Deep Science Publishing
https://doi.org/10.70593/978-93-49307-44-5_9

DeepScience
Open Access Books

Chapter 9: Managing compliance, fairness, and transparency while deploying artificial intelligence systems in regulated industries

9.1. Introduction

Artificial intelligence (AI) is being harnessed to complete important tasks across various domains, while concerns about malicious use and unintended side effects are growing. Recent initiatives towards AI governance have also remarked on the trepidation concerning AI technologies. Although such technologies hold immense potential for beneficial use, they can be misused or unintentionally have negative effects, necessitating oversight mechanisms to guarantee compliance with policies and standards on AI. The wide adoption and use of Generative AI (GenAI) systems have led to increasing calls for their regulation amid fears of misuse, harmful outputs, and unfair bias. Regulated industries depend on the decisions made by AI systems that are not only consequential but potentially discriminatory, thus the governance of these hyperparameters is even more consequential. Concerns about impending harm from GenAI outputs have led to calls for regulation, including recent legislative proposals to create “guardrails”. The implementation of measures is of equal interest: indeed, guidelines abound but many implementation details are unknown.

This article combines insights from AI policy experts and practitioners in regulated industries on challenges and solutions around the implementation of compliance, fairness, and transparency measures. Implementation challenges regarding compliance with AI regulations and aims for fair and transparent AI systems are examined, along with opportunities for AI developers and teams to address such challenges. Several implementation challenges regarding compliance with AI regulations and aims for fairness and transparency in AI systems will be discussed. At the same time, possible

entry points for AI developers and teams to respond to such challenges will be shared in the hope of facilitating further discussion and collaboration. Many organizations are confronting both similar challenges and possible entry points to address them in comparable manners. Collaboration on methods and practices of innovation, from both the technical and design perspectives, could enhance governance of AI systems that serve the public interest.

Fig 9.1: AI Compliance

9.1.1. Background and Significance

Another major blind spot is the cause of harmful outcomes of AI systems – the AI value chain including data, algorithms, and decision-making processes. In a bottom-up approach detailing the AI lifecycle, from conception and development to deployment and monitoring of AI systems, consumer protection can be enforced through data protection, transparency, audit, discrimination, propagation of harm, accountability, and liability. AI is a mixture of statistics and computer science that comes with a price. In other words, the nature of AI entails opaque systems that can bereave humans of understanding the cause and effect of an algorithmic decision, thus defeating the goals of transparency and accountability.

This makes the burden-of-proof to demonstrate disparate impact of AI systems on the part of the harmed individual high. Moreover, AI systems are large and complex, and there are a multitude of reasons for alleged discrimination against AI systems. A statistical discrepancy between input and output data is indicative of discrimination but by no means is definitive evidence of discrimination unless random experiments are conducted, which is hard to simulate with modern-day AI systems. Consequently, civil recourse avenues in claiming and litigating bias against AI systems fare poorly in comparison to other forms of consumer protection in finance, health insurance, and employment.

9.2. Understanding AI in Regulated Industries

AI is frequently deployed in a variety of industries, such as finance, healthcare, and media (Dhawan, 2024; Cyriac et al., 2025; Klein, 2025). Particularly in these areas, AI systems are often required to comply with numerous governmental regulations and internal policies, which adds some constraints on possible model structures, training strategies, and deployment methodologies. Although these datasets often originate from different departments or institutions and have substantially different characteristics, they are usually all relevant for the model's purpose. Moreover, while complying with these regulations and constraining the available model designs, they also introduce issues such as data hoarding, capital misallocation, and market manipulation. Violation of regulations put in place to safeguard against these problems could lead to huge penalties for the institution employing the AI systems as well as possible criminal charges for the employees responsible for maintaining compliance.

AI systems in regulated industries are legally required to satisfy a wide array of fairness, transparency, and accountability principles. For instance, in finance, models used for credit adjudication must meet the Fair Credit Reporting Act requirements. In healthcare, AI systems are often subject to HIPAA regulations, which govern the privacy and technical security of patient information. For AI algorithms in finance and healthcare, government regulations require evaluation and auditing tools to guarantee fairness and transparency. However, these regulations have not been automated into black-box auditing or monitoring mechanisms that could be employed by low-tier institutions, where a dominant majority of innovation occurs. Both the institutions being regulated and the regulators themselves have general guidelines that wait to be compiled into the machine code needed to build regulations into the deployment process.

9.2.1. Defining AI Systems

The European Union has started to regulate AI by setting an AI Act to create a single regulatory framework for these systems. The proposal focuses on accountability and oversight relevant to their impact, alongside a risk-based categorization. Proportionality of obligations with a focus on compliance for the highest-risk systems. However, the topic is very wide, and the interaction with multiple technological, societal, and ethical dimensions of the data economy is expected to be challenging in its implementation.

Paradox of artificial intelligence systems. Due to their ability to categorize, process, and optimize solutions, systems based on machine learning applied in the public interest domain, e.g., for recruitment, financing, policing, are increasingly proposed, developed, and deployed. However, with great power comes great responsibility. A growing number of discrimination and unfairness concerns have arisen across various application domains. In turn, this has led to an increase in regulated and self-regulated domains and discussions surrounding risk, accountability, and explainability. Nevertheless, a unified regulatory framework has not been established. Starting with the synthesis of insights into fairness and anti-discrimination concerns across various domains and perspectives, and their intersection with current and anticipated regulations, ideally leading to AI regulations aligned with the European values on fairness would be addressed.

General introduction to modern AI systems. Attempts to provide simulation methodologies for training, validation, and benchmark of AI systems often focusing on highly data-driven AI approaches. Increased need to understand the underlying decision processes of AI systems and have comprehensive technical assessments regarding their compliance with regulations. Scaling up AI legislation and regulation alongside the technology development through audits and impact assessments. Acknowledge that the concern surrounding AI systems is just beginning to be addressed with proper regulations and compliance frameworks.

9.3. Compliance Frameworks for AI Deployment

AI systems continue to mature and provide significant ethical and regulatory challenges for financial service players and regulators alike (Financial Times, 2023; Rowe, 2024). Traditional IA governance frameworks have shown limitations with an inability to keep pace with the fast-evolving AI landscape and complexities, the widespread use of externally sourced AI models, and neural networks that are opaque by design. There are clearly still very significant regulatory gaps in the deployment of AI in LOBs. The AI governance framework will require significant augmentation to bring more comprehensive coverage and enforcement of compliance across both AI models and financial service players. This increased coverage is expected to introduce a new class

of compliance professionals better versed in AI. This could be a recent focus area in graduate school education. Outside of training, AI players have started to introduce a new tier of AI compliance personnel. They come with this added scrutiny and diligence imposed on the findings and consequences of the usage of AI on MS processing and recommendations. They also possess the ability to apply a legal lens to newer aspects of AI compliance and legality such as data availability and usage. Most of the described compliance personnel would report formally to an independent compliance team under the Chief Compliance Officer within the second line of defense.

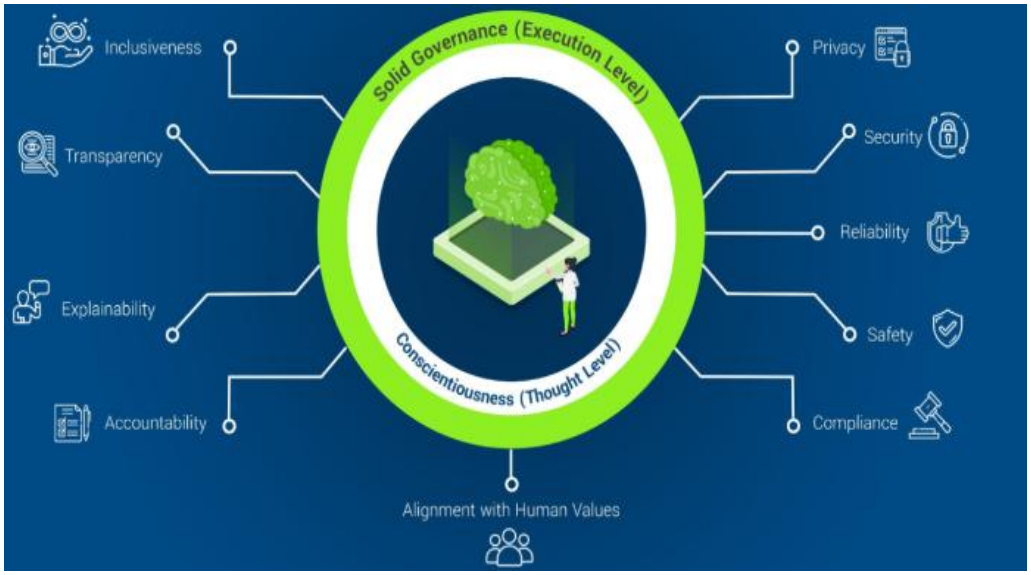

Fig 9.2: Responsible AI Framework

In addition, AI governance will require the usage of more complex and modern compliance tools. AI governance frameworks have started to appear, but they often have limited coverage and enforcement capabilities. This requires the provisioning of more deeply integrated enterprise risk platforms, which can leverage internal and external data to take a far more holistic view of compliance and related risks across multiple financial service players and AI models. AI governance will increasingly rely on tools such as algorithmic transaction surveillance for both supervised and unsupervised usage. These tools will allow the more thorough capture of systemic AI risks across player instances, models, and instances of model processing. It is expected that AI players and regulators will need to partner closely to implement more comprehensive and effective AI compliance tooling for this purpose.

9.3.1. Legal and Regulatory Standards

The AI Act differentiates between four tiers of risk. High-risk AI systems to be regulated by the AI Act must be classified in domains that either entail a significant risk to the health and safety or the fundamental rights of persons, or are likely to affect public interest in other domains. These regulated areas include ultrasound and temperature scanning devices for the evaluation of thermal risk, exoskeletons for risk of musculoskeletal disorders, service robots with social interaction for risk of negative psychological influence, and AI-enabled recruitment systems. While exoskeletons, temperature scanners, and segmentation or prediction algorithms for images and sounds belong to low-risk computer vision applications, which would fall outside the scope of regulation by the AI Act, AI-foundation models have high stakes, raising questions about public safety and accepted notions of fundamental rights. Understanding AI technology in practical terms and becoming a competent regulator is difficult. Especially fundamental rights are transient and subject to social agreements, requiring close collaboration with civil society to gain insight into the safety and soundness of AI, filtering pernicious systems. With traditional scientific and technological culture unable to process these changes, solutions outside existing categories and ideas need to be developed.

The AI Act demands compliance with legal and regulatory standards. For example, rendering models explainable could partner AI technology with Human-in-the-loop (HITL) notions to embed human expertise into decision-making processes. However, while ways of enhancing algorithmic fairness or compliance exist and work arithmetically, legal standards and outcomes are subject to transformation. Instead, it is proposed to solidify and render transparent the fairness expert judgment processes through loop models. A system could be paired with a human that could sort data examples and with at least a part of their deliberation either directly recorded or modeled. Any empire built on smoke disappears in the light; the same applies to decision processes. While rendering AI algorithms similar or compliant with existing laws, holding systems accountable for the impacts of decisions entails scrutiny of the fairness reasoning. Fairness experts and legal verdicts are themselves questioned for ambiguities, incompleteness, and other incompliance issues. AI-based systems have limitations in being engineered to guarantee compliance with something as broadly agreed on as fairness, unless lofty ultimate definitions of fairness are agreed on and machinery producing, verifying, and assuring compliance of fair actions are conceived. Expecting accountability from systems producing fair actions, judgments, and results would only be expectable from an oracle that inspects the depth of the ocean or precisely forecasts rainfall 20 years in advance.

9.4. The Importance of Fairness in AI Systems

In many regulated industries, AI systems are expected to be fair, in addition to being compliant and transparent. A particular focus is on the fair treatment of individuals, where bias exists if those individuals are treated systematically unfairly with respect to some protected characteristics such as race, gender, or age. But this notion of fairness is challenging and highly nuanced. Many different definitions fit within the fair treatment framework, with different predicates expressing a notion of classification, prediction, or treatment bias. There are various ways in which these definitions can be instantiated, expressed in terms of egalitarian principles such as statistical parity, equality of treatment, and conditional demographic equality.

Even when definitions of bias are aligned and this bias is demonstrably present in an AI system, the decision about how to respond to this imbalance is not obvious. Bias in AI systems can arise through many sources across their lifecycle, and can be addressed or mitigated in a range of technical or regulatory ways that may or may not be feasible or ethical. With AI systems operating in various high-risk sectors and interacting with individuals and communities across disparate social and cultural borders, the actions recommended or required to ameliorate fairness disparities will vary widely in form and feasibility.

9.4.1. Defining Fairness in AI

The Fairness in AI Working Group of the Partnership on AI undertook multiple initiatives with the aim of exploring AI Fairness concerns and articulating a common framework for discussing them. Participants included a range of organizations from academia, industry, the public sector, and civil society. All discussions were oriented around the following Primary Question: What does Fairness in AI mean? What issues, risks, or concerns must be brought to light and/contextualized (ideally, framed with examples)? Included with the Primary Question were the following Auxiliary Questions: What are the bounds of this conceptual space? What questions or issues fall outside scope? What are the criteria / attributes / principles for a good definition of Fairness in AI? The notion of AI Fairness being described here can be better understood if it is initially presented in terms of what it is not. AI Fairness cannot be understood in general or abstract terms: Without context and specification of the applied setting/purpose, it is empty rhetoric. AI fairness must come with a contextual specification detailing (1) the application domain (i.e., the industry* and/or sector); (2) the socio-technical context (which individuals/groups are involved or represented, and in what way); (3) the considerations of concern (desiring AI systems to be fair in what manner). Fairness must be understood as a concern in AI case settings; one cannot adjudicate whether an AI system is fair or not outside the context of that system and its application. Each of these

different contexts indicates different sets of relevant fairness concerns. The definition of fairness that the group advanced must now be contextualized in this manner. A context-independent fairness principle, "discriminatory non-harm" was articulated to address this conceptual gap. Defined as "the non-discriminatory design, marketing, audit, and use of AI systems", this is a common principle of fairness that applies across all considered scenarios.

9.5. Transparency in AI Systems

While deploying AI systems in regulated industries, obtaining transparency in both regulated and unregulated contexts is challenging. Here, the paper presents the concept of transparency and its categories, with insights into how transparency can be acquired through regulatory means.

Transparency means an observable process that thus can be questioned or scrutinized and is inherently tied to a relationship and the expression thereof. It is identified as one of the tools to govern AI systems through the provision of specific information to stakeholders controlling the whole spectrum of the AI life-cycle. Thus, the paper discusses the choices to be made to design transparency frameworks to comply with the right to transparency. In addition, the limits of transparency are discussed, focusing on the degrees of technical transparency. This degrees-based schema interprets existing concrete transparency frameworks regarding the depth of insights into the decision-making process of AI systems.

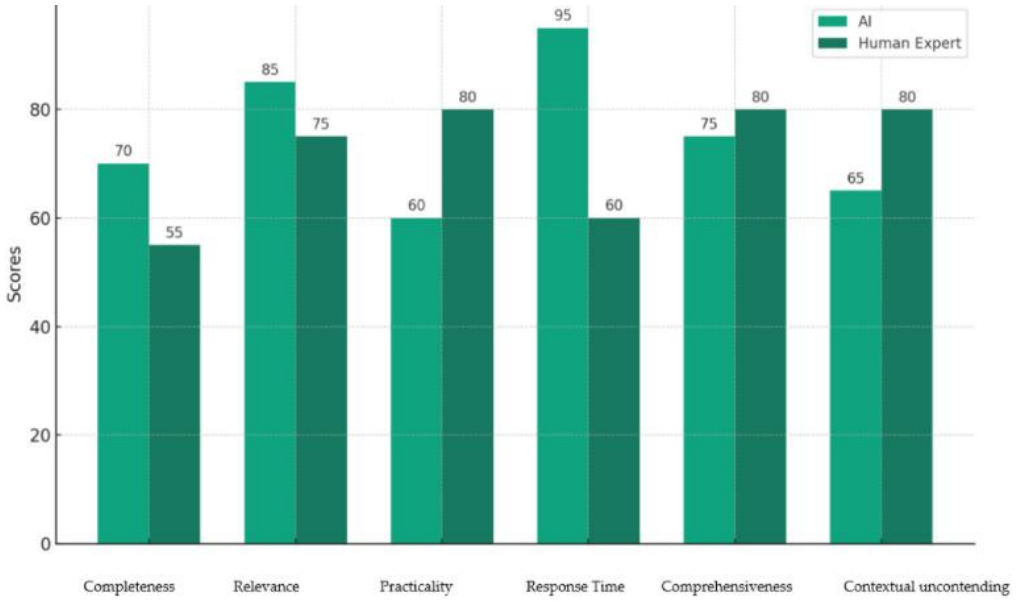

Fig : Navigating the Power of Artificial Intelligence in Risk Management

Decades of research on transparency in sociotechnical systems has brought to the fore multiple conceptualizations and classifications of transparency relevant to AI systems. First, transparency is categorized into different types, then the characteristics of transparent AI systems are introduced. Last, the fundamental process underlying transparency is put forward. Through this clarification, the opportunities and challenges of transparency in AI systems are illuminated. Conclusively, this research aims to offer theoretical refinements that resonate with both socio-technical and engineering audiences and can support further avenues of research on transparency in AI systems. Ultimately, transparency in these systems can better enable fairness and accountability in their design, development, and deployment while facilitating an appropriate degree of trust in them across very different contexts and applications.

Transparency is mostly seen as an advance of algorithmic or data transparency. In simple terms, transparency in AI systems indicates the quality of being open to information rather than black-boxed upon. A deeper contention is that due to the weighting of ethical considerations or working principles, AI systems may be actively designed as opaque. Hence, a degree of transparency is set in a dynamic tension with privacy, security, or competitiveness. Provided the needed information, the AI systems inform adequately how they function or why were outputs, in agreement with stakeholder expectations.

9.5.1. The Role of Transparency

The comparatively recent movement towards “transparency by design” for AI systems is aimed at addressing the demand side of transparency by taking both the audience and context into account. Transparency is increasingly seen as a crucial element in its own right. Efforts to promote “transparency by design” for AI systems have mushroomed across different spheres, building on the so-called “transparency by design” idea in quality systems. Future modes of governance are likely to see AI development required to disclose technical information, documents, and tests to regulators and users. However, despite the increasing popularity of the idea, it remains theoretically ambiguous, and its implementation practically controversial.

Human-AI systems can affect high stakes like human dignity, social inclusion, and public order, and humans are sceptical of transparency if not prefer non-transparency. The critics start with a narrower technological focus than is common in the literature on explainable AI. Transparency by design is expected to prevent systemic failures and largely avoid nondisclosure. For instance, to govern algorithmic decisions, the design should disclose their purpose, sourcing of input, and data origin. Transparency by design is expected to provide ordinary users with encapsulated means to understand the more complex background of decisions made by AI. In future modes of governance, the pressure for transparency may shift from top-down initiatives to bottom-up efforts.

In the compliance context, norms traditionally were understood as the duty to be open to regulators and to disclose information instead of profiting from it. Systematic transparency and information to be disclosed based on specifications/rendering could be managed and regulated with compliance functioning as a literal monitor. However, the implementation of transparency by design would ultimately constitute a powerful way to subvert or creatively destroy the original compliance function for governance.

9.6. Risk Management in AI Deployment

AI systems are of growing societal importance and risk, starting widely followed debates on regulation. In view of the rapidly growing prevalence and importance of AI systems, to mitigate harms arising from the use of these systems, the AI deployment sector has a growing, increasingly urgent need for systems and mechanisms to manage the risks associated with deploying these systems. The nature of this need ultimately motivates this research endeavor, which seeks to develop a technology-agnostic framework, methodology, and toolkit for assessing the systematic and comprehensive risk of deploying AI systems.

The AI deployment problem consists of these core interrelated parts: (1) a deployed AI system, (2) inert data, (3) an end goal, (4) functions of the AI system that support the end goal, and (5) a domain in which the AI system is correlated with the data. Different organizations deploying AI systems can have different end goals for the systems to be deployed as well as important correlations between data and AI systems that apply specifically to their actions. In connection with launched AI systems, the involved organizations and AI system developers are deeply concerned with preventing and mitigating known risks of publicly deployed AI systems, including risks of unfairness, lack of AI system competence, and lack of proper care taken to mitigate such risks.

9.6.1. Identifying Risks Associated with AI

AI-based systems promise to provide organizations, individuals, and society with substantial value. This value can be obtained by optimizing costs and revenues, enhancing employee, customer, and community experience, advancing innovation, extending human capabilities, providing insights based on vast collections of data, or augmenting abilities through automation. However, many significant attendant risks have also been identified. Fraud and misinformation can be propagated more widely, decisions affecting a person’s life can be inaccurate and opaque, and malicious deep fakes can be created with unprecedented ease. Alternatively, a lack of access to the technology can result in the entrenchment of current power models or the inability for healthcare or education delivery in resource-limited environments.

Organizations want to realize the benefits of the positive capabilities of AI technology while reducing the risks. This is at the forefront of C-level and board-level focus in reviewing the strategic direction of the organization. Existing regulations worldwide are attempting to address these risks to society with considerable pressure to ensure compliance. Regulators and standards bodies are also issuing guidance and standards as rapidly as possible. The same objective is behind the current prioritization of these topics, with substantial effort given to codifying the standards, processes, and policies necessary to satisfy the regulatory requirements, including substantive measures to demonstrate compliance.

The best way to reduce the risks is to implement comprehensive AI lifecycle governance where the policies and procedures are enforced on stakeholder activities during the design, development, deployment, and monitoring of an AI system. The difficulty in this case is that organizations often need to identify the risks of deploying an already-built model without knowledge of how it was constructed or access to its original developers. Questions to answer include: What was the data used to train the system, and what risks does such data pose? What safeguards against bias and discrimination were employed, if any? What testing and validation procedures were used to demonstrate reliability and performance? Was the model robust against conceptual drift following deployment and monitoring? Was the system explainable in order to understand 1) why a decision was reached and 2) determine whether and how models could be updated or retrained as knowledge was developed or changed? When detection errors occurred, why did they happen, and how similar were the detection test samples leading up to it?

9.7. Conclusion

With the changes in mechanism, methods, and deployment challenges of these AI technologies, companies with existing diasporas of AI systems are left with demanding challenges in complying with evolving regulatory frameworks, maintaining current fair and transparent practices, and adopting reasonable explainability methods. The recent regulations on AI system technologies in heavily regulated industries focus on compliance, fairness, and transparency. There are concrete use cases where regulatory requirements and/or fair and transparent practices imposed by organizations have changed dramatically, compelling AI system adaptations after deployment. While these are well-documented challenges for emerging AI technologies in general, healthcare and finance industries face unique pressures to ensure compliance and minimize adverse outcomes for affected groups. Limited opportunities are available for AI practitioners to understand these regulatory and corporate considerations without prior specialized knowledge or experience.

A deeper understanding of changes in compliance processes and an analysis of the external and regulatory contexts surrounding these changes may help organizations devise better mitigation strategies. Compliance and audit firms are equipped to assess current practices, map them to a new regulatory framework, evaluate potential exposure to relevant regulations and standards, and assist with change management decisions. Nonetheless, these firms do not develop compliance processes or technologies, as this would pose a conflict of interest. Further, the process mining community does not have in-depth, specialized knowledge of technical details and expected processes for compliance with evolving regulatory frameworks in finance and healthcare.

Nine case studies drew on insights from these compliance firms as well as interactions with top-tier healthcare and finance companies. Rigorous thematic analysis of qualitative data from these interactions identified over 100 open questions and challenges and synthesized them into a manageable set of topics. Organizing this material into a workshop agenda will support agreed-upon key topics for discussions among five practitioner-led panels, complemented by invited talks on auditing AI technologies.

9.7.1. Emerging Technologies

Consequently, society's expectations of transparency, interpretability, and effective governance of machine learning applications is becoming increasingly strict. As the use of these technologies proliferates across a wide range of decisions, including hiring, lending, job performance monitoring, product recommendations, criminal risk assessment, healthcare treatment recommendations, and more, there is growing awareness that these emerging technologies must be designed and deployed in ways that are fair and nondiscriminatory and that such fairness can be governed and regulated rigorously. There is a substantial body of academic research that is focused on fairness in machine learning and algorithmically delivering fair outcomes. However, much of this research does not directly address the complexities of algorithmic systems in practice. Specifically, there are complexities associated with legal and compliance implications surrounding fairness, as well as complexities of the design of algorithmically-delivered workflows that bear on fairness. Compliance with Trustworthy AI governance best practices and regulatory frameworks is an inherently fragmented process, resulting in process uncertainties and compliance gaps that may expose organizations to reputational and regulatory risks.

There are complexities associated with meeting specific dimensions of Trustworthy AI best practices such as data governance, conformance testing, quality assurance of AI model behaviors, transparency, accountability, and confidentiality requirements. Various custom-developed business processes can be followed to manage AI compliance from the perspectives of process steps, compliance roles, responsible tools, and

supporting datasets. Fundamental challenges exist including: (i) How to gain fact-based visibility into the AI compliance process execution and the management of process uncertainties? (ii) How to surface detailed compliance bottlenecks? (iii) How to analyze, remediate, and monitor the uncertainty in AI regulatory compliance processes?

This paper provides a comprehensive introduction to the concepts of Process Mining and Regulatory Technology and presents RAIView, a new process mining tool for algorithmic model compliance, which exposes the adherence of algorithmic models to these three regulatory dimensions. This paper also showcases its application in the financial services industry.

References

Klein, C. (2025). Agentic AI is a 'big next step' in AI's evolution. Axios. Retrieved from axios.com

Dhawan, R. (2024). AWS is helping financial giants like JPMorgan and Bridgewater with their AI ambitions. Business Insider. Retrieved from businessinsider.comBusiness Insider

Financial Times. (2023). AI in banking, payments and insurance. Financial Times. Retrieved from ft.comFinancial Times

Cyriac, T., Regenstein, J., & McConnell, S. (2025). Agentic AI in Financial Services and Insurance. Snowflake. Retrieved from snowflake.comSnowflake AI Data Cloud

Rowe, T. (2024). How Agentic AI is Transforming the Banking Industry. Intelligent Core™. Retrieved from intelligentcore.io