Trustworthy Artificial Intelligence and Process Mining: Challenges and Opportunities

Andrew Pery[1], Majid Rafiei[2] ✉, Michael Simon[3], and Wil M.P. van der Aalst[2]

[1] ABBYY, Ottawa, Canada
[2] Chair of Process and Data Science, RWTH Aachen University, Aachen, Germany
[3] XPAN Law Partners, Boston, USA

Abstract. The premise of this paper is that compliance with Trustworthy AI governance best practices and regulatory frameworks is an inherently fragmented process spanning across diverse organizational units, external stakeholders, and systems of record, resulting in process uncertainties and in compliance gaps that may expose organizations to reputational and regulatory risks. Moreover, there are complexities associated with meeting the specific dimensions of Trustworthy AI best practices such as data governance, conformance testing, quality assurance of AI model behaviors, transparency, accountability, and confidentiality requirements. These processes involve multiple steps, hand-offs, re-works, and human-in-the-loop oversight. In this paper, we demonstrate that process mining can provide a useful framework for gaining fact-based visibility to AI compliance process execution, surfacing compliance bottlenecks, and providing for an automated approach to analyze, remediate and monitor uncertainty in AI regulatory compliance processes.

Keywords: AI ethics · Fairness · Artificial intelligence · Trust mining · Process mining

1 Introduction

AI-based technologies are becoming pervasive, impacting virtually every facet of our lives. While AI has a lot of promise, not all of its impacts are good. There is growing evidence that AI models can embed human and societal biases and deploy them at scale. As such, the ever-increasing growth of AI highlights the vital importance of balancing AI utility with the fairness of outcomes, thereby engender a culture of trustworthy AI. Fairness is the foundation for Trustworthy AI. Intuitively, fairness seems like a simple concept. However, it embodies consideration of a number of dimensions, such as trade-offs between algorithmic accuracy versus human values, demographic parity versus policy outcomes and power-focused questions such as who gets to decide what is fair.

These are vexing challenges for AI developers, policy-makers and consumers alike. For AI developers, clarity of what constitutes AI fairness is a key consideration given the juxtaposition of ethical, legal, and reputational issues. For policy-makers and regulators, the challenge is how to promote innovation while

protecting consumers from the harmful impacts of AI. For consumers of AI, its about trustworthiness, whether they can rely upon AI outputs to be accurate and transparent, with safeguards in place to protect them from adverse outcomes.

This paper explores the challenges and opportunities associated with fostering a culture of Trustworthy AI, with particular focus on: (1) The current state of Trustworthy AI, including a survey of key industry and standards organization initiatives with emphasis on the proposed EU Artificial Intelligence Act, (2) The relationship between Trustworthy AI and Responsible Data Science (RDS), and (3) Contribution of trust aware process mining to facilitate a data-driven analytical framework to surface uncertainties, variabilities, and vulnerabilities in Trustworthy AI compliance processes.

The remainder of the paper is organized as follows. In Section 2, we define the contours of Trustworthy AI principles. In Section 3, we explore the proposed EU Artificial Intelligence Act (AIA) that intends to operationalize and implement rigorous risk-based prescriptive processes for ensuring a culture of Trustworthy AI. In Section 4, we map the relationship between RDS and Trustworthy AI, including a discussion of challenges associated with contextualizing AI fairness as a foundation for Trustworthy AI. In Section 5, we discuss the applications and benefits of process mining as an important tool to enable organizations to make data-driven decisions relating to the obligations and conformance requirements inherent in the proposed EU AI regulation.

2 Trustworthy AI

Surveys reveal an undercurrent of pervasive distrust of AI systems. Cathy O'Neil, a leading advocate for AI algorithmic fairness, highlighted three main reasons behind consumer distrust of AI: *opacity*, *scale*, and *damage* [12]. Fairness is the foundation for trustworthy AI. It is the connective tissue that binds together the principles of ethical use, interpretability, transparency, accountability, and confidentiality that engenders trust and promotes the use of AI for social good. Trustworthy AI is a governance framework designed to mitigate potential adverse impacts on consumers as AI is poised to profoundly and indelibly change our lives. As mentioned in [17], Trustworthy AI is changing the dynamic between user and system into a relationship.

2.1 Achieving Trust in AI

Trustworthy AI starts with human agency and autonomy. Trust in AI systems is enhanced when there is a human-in-the-loop who monitors the overall performance of AI systems and when circumstances dictate, remediates potential adverse outcomes. Trust in AI is strengthened by giving users the ability to make informed decisions about the impact of AI on their personal and economic well-being.

AI is perceived by consumers to be *a black box*. Data inputs to the AI systems, their learning models, and how they arrive at decisions are neither visible, nor

understood by consumers. Furthermore, many AI developers defensively protect their algorithms as proprietary and a competitive differentiator. *Interpretability* and *explainability* of AI are two important elements that strengthen trust in AI. Interpretability of AI provides insight into the cause and effect between inputs and outputs of an AI system and how AI predicts outcomes. Explainability of AI goes one step further by providing users with not only insight into how AI models work but also traceability of AI decisions and documentation relating to the process of data gathering, labeling, and methods used for training AI algorithms.

Consumers have limited recourse to hold AI developers accountable for the adverse impacts of AI systems. While there is sectoral legislation, e.g., Section 5 of the FTC (Federal Trade Commission) Act[4], available for consumers to remedy disparate treatment attributable to AI systems it is an onerous process to prevail. Moreover, for the disparate impact, the burden of proof requires statistical analysis that a protected class is treated differently from others, which is hardly something that would be accessible to average consumers. For these reasons, accountability, including redress mechanisms in the event of demonstrated harmful impact need to be addressed to achieve trust in AI.

2.2 The Emergence of Trustworthy AI Principles

We can see efforts being made, to varying degrees, that recognize and deal with issues relating to trust in AI by the data sciences community (see Section 4), standards organizations, e.g., IEEE [16], NIST (National Institute of Standards and Technology) [13], and by public sector organizations.

In 2019, OECD member countries adopted OECD Council Recommendation on Artificial Intelligence[5] consisting of five principles of human centered values of fairness of AI, inclusive investments in AI, transparency, accountability, and robustness of AI systems. The OECD recommendations were subsequently endorsed by the G20 with particular reference to the view that the "digital society must be built on trust among all stakeholders including governments, civil society, international organizations, academics, and businesses through sharing common values and principles including equality, justice, transparency, and accountability taking into account the global economy and interoperability".

While Trustworthy AI principles serve as a helpful framework, they are just that. Adherence to Trustworthy AI is fragmented at best and they lack effective enforcement mechanisms to safeguard against potentially harmful impacts. For this reason, the momentum has shifted towards the regulation of AI: "The calls for modest regulation that lets industry take the lead are part of a failed regulatory philosophy, one that saw its natural experiment over the past several decades come up lacking. AI is too important and too promising to be governed in a hands-off fashion, waiting for problems to develop and then trying to fix them after the fact".[6]

[4] https://www.federalreserve.gov/boarddocs/supmanual/cch/ftca.pdf

[5] https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[6] https://www.brookings.edu/research/ai-needs-more-regulation-not-less/

3 The Proposed EU Regulation of AI

On April 20, 2021 the European Commission released the proposal for the regulation of artificial intelligence[7], the ambition of which is to balance the socio-economic benefits of AI and new risks or negative consequences for individuals or society. The proposed Artificial Intelligence Act (AIA) takes a risk-based approach to regulate AI by fostering an "ecosystem of trust that should give citizens the confidence to take up AI applications and give companies and public organisations the legal certainty to innovate using AI". In the following, we demonstrate five governing principles for trustworthy AI proposed by AIA.

3.1 Scope of the Proposed Regulation

The proposed AIA applies to all providers, i.e., natural or legal persons, public authorities, agencies, or any other body that develops an AI system, that places or makes available on the market or puts into service AI systems or services in the EU (cf. Article 3). The AIA also assigns responsibility to users, importers, distributors, and operators who make use of or make substantial modifications to the functionality and performance of AI systems (cf. Article 26-29). The geographic scope for the AIA will operate irrespective of whether such providers are established in the EU or a third country, and so will cover where the system users are in the EU or the output of the systems is used in the EU (cf. Article 2). AI systems under the regulation encompass a wide range of methods and algorithms including supervised, unsupervised, and reinforcement machine learning for a given set of human-defined objectives that generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with (cf. Article 3).

3.2 Risk-Based Approach

The foundation of the AIA is a risk-based approach that classifies AI systems into three categories based on a combination of factors that include the intended purpose, the number of impacted persons, and the potential risk of harms (cf. Article 5-7):

- Prohibited AI: Systems that use subliminal techniques that cause physiological or psychological harm, exploit vulnerable groups, effectuate social scoring by public authorities that may result in discrimination or unfavorable treatment, and remote biometric systems used by law enforcement in public spaces (subject to well-defined exceptions) (cf. Article 5).
- High Risk: Annex III provides a list of systems that are used in critical infrastructures, educational or vocational training, human resources, essential private and public services, law enforcement, migration, asylum and border control management, and administration of justice and democratic processes (cf. Article 7).

[7] https://ec.europa.eu/commission/presscorner/detail/en/ip_21_1682

– Low Risk: While not explicitly named (we use the term *low risk* of our own choosing), by default, all systems not categorized as *prohibited* or *high-risk*. Providers of such systems are encouraged to institute responsible use of AI best practices on a voluntary basis (cf. Article 69).

3.3 Promote Fair and Trustworthy AI Best Practices

The AIA sets forth a comprehensive legislative mandate to ensure fairness in the application of AI systems that safeguards fundamental human values and promotes socio-economic rights. Some of these mandates are as follows: obligation on providers to implement appropriate risk management measures throughout the entire lifecycle of AI systems (cf. Article 9), rigorous data governance processes (cf. Article 10), technical documentation, and record-keeping processes to enable monitoring of compliance (cf. Article 11-12), transparency that enables full interpretation of outputs (cf. Article 13), and Human-in-the-loop oversight (cf. Article 14).

3.4 Transparency and Accountability

According to the AIA, providers of AI systems will be required to implement a range of processes to ensure full transparency into and accountability for AI systems (cf. Article 19-23) such as (1) conformity assessment and certification processes, (2) auditability, including accessible event logs, and (3) Explainability, potentially to coordinate with the human-in-the-loop for adjudication and remediation.

3.5 Enforcement

The AIA incorporates an onerous enforcement mechanism that even surpasses the fines under the GDPR (cf. Article 71). Some examples are as follows: up to €10m or 2% of the total worldwide annual turnover for the supply of incorrect, incomplete or misleading information to the authorities, up to €20m or 4% of the total worldwide annual turnover for non-compliance with any other AIA requirement or obligation, and up to €30m or 6% of the total worldwide annual turnover for violations of prohibited practices.

While the proposed AIA is far from ratification and still subject to vigorous debate within the EU Parliament and Council, the momentum towards its adoption is inevitable. Like the GDPR, the AIA will serve as a model for other jurisdictions that will seek to finally exert control over what has been the unregulated, hyperbolic growth of AI across the globe.

4 Responsible Data Science and Trustworthy AI

Responsible Data Science (RDS) is a discipline that is influential in shaping Trustworthy AI best practices. RDS refers to the collection of techniques and

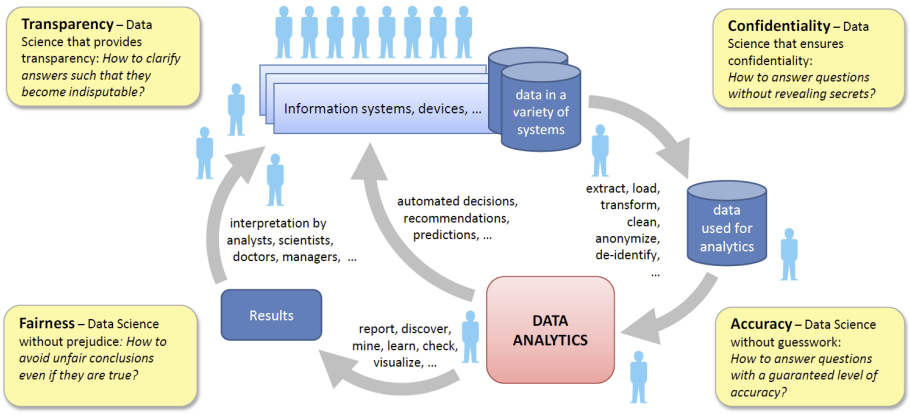

Fig. 1: The data science pipeline facing the four FACT challenges [2].

approaches trying to reap the benefits of data science and big data while ensuring *fairness*, *accuracy*, *confidentiality* and *transparency* [2]. To minimize adverse AI outcomes of AI the role of RDS is to: (1) Avoid unfair conclusions even if they are true, i.e., the fairness principle, (2) Answer questions with a guaranteed level of accuracy, i.e., the accuracy principle, (3) Answer questions without revealing secrets, i.e., the confidentiality principle, and (4) Clarify Answers such that they become indisputable, i.e., the transparency principle.

RDS applies a methodology throughout the entire life cycle of information to support trustworthy AI best practices by applying these four principles of fairness, accuracy, confidentiality, and transparency to the *data science pipeline* resulting in rigorous data governance as illustrated in Figure 1.

RDS delivers a robust framework for the ethical design of AI systems that addresses the following key areas: (1) Unbiased outcomes through the application of appropriate fairness constraints to the training data, (2) Algorithmic outcomes interpreted in a manner that is meaningful to end users, (3) Resilience in how AI systems deliver accurate results and respond to change in inputs, (4) Accountability for the system's outcomes, and (5) Safeguarding the confidentiality of training data through privacy enhancing measures. However, providing each aspect of RDS has its own challenges from contextualizing the aspect to implementing it in data science and AI systems. In [6], the authors describe the challenges regarding the *confidentiality* aspect for *process mining* which combines process and data science. In the following, we provide the challenges regarding the *fairness* aspect.

4.1 Contextualizing Fairness in AI Systems: Challenges

The idea of *fairness* is somewhat amorphous. At its highest level of abstraction, fairness is a normative concept that comes from our conscience. Dator defines a fair system as follows: "What is fairness then? We all have desires and we

want people to treat us according to those desires. We also know that people around us have similar desires and want to be treated accordingly. Fairness is closely related to fair play so it seems logical to conclude that a fair system is a system where everybody is treated in a similar way" [4]. There are a number of challenges associated with contextualizing and applying such a high level of abstraction to a more concrete algorithmic AI fairness framework.

First, fairness may be influenced by cultural, sociological, economic, and legal considerations. What may be considered as fair in one culture may be perceived as unfair in another. Unequal distribution of opportunity may require the application of distributive fairness that levels the playing field. For example, in the context of credit applications, there ought to be an equal probability of loan eligibility by ensuring that AI algorithmic outcomes do not discriminate against members of protected groups [3]. There are other instances where the application of corrective fairness may be necessary, for example, to remedy adverse impacts in the administration of justice, housing, education, and employment.

Second, equality does not necessarily result in the fairness of outcomes. While under Human Rights legislations disparate treatment on the basis of race, gender, nationality, disability, and sexual orientation is prohibited there may still be instances of adverse outcomes, based on other facially-neutral variables that cause a disparate impact, i.e., unintentional discrimination [5]. Consider Amazon's free same day delivery service based on an AI algorithm that included attributes, such as distance to the nearest fulfillment center, local demand in designated zip code areas, and frequency distribution of prime members to determine profitable locations for free Same-Day Delivery. The algorithm was found to be biased against minorities even though race was deemed not to be a factor in the determination of same day delivery, and minority residents in the selected zip codes were *about half as likely* to be eligible as white residents.[8]

The third challenge is balancing algorithmic fairness with fairness outcomes [10]. In this context, fairness encompasses policy and legal considerations, and leads us to ask: *what ought to be fair?* For example, in the context of hiring practices, what ought to be a fair percentage of women in management positions that AI algorithms should incorporate as thresholds to promote gender parity?

The fourth challenge relates to trade-off in balancing demographic parity with the utility of outcomes. For example, if AI algorithms remove disparate impact in the incarceration of minorities, how would that impact broader policy considerations such as the administration of justice?

Finally, fairness implicates issues of power. Before we can decide what is fair, we need to decide who gets to decide that. The conundrum we must confront is that the minority groups who are so typically the victims of algorithmic bias are rarely given a seat at the table when it is time to define what is fair. The unfortunate result is that far too often, the definition of fairness is simply what those already in power need it to be to maintain that power.

[8] https://eu.usatoday.com/

4.2 Implementing Fairness: Challenges for Data Scientists

Fairness constraints need to be considered in the context of specific use cases and for desired outcomes. Bias may be introduced at various levels within an AI system. Training data may introduce proxies that discriminate. Historical bias may unconsciously result in adverse outcomes, for example through word embeddings [4]. Representation bias through under or, over representation of training data may produce disparate impacts. The algorithms may not sufficiently adjust for fairness constraints. Inadequate testing for disparate treatment and impact may have adverse consequences for protected groups. While some argue that AI algorithms in fact minimize bias there is compelling evidence that they can and often amplify biases. Examples span facial recognition, criminal justice, hiring practices, and loan approvals [9].

Regardless of any contextualization, any definition, and any implementation approach of the fairness which is the cornerstone for Trustworthy AI, what is essential is to gain visibility to and remediate potential gaps in Trustworthy AI compliance processes. In the next section, we demonstrate how process mining could play a role in fulfilling such requirements.

5 Process Mining for Promoting Trustworthy AI

Compliance with the proposed EU AIA requires an understanding of process execution and interactions between multiple internal and external stakeholders, risk assessment of diverse systems of record that incorporate AI systems, and cooperation with various regulatory bodies and standards organizations.

The proposed AI regulation operationalizes and codifies trustworthy AI principles with prescribed mandates to institute *appropriate data governance and management practices*. The governance mechanism is complex and requires human and systems-based interactions between diverse internal and external stakeholders and EU and national regulators. Monitoring conformance with AIA is delegated to national supervisory authorities, they are empowered to order companies to take corrective actions, access all information, documentation, and data required to enforce compliance with the proposed regulation.

Given the complexity and variability of interactions implicit in achieving compliance with the proposed regulation it is our contention that *process mining* can be a valuable tool to help organizations gain visibility to various dimensions of prescribed process flows stipulated by the regulation, accelerate the analysis of how information flows, surface process bottlenecks, visualize interactions generated by event logs from disparate systems of record that may reveal areas of compliance and reputational risks. Process mining bridges the gap between data science and process science using event data captured from different types of information systems [1]. It is a data-driven approach that enables organizations to gain insight into interactions between people, systems, and organizations based on "as-is" visualization of process execution.

There are many techniques and activities in the context of process mining. However, the three main types of activities in process mining are *process discov-*

ery, *conformance checking*, and *enhancement*. Process discovery techniques take an event log and discover a process model without using any other information. Conformance checking techniques take a process model and an event log of the same process to check whether reality, as recorded in the event log, conforms to the model and vice versa. Enhancement techniques are used to extend or improve a given process model using the information about the process recorded in some event logs [1].

Process Mining can facilitate compliance with AIA by many functionalities such as: (1) Surfacing AI regulatory compliance process gaps and uncertainties, (2) Capturing user interactions performing compliance tasks, (3) Comparing process execution variations, (4) Highlighting compliance task outliers and errors, (5) Identifying potential root causes for improper execution, (6) Real-time monitoring of processes to ensure conformance to prescribed process execution paths, and (7) Triggering alerts in the event of non-compliant process tasks or changes in conditions. Furthermore, the AIA proposed regulation is inherently collaborative in nature wherein process execution spans across different organizations.

As discussed in [11], in collaborative processes where different organizations execute different parts of a shared process, the internal activities carried out by each organization are beyond the control of the other collaborators resulting in uncertainty regarding process execution. Whenever there is uncertainty in a process, there is a need for trust. Hence, collaborative business processes are especially trust-intensive. In such trust-intensive environments, process mining can be used to clarify the flow of activity execution among several organizations.

Compliance with AIA constitutes a number of interdependent steps. Performing these steps may involve variabilities in process execution paths and hand off between different stakeholders and prescribed conformance obligations to meet Articles 16-23 and Annex VII of the AIA:

- Step 1: R&D teams develop and bring to market AI systems in accordance with the risk classification system defined by the proposed regulation. If it is a high-risk AI system then a priori conformance assessment must be undertaken and a declaration of conformity must be submitted to the appropriate National Supervisory Authority. Then the AI system may be placed on the market.
- Step 2: Legal and Compliance teams must institute compliance measures in accordance with Chapter 2 of the proposed regulation that ensures adherence to data governance, accountability, transparency, accuracy, robustness, and cybersecurity provisions.
- Step 3: Data Science teams must undertake continuous monitoring of AI systems, collect data on the system's operation and take corrective action if needed. The post-market monitoring system must actively and systematically collect, document, and analyze relevant data provided by users.
- Step 4: Customer-facing functions such as Sales, Marketing, and Support, are responsible for providing clarity and certainty as to the expected AI system inputs and outputs in a way that users are informed that they are interacting

Artificial Intelligence Act (AIA)	Process Mining		
	Process discovery	**Conformance checking**	**Enhancement**
Article 9: Risk management system shall be established, implemented, documented, and maintained. *Step 1: Assessment of conformance with risk-based classification of AI systems.*	• "As-is" visualization of event logs along with risk management life cycle.	• Conformance verification that the risk management system is compliant with the prescribed requirements of Chapter 2 of AIA. Also, case analysis of specific sub-processes by drilling down to identify unexpected process execution deviations.	• Analyzing potential bottlenecks and their root causes and how do they impact downstream risk management processes.
Article 10: Data governance relating to data preparation processing operations, e.g., annotation, labelling, cleaning, enrichment, aggregation, and possible biases. *Step 2: Institute compliance measures.*	• Discovery of process steps relating to data gathering, labeling, and data governance.	• Protocol analysis of adherence to data governance rules that must be followed and identify processes that fail to meet those conditions and display protocol violations or trigger an alert.	• Continuous monitoring of data governance processes relating to implementation of fairness measures and assist in triggering remediation processes relating to identification of any possible data gaps and how they may be mitigated.
Article 12: Record-keeping that enables the automatic recording of events relating to the operation of high-risk systems. *Step 3: Record keeping and traceability of adverse impacts.*	• Validation and discovery of record keeping process execution steps and visualize processes relating to accessing training data sets.	• Audit whether documentation operations relating to the performance of high-risk AI systems are in conformance with the record keeping provisions of AIA.	• Monitoring the processes relating to the collection and documentation of the performance of high-risk AI systems.
Article 13. Institute transparency processes to enable users to interpret the system's output and include concise, complete, correct, and clear information that is relevant, accessible and comprehensible to users. *Step 4: Implement transparent communications with AI users.*	• Discovery techniques can be used to make the process of decision making transparent for users in case of objections.	• Conformance checking techniques can be used to check whether the provided transparency comply with the transparency requirements imposed by regulations.	• Enhancement techniques can be used to verify where transparency was already requested by the similar users and automatically generate and recommend transparent end-to-end process to engender trust.
Article 14. Human oversight that prevents or minimises the risks of AI adverse outcomes. *Step 5: Implement and adhere to AI Governance policy.*	• Discover complicated part of processes where human oversight could be helpful.	• Using conformance checking techniques to verify the effectiveness of human oversights.	• Actively monitoring the risk management system to learn where and when automatic risk assessment systems fail, and there is a need to involve human oversights.
Article 17. Quality management system that ensures compliance with the Regulation. It shall be documented in a systematic and orderly manner in the form of written policies, procedures and instructions. *Steps 1-5 End to end processes to mitigate and demonstrate Trustworthy AI compliance.*	• Build out event logs and "digital twin" of end-to-end quality management processes from reconstructed process instances across multiple back-end systems.	• Conformance verification that the established quality management system is compliant with Art 19 of AIA.	• Proactive Monitoring of procedures related to the reporting of serious incidents (Art 21), and communication with national competent authorities (Art 23).

Fig. 2: Process mining cadence to meet AIA prescriptive compliance obligations.

with an AI system, augmented with human oversight who monitor their operation and be able to decide, to override or reverse the output of the high-risk AI system.
- Step 5: Implementation of a Quality Management System with auditable and traceable documentation relating to the techniques, procedures for the design, of the high-risk AI systems, including procedures for data management, data analysis, data labeling, data storage, data aggregation, data retention and report serious incidents that may result in adverse outcomes.

Figure 2 further maps the compliance steps, the obligation provisions of the AIA, and process mining functionality to support Trustworthy AI. The figure illustrates how process mining techniques can facilitate AIA obligations. The FACT challenges of RDS are also taken into consideration in process mining as a subdiscipline called Responsible Process Mining (RPM) which is recently receiving increasing attention [15,14,7,8].

6 Conclusion

Trustworthy AI engenders a climate of trust essential for achieving sustainable competitive advantages in an intensely competitive environment where the application of AI is a disruptive force. The proposed EU regulation of AI is a comprehensive prescriptive measure which imposes onerous obligations, redress mechanisms on AI developers and businesses deploying AI systems. To mitigate compliance, reputational, and business risks process mining is poised to provide a data-driven approach to discover how existing Trustworthy AI compliance processes work, surface and remediate process bottlenecks, visualize different pathways of process execution and identify and remediate variations from prescribed protocols. Process mining can be a useful toolbox for ensuring that certain AI systems are designed and developed in accordance with common necessary requirements before they are put on the market and operationalized through harmonized technical standards.

Acknowledgments

Funded under the Excellence Strategy of the Federal Government and the Länder. We also thank the Alexander von Humboldt Stiftung for supporting our research.

References

1. van der Aalst, W.M.P.: Process Mining - Data Science in Action, Second Edition. Springer (2016). https://doi.org/10.1007/978-3-662-49851-4
2. van der Aalst, W.M.P.: Responsible data science: Using event data in a "people friendly" manner. In: Hammoudi, S., Maciaszek, L.A., Missikoff, M., Camp, O., Cordeiro, J. (eds.) Enterprise Information Systems - 18th International Conference, ICEIS 2016, Rome, Italy, April 25-28, 2016, Revised Selected Papers. Lecture Notes in Business Information Processing, vol. 291, pp. 3–28. Springer (2016). https://doi.org/10.1007/978-3-319-62386-3_1

3. Binns, R.: On the apparent conflict between individual and group fairness. In: Proceedings of the 2020 conference on fairness, accountability, and transparency. pp. 514–524 (2020)
4. Dator, J.: Chapter 3. What Is Fairness?, pp. 19–34. University of Hawaii Press (2006). https://doi.org/doi:10.1515/9780824841966-004, https://doi.org/10.1515/9780824841966-004
5. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., Zemel, R.: Fairness through awareness. In: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference. p. 214–226. ITCS '12, Association for Computing Machinery, New York, NY, USA (2012). https://doi.org/10.1145/2090236.2090255, https://doi.org/10.1145/2090236.2090255
6. Elkoumy, G., Fahrenkrog-Petersen, S.A., Sani, M.F., Koschmider, A., Mannhardt, F., von Voigt, S.N., Rafiei, M., von Waldthausen, L.: Privacy and confidentiality in process mining - threats and research challenges. CoRR **abs/2106.00388** (2021), https://arxiv.org/abs/2106.00388
7. Elkoumy, G., Pankova, A., Dumas, M.: Mine me but don't single me out: Differentially private event logs for process mining. CoRR **abs/2103.11739** (2021), https://arxiv.org/abs/2103.11739
8. Fahrenkrog-Petersen, S.A., van der Aa, H., Weidlich, M.: PRIPEL: privacy-preserving event log publishing including contextual information. In: Business Process Management - 18th International Conference, BPM. Lecture Notes in Computer Science, vol. 12168, pp. 111–128 (2020)
9. Grother, P., Ngan, M., Hanaoka, K.: Face recognition vendor test part 3: Demographic effects (2019)
10. Kleinberg, J., Ludwig, J., Mullainathan, S., Rambachan, A.: Algorithmic fairness. In: Aea papers and proceedings. vol. 108, pp. 22–27 (2018)
11. Müller, M., Ostern, N., Koljada, D., Grunert, K., Rosemann, M., Küpper, A.: Trust mining: Analyzing trust in collaborative business processes. IEEE Access **9**, 65044–65065 (2021). https://doi.org/10.1109/ACCESS.2021.3075568
12. O'neil, C.: Weapons of math destruction: How big data increases inequality and threatens democracy. Crown (2016)
13. Phillips, P., Hahn, A., Fontana, P., Broniatowski, D., Przybocki, M.: Four principles of explainable artificial intelligence (2020)
14. Rafiei, M., van der Aalst, W.M.P.: Group-based privacy preservation techniques for process mining. Data Knowl. Eng. **134**, 101908 (2021). https://doi.org/10.1016/j.datak.2021.101908
15. Rafiei, M., van der Aalst, W.M.P.: Privacy-preserving continuous event data publishing. CoRR **abs/2105.11991** (2021), https://arxiv.org/abs/2105.11991
16. Shahriari, K., Shahriari, M.: Ieee standard review — ethically aligned design: A vision for prioritizing human wellbeing with artificial intelligence and autonomous systems. In: 2017 IEEE Canada International Humanitarian Technology Conference (IHTC). pp. 197–201 (2017). https://doi.org/10.1109/IHTC.2017.8058187
17. Stanton, B., Jensen, T.: Trust and artificial intelligence (2021-03-02 05:03:00 2021), https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=931087