IRIS AperTO

UNIVERSITÀ DEGLI STUDI DI TORINO

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Emerging challenges in legal informatics from machine learning to LLMs - Preface to the proceedings of the 1st PLC workshop

This is the author's manuscript

Original Citation:

Availability:

This version is available http://hdl.handle.net/2318/2122301 since 2026-02-09T14:37:55Z

Publisher:

CEUR-WS

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

21 November 2026

Emerging challenges in legal informatics from machine learning to LLMs - Preface to the proceedings of the 1st PLC workshop

Laura Genga[1], Hugo A. López[2] and Emilio Sulis[3,*]

[1]School of Industrial Engineering - Eindhoven University of Technology, Eindhoven, The Netherlands
[2]Technical University of Denmark, Kgs. Lyngby, Denmark
[3]Computer Science Department, University of Torino, Italy

Abstract

The integration of Artificial Intelligence techniques, machine learning and large language models into legal informatics offers innovative potential, from enhancing legal research efficiency to supporting legal reasoning. These advancements introduce significant challenges, including issues related to data privacy, bias in legal datasets, and the interpretability of complex algorithms in legal contexts. Emerging challenges involve reliability, fairness, and ethical considerations in AI-driven legal applications. The research contributions presented at a recent workshop on Processes, Law and Compliance aim to deepen these issues for the development of AI applications in the field of legal informatics.

Keywords

Legal Machine Learning Challenges, AI-Driven Legal Informatics, Legal Event Logs,

1. Legal Informatics

Research in legal informatics has grown significantly in recent decades, driven in large part by the proliferation of advanced information systems that are increasingly capable of recording, organizing, and analyzing vast amounts of data generated by legal processes [1]. These systems allow for the systematic analysis of the different steps of interest in a legal process, including the storage of legal documents as texts of tenders, court judgments, public procurements.

Artificial Intelligence (AI) techniques provide a valuable tool for analyzing legal data to obtain valuable information to support the work of both government agencies and private companies [2]. Research in legal informatics has been directly linked to the applications of AI [3]. Examples of AI-driven systems include the intersection of Machine Learning (ML), Process Mining (PM), and Natural Language Processing (NLP) techniques [4, 5, 6]. Nevertheless, AI systems are typically considered black boxes, i.e. posing explainability issues [7]. In fact, AI and data-driven techniques do not provide full transparency of how processes and law intersect [8].

Within this framework of relevant opportunities and critical issues, research is facing new challenges in the legislative and information technology application domain. The proceedings of the workshop Processes, Laws, and Compliance confirm this heterogeneity of aspects. The current section summarizes the main areas of interest, while the next describes the organization and program of the workshop.

Automated process-oriented analysis and digital law The intersection of technology and law has given rise to a research area focused on the automated, process-oriented analysis of legal systems and digital law. The recent discipline of PM combines data science and process management to analyze and optimize real-world business processes based on event log data [9]. This emerging field explores how data-driven and computational approaches can enhance the understanding, modeling, and application

PLC - Processes, Laws and Compliance workshop, in conjunction with ICPM 2024, October 14, 2024, Lyngby, Denmark
*Corresponding author.
l.genga@tue.nl (L. Genga); hulo@dtu.dk (H. A. López); emilio.sulis@unito.it (E. Sulis)
0000-0002-9421-8566 (L. Genga); 0000-0001-5162-7936 (H. A. López); 0000-0003-1746-3733 (E. Sulis)

© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

of complex legal processes. Research investigated traditional legal workflows and facilitated compliance improvement, predictive analytics, and performance analysis [10]. Central to this research are methods that automate legal processes in legal systems, or consider compliance-oriented designs that align with codified standards and judicial expectations [11]. In fact, legal documents such as laws, guidelines, standards contain information about the underlying legal processes. PM techniques allow for the discovery of process behaviour within legal artifacts. By analyzing variations in how laws and legal standards are applied or interpreted, these studies provide valuable insights on legal workflows and expose inefficiencies. In addition, variant analysis of different legal process executions allows a nuanced view of discrepancies and divergences that can inform better policy development and procedural refinement [12].

Compliance and formal representation of laws Author Agreements

A relevant research area involves the compliance between formal representations of laws and their implementation in practical settings [13]. Formal representations, such as digital encodings of legal rules, can be compared with actual case applications to identify and resolve compliance issues [14]. This line of research enables the development of systems that are compliant by design, where digital platforms are pre-configured to follow legal norms [15]. Approaches as "Rules as Code" involve encoding regulations and standards directly into software, enabling automated systems that inherently comply with legal requirements [16]. The paradigm of compliance-by-design has the potential to transform industries that depend on regulatory adherence, ensuring that digital systems can automatically align with complex and evolving legal standards.

Advances in legal modeling and Natural Language Processing The complexity of law also requires advanced techniques for modeling legal norms [17], while Law includes a set of rules, exceptions, and interpretations. Research has focused on developing models that capture the conditional, hierarchical, and interpretative nature of legal norms, which aim to bridge the gap between rigid digital structures and the flexible, context-sensitive needs of legal reasoning. NLP techniques enable machines to interpret and extract meaning from legal texts facilitating practical applications, from automating document review to extracting legal clauses, allowing legal professionals to manage large volumes of documents efficiently [18]. Moreover, NLP techniques are increasingly being used to support legal reasoning, offering assistance in interpreting statutory language, analyzing court opinions, and comparing legal standards across jurisdictions [19].

Visualization and relations in legal data engineering With the growing volume of digital legal documents and the complexity of legal processes, visualization and simplification techniques have gained prominence. Researchers are developing tools that make legal processes more understandable and accessible, by adopting user-friendly formats that can be easily interpreted by legal practioners [20]. Such tools play a crucial role in enhancing transparency, aiding public understanding, and supporting effective decision-making within legal settings. Furthermore, advancements in information retrieval and legal knowledge extraction enhances capabilities for finding relevant legal references, similar documents, and previous cases, creating a more interconnected legal knowledge base [21].

2. Outline and rationale for the PLC workshop

The first international workshop Processes, Laws and Compliance (PLC) intended to provide a forum to facilitate the exchange of research findings and ideas on data-driven and process-oriented techniques and practices in the legal domain, fostering collaboration between interdisciplinary experts, researchers, and practitioners working in IT and law. The workshop has been held in conjunction with the 6th International Conference on Process Mining (ICPM 2024), Technical University of Denmark, Lyngby, October 14, 2024.

The program of the first edition of the PLC workshop included oral presentations of six research papers out of nine contributions received and accepted at the end of a peer-reviewed process, as well as three showcase contributions. The workshop has been opened by an invited talk by Prof. Dr. Stefanie Rinderle-Ma (Technical University of Munich), titled "How can Large Language Models support process mining and compliance checking?". The keynote speech explored how LLMs can enhance PM and compliance checking by automating the interpretation of complex regulations, extracting insights from unstructured documents, and identifying patterns or deviations in process logs, thereby improving accuracy and efficiency in compliance verification.

In the afternoon session of the workshop, participants focused into central and highly relevant topics for the future of the field. A first topic, "Interdisciplinary Challenges on Processes, Law, and Compliance" prompted a discussion on the emerging difficulties in integrating business processes, legal frameworks, and compliance requirements, emphasizing the need for collaboration between experts from various disciplines. Moreover, a session on "Opportunities and Ideation: Possible Futures and Affordances in Digital Compliance" involved participants to explore future possibilities, innovations, and practical applications in digital compliance, fostering a forward-thinking approach to technological and regulatory advancements.

Research themes. The discussion focused on the following four themes and related research questions:

- *Theme 1: AI for the legal sector. Benefits and challenges.* How can we rethink the legal sector by leveraging process and data-driven techniques? What are the legal processes that are calling for support from digital technologies, and how could data and process-driven techniques help to improve them? What are the enablers, and what are the challenges of applying these techniques to this domain? Are there relevant case studies already?
- *Theme 2: Risk and Compliance Formalization.* How can organizations ensure automated processes are both efficient and compliant, especially when regulations require human judgment or interpretation? How do legal and process jargon align? For instance, does a legal violation correspond to violations in a process mining sense? What is undesired behavior and what is the difference from undefined behavior in laws? What is the legal implication of concepts such as deviations, workarounds, or anomalies for compliance?
- *Theme 3: Adoption of Compliance Frameworks.* Automating compliance checks through BPM systems is desirable but challenging. How should a Compliance Checking Framework be designed and implemented? Compliance by Design (CdB) or Compliance by Auditing? Is the dream of CbD attainable? What do we need to make it happen? If not, what are the challenges in Audits? What are the factors that impede the adoption of compliance technologies in the industry?
- *Theme 4 The Human Factor in Compliance.* Compliance technologies aim to support legal specialist in their certification and auditing techniques. What are the gaps in: i. Generating (mathematical) specifications from legal behavior that correspond to what is expressed in a law? ii. Explaining the output of process/data-driven technologies in a way that corresponds to legal argumentation for compliance officers? What are the requirements to implement tools for non-technical users (e.g., low-code tools) and how far are we?

We thank all the contributing speakers, the members of our Program Committee for timely providing their reviews, and the ICPM Workshop chairs Andrea Delgado and Tijs Slaats for their support.

Workshop Organizers The PLC workshop has been organized by the following co-chairs: Laura Genga (Technical University of Eindhoven, Netherlands), Hugo A. López (Technical University of Denmark, Denmark), and Emilio Sulis (University of Turin, Italy).

Program Committee The Program Commitee of the workshop that also carried out the reviews of the articles consisted of the following researchers:

- Davide Audrito (University of Bologna)
- Chiara Di Francescomarino (University of Trento)
- Chiara Gallese (University of Turin)
- Roberto Nai (University of Torino)
- Barbara Pernici (Politecnico di Milano)
- Stefanie Rinderle-Ma (Technical University of Munich)
- Livio Robaldo (University of Swansea)
- Massimiliano Ronzani (FBK Trento)
- Giovanni Siragusa (University of Turin)
- Han Van der Aa (University of Vienna)
- Andrea Vandin (Sant'Anna School of Advanced Studies, Pisa)
- Karolin Winter (Technical University of Eindhoven)

Workshop Website Further information on the topics, schedule, keynote presentation, and further developments of the PLC Workshop can be found at the website: https://sites.google.com/view/plc-workshop-2024/home.

Workshop Proceedings This volume includes post-conference papers from the PLC workshop. In particular, the authors of the six research works agreed to include their paper in the workshop proceedings. In addition, we invited an author to present a more extensive discussion of his showcase.

References

[1] H. Surden, Machine learning and law: An overview, Research Handbook on Big Data Law (2021) 171–184.

[2] F. J. Bex, Ai, law and beyond. a transdisciplinary ecosystem for the future of ai & law, Artificial Intelligence and Law (2024) 1–18.

[3] T. J. M. Bench-Capon, M. Araszkiewicz, K. D. Ashley, K. Atkinson, F. Bex, F. Borges, D. Bourcier, P. Bourgine, J. G. Conrad, E. Francesconi, T. F. Gordon, G. Governatori, J. L. Leidner, D. D. Lewis, R. P. Loui, L. T. McCarty, H. Prakken, F. Schilder, E. Schweighofer, P. Thompson, A. Tyrrell, B. Verheij, D. N. Walton, A. Z. Wyner, A history of AI and law in 50 papers: 25 years of the international conference on AI and law, Artif. Intell. Law 20 (2012) 215–319. doi:10.1007/S10506-012-9131-X.

[4] L. Robaldo, S. Batsakis, R. Calegari, F. Calimeri, M. Fujita, G. Governatori, M. C. Morelli, F. Pacenza, G. Pisano, K. Satoh, I. Tachmazidis, J. Zangari, Compliance checking on first-order knowledge with conflicting and compensatory norms: a comparison among currently available technologies, Artif. Intell. Law 32 (2024) 505–555. doi:10.1007/S10506-023-09360-Z.

[5] M. Medvedeva, M. Vols, M. Wieling, Using machine learning to predict decisions of the european court of human rights, Artif. Intell. Law 28 (2020) 237–266. doi:10.1007/S10506-019-09255-Y.

[6] R. Nai, E. Sulis, I. Fatima, R. Meo, Large language models and recommendation systems: A proof-of-concept study on public procurements, in: Natural Language Processing and Information Systems - 29th International Conference on Applications of Natural Language to Information Systems, NLDB 2024, Turin, Italy, June 25-27, 2024, Proceedings, Part II, 2024, pp. 280–290. doi:10.1007/978-3-031-70242-6_27.

[7] R. Meo, R. Nai, E. Sulis, Explainable, interpretable, trustworthy, responsible, ethical, fair, verifiable AI... what's next?, in: S. Chiusano, T. Cerquitelli, R. Wrembel (Eds.), Advances in Databases and Information Systems - 26th European Conference, ADBIS 2022, Turin, Italy, September 5-8, 2022, Proceedings, volume 13389 of *Lecture Notes in Computer Science*, Springer, 2022, pp. 25–34. doi:10.1007/978-3-031-15740-0_3.

[8] A. Bibal, M. Lognoul, A. de Streel, B. Frénay, Legal requirements on explainability in machine learning, Artif. Intell. Law 29 (2021) 149–169. doi:10.1007/S10506-020-09270-4.

[9] W. M. P. van der Aalst, Process Mining - Data Science in Action, Second Edition, Springer, 2016. doi:10.1007/978-3-662-49851-4.

[10] R. Nai, E. Sulis, L. Genga, Automated analysis with event log enrichment of the european public procurement processes, in: T. P. Sales, J. Araújo, J. Borbinha, G. Guizzardi (Eds.), Advances in Conceptual Modeling - ER 2023 Workshops, Lisbon, Portugal, November 6-9, 2023, Proceedings, volume 14319 of *LNCS*, Springer, 2023, pp. 178–188. doi:10.1007/978-3-031-47112-4_17.

[11] R. Nai, R. Meo, G. Morina, P. Pasteris, Public tenders, complaints, machine learning and recommender systems: a case study in public administration, Comput. Law Secur. Rev. 51 (2023) 105887. doi:10.1016/J.CLSR.2023.105887.

[12] A. J. Unger, J. F. dos Santos Neto, M. Fantinato, S. M. Peres, J. Trecenti, R. Hirota, Process mining-enabled jurimetrics: analysis of a brazilian court's judicial performance in the business law processing, in: J. Maranhão, A. Z. Wyner (Eds.), ICAIL '21: Eighteenth International Conference for Artificial Intelligence and Law, São Paulo Brazil, June 21 - 25, 2021, ACM, 2021, pp. 240–244. doi:10.1145/3462757.3466137.

[13] H. A. López, S. Debois, T. Slaats, T. T. Hildebrandt, Business process compliance using reference models of law, in: H. Wehrheim, J. Cabot (Eds.), Fundamental Approaches to Software Engineering - 23rd International Conference, FASE 2020, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2020, Dublin, Ireland, April 25-30, 2020, Proceedings, volume 12076 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 378–399. doi:10.1007/978-3-030-45234-6_19.

[14] I. A. Amantea, L. Robaldo, E. Sulis, G. Boella, G. Governatori, Semi-automated checking for regulatory compliance in e-health, in: 25th International Enterprise Distributed Object Computing Workshop, EDOC Workshop 2021, Gold Coast, Australia, October 25-29, 2021, IEEE, 2021, pp. 318–325. doi:10.1109/EDOCW52865.2021.00063.

[15] S. Debois, H. A. López, T. Slaats, A. A. Andaloussi, T. T. Hildebrandt, Chain of events: Modular process models for the law, in: B. Dongol, E. Troubitsyna (Eds.), Integrated Formal Methods - 16th International Conference, IFM 2020, Lugano, Switzerland, November 16-20, 2020, Proceedings, volume 12546 of *Lecture Notes in Computer Science*, Springer, 2020, pp. 368–386. doi:10.1007/978-3-030-63461-2_20.

[16] T. Athan, G. Governatori, M. Palmirani, A. Paschke, A. Z. Wyner, Legalruleml: Design principles and foundations, in: W. Faber, A. Paschke (Eds.), Reasoning Web. Web Logic Rules - 11th International Summer School 2015, Berlin, Germany, July 31 - August 4, 2015, Tutorial Lectures, volume 9203 of *Lecture Notes in Computer Science*, Springer, 2015, pp. 151–188. doi:10.1007/978-3-319-21768-0_6.

[17] E. Sulis, L. D. Caro, R. Nanda, Introduction for computer law and security review: special issue "knowledge management for law", Comput. Law Secur. Rev. 52 (2024) 105949. doi:10.1016/J.CLSR.2024.105949.

[18] F. Yu, L. Quartey, F. Schilder, Exploring the effectiveness of prompt engineering for legal reasoning tasks, in: A. Rogers, J. L. Boyd-Graber, N. Okazaki (Eds.), Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023, Association for Computational Linguistics, 2023, pp. 13582–13596. doi:10.18653/V1/2023.FINDINGS-ACL.858.

[19] J. Collenette, K. Atkinson, T. J. M. Bench-Capon, Explainable AI tools for legal reasoning about cases: A study on the european court of human rights, Artif. Intell. 317 (2023) 103861. doi:10.1016/J.ARTINT.2023.103861.

[20] M. Hagan, Legal Design as a Thing: A Theory of Change and a Set of Methods to Craft a Human-Centered Legal System, Design Issues 36 (2020) 3–15. doi:10.1162/desi_a_00600.

[21] S. Castano, A. Ferrara, E. Furiosi, S. Montanelli, S. Picascia, D. Riva, C. Stefanetti, Enforcing legal information extraction through context-aware techniques: The ASKE approach, Comput. Law Secur. Rev. 52 (2024) 105903. doi:10.1016/J.CLSR.2023.105903.