Technol. (2023).
[27] J. Casas-Ramos, S. Winkler, A. Gianola, M. Montali, M. Mucientes, M. Lama, Efficient conformance checking of rich data-aware declare specifications, in: Proceedings of BPM 2025 - 23rd International Conference on Business Process Management, volume 16044 of *Lecture Notes in Computer Science*, Springer, 2025, pp. 88–105. URL: https://doi.org/10.1007/978-3-032-02867-9_7. doi:10.1007/978-3-032-02867-9_7.
[28] A. Biere, M. Heule, H. van Maaren, T. Walsh, Handbook of Satisfiability, IOS Press, 2021.
[29] C. Barrett, R. Sebastiani, S. Seshia, C. Tinelli, Satisfiability modulo theories, in: Handbook of Satisfiability, 2021.
[30] W. van der Aalst, J. Carmona, Process Mining Handbook, 2022.
[31] M. Dumas, M. L. Rosa, J. Mendling, H. A. Reijers, Fundamentals of Business Process Management, Second Edition, Springer, 2018. URL: https://doi.org/10.1007/978-3-662-56509-4.
[32] M. Reichert, Process and data: Two sides of the same coin?, in: Proc. OTM Conferences, 2012.
[33] P. Felli, A. Gianola, M. Montali, A. Rivkin, S. Winkler, Cocomot: Conformance checking of multi-perspective processes via smt, in: Proc. BPM, 2021.
[34] P. Felli, A. Gianola, M. Montali, A. Rivkin, S. Winkler, Multi-perspective conformance checking of uncertain process traces: An smt-based approach, Eng. Appl. Artif. Intell. 126 (2023) 106895. URL: https://doi.org/10.1016/j.engappai.2023.106895.
[35] M. Hashmi, G. Governatori, H.-P. Lam, M. Wynn, Are we done with business process compliance: state of the art and challenges ahead, Knowl. Inf. Syst. (2018).
[36] H. Groefsema, N. van Beest, M. Aiello, A formal model for compliance verification of service compositions, IEEE Transactions on Services Computing (2018).
[37] D. Calvanese, S. Ghilardi, A. Gianola, M. Montali, A. Rivkin, Formal modeling and smt-based parameterized verification of data-aware bpmn, in: Proc. BPM, 2019.
[38] S. Ghilardi, A. Gianola, M. Montali, A. Rivkin, Petri net-based object-centric processes with read-only data, Inf. Sys. (2022).
[39] S. Ghilardi, A. Gianola, M. Montali, A. Rivkin, Safety verification and universal invariants for relational action bases, in: Proc. IJCAI, 2023.
[40] H. A. López, S. Debois, T. Slaats, T. T. Hildebrandt, Business process compliance using reference models of law, in: H. Wehrheim, J. Cabot (Eds.), FProceedings of FASE 2020, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2020, volume 12076

of *Lecture Notes in Computer Science*, Springer, 2020, pp. 378–399. URL: https://doi.org/10.1007/978-3-030-45234-6_19.
[41] H. Leopold, Natural Language in Business Process Models - Theoretical Foundations, Techniques, and Applications, Springer, 2013.
[42] W. van der Aa, J. Carmona, H. Leopold, J. Mendling, L. Padró, Challenges and opportunities of applying natural language processing in business process management, in: Proc. COLING, 2018.
[43] Y. Fontenla-Seco, M. Lama, V. González-Salvado, C. Peña-Gil, A. Bugarín Diz, A framework for the automatic description of healthcare processes in natural language: Application in an aortic stenosis integrated care process, J. Biomed. Informatics 128 (2022).
[44] H. van der Aa, H. Leopold, H. Reijers, Detecting inconsistencies between process models and textual descriptions, in: Proc. BPM, 2015.
[45] Y. Fontenla-Seco, S. Winkler, A. Gianola, M. Montali, M. Lama Penín, A. Bugarín Diz, The droid you're looking for: C-4pm, a conversational agent for declarative process mining, in: Proc. BPM, 2023.
[46] Z. Su, J. Zhang, X. Qu, T. Zhu, Y. Li, J. Sun, J. Li, M. Zhang, Y. Cheng, Conflictbank: A benchmark for evaluating the influence of knowledge conflicts in llms, Advances in Neural Information Processing Systems 37 (2024) 103242–103268.
[47] F. Wang, X. Wan, R. Sun, J. Chen, S. Ö. Arik, Astute RAG: overcoming imperfect retrieval augmentation and knowledge conflicts for large language models, arXiv preprint arXiv:2410.07176 (2024).
[48] J. Vasilakes, C. Zerva, M. Miwa, S. Ananiadou, Learning disentangled representations of negation and uncertainty, in: Proceedings of ACL, 2022.
[49] J. J. Jia, Z. Yuan, J. Pan, P. McNamara, D. Chen, Decision-making behavior evaluation framework for llms under uncertain context, Advances in Neural Information Processing Systems 37 (2024) 113360–113382.
[50] Q. Dong, Y. Liu, Q. Ai, Z. Wu, H. Li, Y. Liu, S. Wang, D. Yin, S. Ma, Unsupervised large language model alignment for information retrieval via contrastive feedback, in: Proc. ACM SIGIR, 2024.
[51] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, Y. Cao, React: Synergizing reasoning and acting in language models, in: Proc. of ICLR, 2023.
[52] J. Ahn, R. Verma, R. Lou, D. Liu, R. Zhang, W. Yin, Large language models for mathematical reasoning: Progresses and challenges, in: Proc. EACL (SRW), 2024.
[53] J. Zhou, C. Staats, W. Li, C. Szegedy, K. Weinberger, Y. Wu, Don't trust: Verify–grounding llm quantitative reasoning with autoformalization, in: Proc. ICLR, 2024.
[54] L. Geatti, A. Gianola, N. Gigante, Linear temporal logic modulo theories over finite traces, in: Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, ijcai.org, 2022, pp. 2641–2647. URL: https://doi.org/10.24963/ijcai.2022/366.
[55] A. Gianola, M. Montali, S. Winkler, Linear-time verification of data-aware processes modulo theories via covers and automata, in: Proc. AAAI, 2024.
[56] A. Pnueli, The temporal logic of programs, in: 18th Annual Symposium on Foundations of Computer Science, Providence, Rhode Island, USA, 31 October - 1 November 1977, IEEE Computer Society, 1977, pp. 46–57. URL: https://doi.org/10.1109/SFCS.1977.32.
[57] L. Geatti, A. Gianola, N. Gigante, S. Winkler, Decidable fragments of ltl_f modulo theories, in: Proceedings of ECAI 2023 - 26th European Conference on Artificial Intelligence, volume 372 of *Frontiers in Artificial Intelligence and Applications*, IOS Press, 2023, pp. 811–818. URL: https://doi.org/10.3233/FAIA230348.