AI is Entering Regulated Territory: Understanding the Supervisors' Perspective for Model Justifiability in Financial Crime Detection

Astrid Bertrand, James R Eagan, Winston Maxwell, Joshua Brand

AI is Entering Regulated Territory: Understanding the Supervisors' Perspective on Model Justifiability in Financial Crime Detection

Astrid Bertrand
Télécom Paris, Institut Polytechnique de Paris
LTCI
Palaiseau, France
astrid.bertrand@telecom-paris.fr

James R. Eagan
Télécom Paris, Institut Polytechnique de Paris
LTCI
Palaiseau, France
james.eagan@telecom-paris.fr

Winston Maxwell
Télécom Paris, Institut Polytechnique de Paris
SES, i3, CNRS
Palaiseau, France
winston.maxwell@telecom-paris.fr

Joshua Brand
Télécom Paris, Institut Polytechnique de Paris
SES, i3, CNRS
Palaiseau, France
joshua.brand@telecom-paris.fr

ABSTRACT

Artificial intelligence (AI) has the potential to bring significant benefits to highly regulated industries such as healthcare or banking. Adoption, however, remains low. AI's entry into complex socio-techno-legal systems raises issues of transparency, specifically for regulators. However, the perspective of supervisors, regulators who monitor compliance with applicable financial regulations, has rarely been studied. This paper focuses on understanding the needs of supervisors in anti-money laundering (AML) to better inform the design of AI justifications and explanations in highly regulated fields. Through scenario-based workshops with 13 supervisors and 6 banking professionals, we outline the auditing practices and socio-technical context of the supervisor. By combining the workshops' insights with an analysis of compliance requirements, we identify the AML obligations that conflict with AI opacity. We then formulate seven needs that supervisors have for model justifiability. We discuss the role of explanations as reliable evidence on which to base justifications.

CCS CONCEPTS

• **Social and professional topics** → *Automation*; **Socio-technical systems**; • **Human-centered computing** → **Empirical studies in HCI**; • **Applied computing** → *Law*.

KEYWORDS

justifiability, explainability, highly-regulated environment, anti-money laundering, AI regulation

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05…$15.00
https://doi.org/10.1145/3613904.3642326

ACM Reference Format:
Astrid Bertrand, James R. Eagan, Winston Maxwell, and Joshua Brand. 2024. AI is Entering Regulated Territory: Understanding the Supervisors' Perspective on Model Justifiability in Financial Crime Detection. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 21 pages. https://doi.org/10.1145/3613904.3642326

1 INTRODUCTION

AI regulation has been rapidly gaining interest due to the advances of generative AI and the emergence of new AI regulations[1]. However, highly regulated industries, such as banking, healthcare, and the military, already have structures in place to deal with technological risks. These domains are characterized by well-established norms, experience in putting principles into practice, a common goal of social welfare, and robust professional accountability mechanisms [84]. In banking, machine learning adoption is on the rise [36], with regulators sometimes encouraging industry players to consider AI to improve the efficiency of their systems [12]. However, little new regulatory guidance has been provided to address the specific risks of AI [37, 104] and firms call for a more proactive regulation approach [36, 109]. Truby *et al.* [109] notes an overall lack of guidance on AI use from "typically cautious financial regulators". Overall, clarification is needed on how current regulatory mechanisms address the risks of AI.

In this paper, we focus on a highly-regulated area, anti-money laundering and countering financing terrorism (AML-CFT). AI applications for AML-CFT, such as unsupervised anomaly detection, have attracted increasing attention from both industry players and academics for their potential to reduce compliance costs and detect new patterns of money laundering that current rule-based systems are not aware of [46, 100]. In experimental conditions, Weber *et al.* [113] has found that these methods can reduce the number of false alerts for money laundering by 20 to 30%. The impact of such technologies is all the more promising as current AML-CFT systems are relatively ineffective [9]. The United Nations Office on Drugs and Crime estimates that between 2 and 5% of global GDP

[1]For example, the developments of the AI, Digital Services and Digital Markets Acts in Europe and the Algorithmic Accountability Act in the US this year [29, 31, 117].

is laundered each year and less than 1% of these funds are seized or frozen [111]. Banks have been increasingly touting the use of artificial intelligence (AI), to the extent that AI use for AML-CFT is entering a tipping point. Big tech companies have also begun to provide AI services for AML-CFT systems within banks, such as Google's collaboration with HSBC which resulted in a 60% reduction of false positive alerts and quadrupling the number of true positives [106].

Kruse *et al.* [57] argue that the primary challenge posed by AI algorithms in the finance industry is related to their opacity. As highlighted by Kuiper *et al.* [58], AI opacity undermines the ability of financial institutions and regulators to control their systems, thereby posing a risk to financial stability, institutional trust and consumer protection [58, 79]. In AML-CFT, concerns of regulators have also focused on the lack of transparency in AI models and on measuring their added value [45]. In October 2022, however, a Dutch court ruling confirmed that the financial institution Bunq could use AI for AML-CFT despite reservations from the regulator [108]. Overall, it is undisputed that a certain level of transparency is required for AI models [77]. It is rarely specified, however, to what extent and why AI explanations should be generated in relation to applicable legal requirements. Moreover, few studies have explored the regulator perspective, despite the fact that they are an essential audience of AI explanations.

In banking, a distinction is made between **regulators**, who are responsible for drafting the rules, and **supervisors**, who verify that the rules are applied. In this paper, we focus on AML-CFT supervisors in France, also referred to as **controllers**, who act as the national public auditors of AML-CFT systems in banks. We strive to understand the supervisors' perspective on AI transparency and justifications in this context of the highly regulated AML-CFT environment in France. Specifically, we leverage two scenarios of promising AI applications from the AML-CFT literature and conceptual design artifacts of AI justifications and explanations [40]. We outline the justification requirements and information needs of supervisors regarding AI systems to help banks better design justifications for AI systems and to help supervisors build relevant explainability and testing solutions for auditing purposes. Grounded in the context of AML-CFT, our study is guided by the following research questions:

- **RQ1**: What are regulatory supervisors' current auditing practices and socio-technical context? (Section 4.1)
- **RQ2:** How does AI opacity conflict with compliance requirements and to what extent can justifiability address these tensions? (Section 4.2)
- **RQ3:** What are the needs of supervisors for justifiability of AI systems? (Section 4.3)

Our study adopts two original approaches. First, the needs and context of regulators, supervisors, and auditors is not currently well understood. By exploring their justification needs, we can reduce regulatory uncertainty around the use of AI. Investigating the supervisor perspective will inform how existing accountability mechanisms can be applied to AI technology. Second, in order to fully understand the objectives and needs of supervisors, it is necessary to consider the legal requirements. As such, we conduct a multi-pronged socio-techno-legal study of these users and their context.

Our contributions can be summarized as follows:

- we describe the socio-techno-legal supervision system and auditing approaches in the AML-CFT context,
- we assess compliance obligations specific to AI-enhanced AML-CFT systems highlighting why the opacity of AI systems may pose problems with regard to AML-CFT obligations,
- we formulate supervisors' needs in terms of model justifications and explanations,
- we demonstrate the complementarity of a dual HCI and legal methodology to fully understand regulatory supervisors' justifiability needs.

We begin by presenting the related literature and the relevant background in AML-CFT. We then describe our methods and findings. We conclude with a discussion regarding the role of explanations for justifiability. We consider explanations' limitations and alternatives such as system-wide testing. We hope that these findings will help AI adopters in finance, and in other highly regulated environments, to design more effective justifications of AI decisions and systems.

2 BACKGROUND AND RELATED WORK

2.1 Terminology

This section clarifies the key terms we use throughout this article and the way the concepts relate to each other. It sets the stage to articulate the role of explanations for justifications, which is discussed in Section 5. Further, our focus on justifiability has implications for notions of auditability and accountability.

Explanation, explainability, explainable AI. Explanations of AI systems are transfers of knowledge about the behavior AI systems [49, 82]. Henin and Le Métayer [49] state that explanations are *"descriptive and intrinsic in the sense that they only depend on the system itself"*. Explainability broadly refers to providing explanations of AI systems to relevant stakeholders to scrutinize AI models in their development, implementation, and deployment stages [50]. Explainable AI (XAI) is the technical arm that aims to provide explainability. Following Markus *et al.* [75] and Gilpin *et al.* [43], an AI system is explainable if it is intrinsically interpretable, or if it is complemented with an interpretable and faithful explanation.

Justification, justifiability. As presented in this paper, regulatory supervisors expect a "justification" by regulatees that an AI system or decision complies with a legal standard, rule, or objective. Justifications are therefore crucial in the process of verifying regulatory compliance, which involves auditability and accountability. According to Henin and Le Métayer [49], a justification, or "justifiability", is an argumentative process that refers to external norms to argue that a decision (or a system) is "good" (or adequate). Justifications are normative and extrinsic [49, 52] as they are grounded in external norms, such as legal requirement. In Section 5, we argue that justifications must also be grounded in intrinsic and accurate information about AI systems implementation, such as through explanations.

Audit, auditability. In the context of a regulated environment, an algorithmic audit is a governance mechanism in which inspectors participate in a field experiment to diagnose compliance risks associated with AI systems in relation to specific regulations[2]. "Auditability" of AI systems enables "the assessment of algorithms, data and design processes"[51] and permits auditors to conclude on the compliance of AI systems [95, 105]. The EU's High Level Expert Group on AI noted the key role of auditability for accountability [51].

Accountability. According to Doshi-Velez *et al.* [24], accountability is "the ability to determine whether a decision was made in accordance with procedural and substantive standards and to hold someone responsible if those standards are not met." Additionally, an important element of accountability is the capacity to demonstrate compliance. Felici *et al.* [33] states: "Accountability involves [...] demonstrating ethical implementation to internal and external stakeholders". We believe that this demonstration element is provided by justifications.

2.2 Understanding user needs for explainability

To inform the design of explainability systems, HCI researchers have relied on cognitive theories about how users explain [7, 23, 44, 66, 68, 82, 99, 112], on interviews [26, 55, 64, 65, 74, 102, 110] and participatory design [18, 92, 112] to learn about users' contexts and needs. These approaches are the starting point for the disciplinary triangulation characteristic of the HCI field between natural science theory, artefact design, and scientific observations to design empowering systems [73]. Some of this work has helped to delineate groups of users [76, 85, 103, 107], or to identify the different questions that users ask of AI systems [27, 64, 66, 67]. Work exploring user needs through interviews have provided detailed insights on specific user groups, contexts and AI applications. For example, Liao *et al.* [64] created a bank of questions that users may have on AI systems, building on 20 interviews with UX and design practitioners. Sun *et al.* [102] conducted workshops with 43 software engineers to explore their explainability needs when using generative AI for code. Ehsan *et al.* [26] interviewed 29 AI users and practitioners to learn about the socio-organizational context of XAI-aided decision making, a perspective they call "Social Transparency". Chazette and Schneider [16] further emphasised that the elicitation of explainability needs should also take into account laws and norms, cultural and corporate values, domain aspects, and organisational constraints such as time and resources [74]. Scenario-based design, [14], in which participants are engaged in a scenario to elicit their feedback, has often been used to understand explainability users in various context [19, 65, 102, 116]. However, no work in the HCI field has addressed the elicitation of explainability needs using both a scenario-based and a legal approach, to the best of our knowledge. Our view is that it is particularly relevant to the study of the needs of regulators.

2.3 Designing AI justifications for compliance

As noted by Hildebrandt [52], explainability is only a small part of the justifiability equation for AI systems and may obscure the bigger picture. However, the notion of legal justification of AI systems has not received as much traction so far; explainability has received much more attention. Specifically, "legal explanations", *i.e.* explanations designed to support the legal compliance process, have been examined by XAI researchers [4, 15, 25]. The requirements of the General Data Protection Regulation (GDPR) [30] to provide users with "meaningful information about the logic involved" have received much attention from explainability researchers [10, 20, 24, 47, 48]. Recent work reviews in detail the legal requirements for explainable AI [10, 24, 88, 93]. Naninni *et al.* [88] highlight that regulations are informed by coarse notions of explanations. Nevertheless, Doshi-Velez *et al.* [24] argue that "legal explanations" are technically feasible, mainly through local explanations and counterfactuals. Bibal *et al.* [10] presents four levels of explanations to meet the different types of requirements: explanation of the main features, of all features, of the features involved in a decision, and of the whole model. However, this interdisciplinary body of work, has not yet adopted a user-centric approach to study the needs of regulators, who are the main end-users of such "legal explanations"[8].

2.4 Auditing AI systems

Some work has emerged to define AI auditing and its role in relation to traditional audits [81, 98, 105] or to outline audit approaches and principles [56, 87, 94, 98]. Sandvig *et al.* [98] first introduced the notion of algorithm audit, with the application of Internet platforms algorithms in mind. Mökander *et al.* [87] summarized the promise of AI auditing in three ideas: it is procedurally regular and transparent, it enables proactivity in addressing AI harms, and it is conducted by independent parties. Koshiyama *et al.* [56] give four main verticals of algorithm auditing: performance and robustness, bias and discrimination, explainability, and privacy. The first vertical encompasses concepts such as resilience to attacks, fallback plan, accuracy, reliability, and reproducibility. They define seven levels of explainability, corresponding to increasing levels of access to information up to the complete "white-box" setup. Raji *et al.* [95] drew lessons for AI auditing from industries including finance. The authors discuss the historical role of internal audits in this domain and their focus on organisational aspects and risks. They also consider financial auditing to be "lagging behind the process of technology-enabled financialisation of markets and firms". The literature on AI auditing is still in its infancy [32], and has so far only focused on definitions and methodological aspects of audits from a theoretical point of view. Very little research has offered qualitative empirical insights on the socio-techno-legal aspects of AI audits.

2.5 The AML-CFT context

2.5.1 Overview. Money laundering is the action of concealing the origin of funds illegally obtained. Terrorist financing is a different process: it involves concealing the destination of funds by raising, storing, moving, and using the money [63]. To detect these financial crimes, AML-CFT laws require banks to carefully control with whom they are engaging in a business relationship and to actively monitor their customers' transactions [9]. This implies that banks map out the money laundering risks to which they are exposed,

[2]Definition adapted from [81, 87, 98]

taking into account their activities and customers, and putting in place a detection system, including an often automated "transaction monitoring system" that flags unusual activities. In general, this rule-based approach begins with an alert first triggered from an automated system usually based on rules (such as "transaction is superior to a certain amount"), then it is quickly reviewed by a human analyst and either closed or passed on to a second level of review. If the alert is still considered suspicious at this stage, a case is created and a more extensive investigation is opened to be reviewed by more experienced analysts. If the suspicion is confirmed, it is reported to the national financial investigative body—TRACFIN in France—which conducts a deeper investigation [54]. If there is evidence of a financial offence, the case is passed on to the law enforcement authorities [3].

2.5.2 Legal requirements. AML-CFT laws propose a risk-based approach, meaning that banks have to identify the risks they are exposed to and take appropriate measures to mitigate them [34]. The risk-based approach to AML-CFT is widely adopted and has been recommended by the Financial Action Task Force (FATF), the intergovernmental organization dedicated to combating money laundering and the financing of terrorism, to its 39 members, which includes 24 non-EU countries [35]. It is also the standard approach in Europe having been recommended by the European Banking Authority [28].

The banking sector also has "internal control" obligations that constitute a set of safeguards enabling financial institutions to control the risks of their activities [95, 101]. EU countries are subject to such requirements under Directive 2013/36/EU. Under these requirements, banks have to implement three "lines of defense" to ensure that their financial activities remain legal: level one corresponds to the day-to-day business operators; level two requires a separate unit responsible for monitoring level one; level three is an audit team that intervenes periodically. If banks fail to comply with these obligations, they can face heavy fines by the national supervisory authority. In France, these fines amounted to several million euros between 2016-2021, sometimes amounting up to 6.5% of the fined banks' revenues [21].

2.5.3 The role of supervisors. Supervisors are agents of regulation. In France, their role is laid down in the regulation[4], and described on the French Regulator's website[5]. Supervisors monitor the compliance of financial institutions with European and national AML-CFT laws. They also influence the development of AML-CFT frameworks by synthesizing gaps, threats, and best practices at the national level. For example, the French supervisor annually reports on the threat posed by money laundering and terrorist financing and often publishes guidelines and thematic reviews detailing the supervisor's expectations and interpretations of the law.

2.5.4 AI for AML-CFT. Banks have only recently begun to explore the use of machine learning in AML-CFT, but it is one of the most impactful applications of AI in banking [39]. AI development is mainly due to two factors. Firstly, AI promises better performance than traditional detection systems, which are based on known scenarios of money-laundering schemes. The most promising use is through unsupervised and reinforced learning that have the potential to detect anomalies which shed light on typologies of money laundering that have not been previously reported [13]. AI can also help set smarter alert thresholds, help human analysts prioritize alert treatment, and enhance the quality and diversity of the data used in criminal investigations [17, 59, 61, 69, 89]. Secondly, AI enables banks to cut costs by alleviating repetitive tasks and reducing the human staff required to review alerts [91, 100].

However, AI is still a relatively recent topic in AML-CFT, and AI-based systems have been subject to few, if any, regulatory audits to date. So far, only a handful of national supervisory authorities have expressed positions on AI. In 2018, the Monetary Authority of Singapore stated to be "in agreement that such advanced technologies can and should be leveraged by banks" [100]. A report on AI for AML in Norway, however, argues that banks "as well as regulators have historically been reluctant to use AI" [91]. The Dutch Central Bank (DNB), in November 2022, was hesitant over machine learning technologies for AML as illustrated in a regulatory sanction [11] but has since cautiously opened the door for its use [53, 100]. The French supervisor has not yet expressed clear guidance on AI but has been generally open to the technology. They have also developed an internal AI-based tool to challenge the performance of banks' systems [62].

2.5.5 Explainability and transparency in AML-CFT. Explainability (XAI) has often been presented as a requirement to meet compliance standards in AML-CFT [1, 5, 39, 42]. In her 2022 speech about technologies to fight financial crime, Elizabeth McCaul, member of the Supervisory Board of the European Central Bank (ECB), presented explainability and transparency as "two of the most important challenges for AI" [77]. However, the specific requirements for explainability and transparency remain vague and general. It is not yet clear which precise legal requirements they would fulfill.

Nevertheless, several efforts to build explainability solutions have emerged in AML-CFT over the past few years. According to Kute *et al.*'s review of AI solutions in AML-CFT [60], 51% of the scientific papers that present a machine learning method for AML also consider the explainability of their solution, such as knowledge-graphs rule-based reasoning approaches [5]. Weber *et al.* [114] identify case studies from the literature where AI and XAI were successfully applied in real financial contexts. The paper also stresses that XAI in AML is under-explored. However, the majority of these contributions are in computer science and do not consider the complex realities of the AML-CFT context.

Some studies have provided more detail on users needs for explainability in AML-CFT. Recent work has emphasized the need to understand why an AI model raised an alert and understand the main features that drove the decision, for the banks' investigators and the national financial investigative bodies [1, 5, 17, 19, 42]. The purpose of this explanation is to provide sufficient evidence about the suspiciousness of a case [60]. Gerlings *et al.* [42] investigated the need for XAI in AML-CFT for banks' investigators and capacity planners. They highlighted the need to explain the reasons for automatic closures of alerts and demonstrate the risk of bias when the scoring of an alert was made visible to the investigators.

[3] c.f. Figure 1 in [60].
[4] In Articles L561-36 to L561-44 of the French Monetary Code.
[5] https://acpr.banque-france.fr/controler/lutte-contre-le-blanchiment-des-capitaux-et-le-financement-du-terrorisme/presentation-du-controle-lcb-ft

However, very few studies have explored user needs from the perspective of supervisors. While Gerlings *et al.* [42] hypothesize that "auditors may require additional information on the model logic", they do not describe the supervisor's explainability requirements in more detail. Kuiper *et al.* [58] explored the perspectives of banks and supervisors in the Netherlands regarding explainability in three financial domains, including AML-CFT. They found that supervisors expected explanations to have a broader scope than banking practitioners, who have a more technical and local understanding of explainability. They did not, however, detail the goals and needs of supervisors for explanations nor justifications and did not consider the legal requirements supervisors expect to see in model explanations.

3 METHODS

This section presents the qualitative methods we used to understand the socio-techno-legal supervision system in AML-CFT and supervisors' needs for model justifiability. We first conducted five semi-structured, scenario-based workshops of two to three participants with 13 supervisors in total. At the beginning of our research, we had initially planned to study the need for transparency and explanation of the models, both for the supervisory authorities and for the banks, but we shifted our focus early on to the supervisory authorities in order to provide a more targeted and in-depth analysis. We nevertheless ran one workshop with participants from a large French bank, which improved our understanding of the existing supervisory mechanism from another perspective: that of regulated entities.

During the workshops, we observed that the participants, particularly the supervisors, consistently referred to legal requirements or regulatory sanction cases when asked about the questions they had about the AI systems and the explanations or justifications they wished to see. This prompted us to find out more about the AML-CFT laws that participants referenced. Additionally, we noticed that the existing scientific or grey literature did not clearly indicate which legal requirements could undermine the use of AI. For that reason, we adjusted our initial research questions and added the RQ2 on how AI opacity conflicts with compliance requirements.

We present below the different methodological building blocks we used in the study, presented in chronological order of implementation. First, we present the procedure, artifacts used, and analysis for the workshops. We then present the methodology we used to complement the analysis of the workshops with regulation-driven needs for algorithmic justifiability. Lastly, we present our findings in post-analysis interviews with two experts in AML-CFT regulation.

3.1 Scenario-based semi-structured workshops

3.1.1 Procedure. All workshops were held in person at the participants' workplace and lasted between 90 and 100 minutes. Participants were not compensated. Upon their arrival, participants were asked to read and complete a paper consent form. The consent form included a description of the purpose and possible risks (mainly confidentiality) of the study, the mitigating measures we implemented to ensure the confidentiality of the recordings and data presented in a publication, and finally their consent to voluntarily participate in this research and to be recorded. They were then asked to answer preliminary questions about their expertise in AML-CFT and their familiarity with AI on a printed form. The interviewer then detailed the workshop agenda.

The workshop questions focused on 4 main themes. First, participants were asked about the existing compliance procedure in AML-CFT in their profession (either controllers or bank practitioners). The following questions addressed the use of AI in AML-CFT to understand participants' impressions of AI. We originally planned this to find out more about how banking supervisors and practitioners envisage AI's future in AML-CFT. However, as the French supervisors were about to publish their position on AI at the time of the study, they considered this information to be too sensitive. We therefore limited the scope of our research to justifiability and explainability needs. We then presented participants with a scenario in which a supervisor controlled an AI-enhanced transaction monitoring system. We asked participants which kind of questions they had about the AI system and what kind of justifications they wanted to see. This scenario-based elicitation approach was used in prior research to understand users' needs for justifications and explanations [64, 65, 96, 102, 116]. Finally, conceptual design artifacts [40] of different explanations and justifications were presented to the participants for fictitious alerts. Participants were invited to discuss the relevance of the justifications and their limitations. As seen in Section 2.5.4, AI's entrance in AML-CFT is a recent topic where regulatory thinking has not yet matured. Some of the questions therefore called for speculative thinking. For this reason, we chose to interview the participants in small groups, so that they could discuss these issues together [86].

3.1.2 Participants. One of the authors had several connections at the French Supervisory Authority to help contact the appropriate directors to obtain the necessary approvals to carry out the research and to connect with controllers. We also learned that the French Supervisory Authority has two departments, one for ongoing monitoring of all financial institutions registered in France and one dedicated to on-site inspections. We used the email lists for these two departments to recruit participants, describing the purpose of the research, the time, location, and agenda of the workshops. In total, we recruited 13 controllers from the French supervisory authority, 6 from the on-site inspections department and 7 from the on-going monitoring department. They had between 1 and 20 years of experience in AML-CFT supervision and their level of familiarity in AI averaged 3.6 out of a Likert scale of 7; two participants had extensive expertise in AI—familiarity level with AI was 7/7.

The participants from the large French bank were recruited by a contact the authors had at the bank with a specific selection criteria for the participants, *i.e.* people specialising in AML-CFT with some previous exposure to AI and, if possible, also to supervisory compliance. In total, six participants took part in the workshop. Three participants' expertise was AML-CFT compliance. The other three participants came from machine learning model development. Naturally, the participants in this study spoke in their individual capacity and their views do not represent the official positions of either the French Supervisory Authority or the Bank that employed them.

Of the 6 workshops, 4 were recorded and 2 were not as some participants did not feel comfortable with being recorded, notably

due to the sensitivity of AML-CFT. However, participants who did not want to be recorded agreed to the interviewer writing notes. One of the unrecorded workshops was with controllers with extensive AI experience, the other was the workshop with banking actors. All participants were French and the quotes presented in this paper were translated from French into English by the authors. Table 4 in the Appendix details the profile of participants.

3.1.3 Artifacts provided. The scenarios featured a fictional character, Eric, whose role was either a controller carrying an on-site mission at a Bank B (for supervisors) or Bank B's head of compliance (for banking practitioners).

We designed two scenarios involving two types of AI-enhanced transaction monitoring systems which have been presented as the most common applications of AI in the scientific literature [13, 42] and in reports from the French supervisory authority [3, 25]. In the first scenario, an unsupervised learning algorithm is used to detect new typologies of financial crime. This algorithm triggers alerts when it identifies a transaction as unusual for certain groups of customers that it has defined. Those alerts come in addition to the ones generated by the bank's traditional rule-based system, which generates alerts based on predefined rules or "scenarios", *e.g. "transaction for this specific customer group is superior to $10.000"*. When an alert is generated, a human analyst examines it and determines whether the identified risk should be addressed by the creation of a new rule in the traditional alert system. The second AI use case involved scoring alerts from Bank B's transaction monitoring system in order to prioritise, redirect, or close them. For high-scored alerts, a Suspicious Activity Report (SAR) was pre-filled automatically with generic information to be sent quickly to the Financial Intelligence Unit (FIU). Only one scenario was used in each workshop. The first use case was used in three workshops and the second in the other three.

For each scenario, we described fictional example alerts triggered by the AI-enhanced AML-CFT system. For instance, the example alert for the first scenario was an alert triggered by the unsupervised AI module. An example alert for the second scenario was an alert considered as low risk and closed by the AI. For these examples, we designed conceptual artifacts [40] of different types of justifications and explanations. Our aim was to encourage participants to comment and imagine possible transparency solutions. We tried to balance the concreteness and openness of these artifacts and to leverage multiplicity in order to get feedback on the concept of these justifications rather than on their design. We chose to show the following justifications and explanations based on what we considered as most common in the literature on XAI for AML-CFT [38, 58, 60, 114].

- **a visualisation of the context** of the alert in the form of graph networks
- **a feature-based explanation** showing the most important variables for the AI-produced decision, their impact (positive or negative) and their weight
- **an uncertainty estimator** showing the probability of the alert to be suspect, as calculated by the algorithm
- **a model documentation structure**, including examples of sections: role of the AI system, training data used, performance evaluations, and choice of parameters.
- **an example-based explanation** presenting similar cases and their outcomes.
- **a certification** of the design, development, evaluation and maintenance of the model by an external body. We added this artifact because it is one of the provisions in the upcoming AI Act relating to high-risk AI systems.

Figure 1 presents the scenarios we showed to participants. The conceptual justification artifacts are presented in Figure 4 of the Appendix.

3.1.4 Analysis. We used a content analysis methodology [6] to analyse the audio transcriptions—including question-answering and think-aloud data—and the notes taken from the workshops. The notes were taken by the interviewer during the workshops and we recognise their limitations. Although they cannot reflect the details and nuances of the participants' thoughts and words, the notes nevertheless capture the general and sometimes strong opinions of the participants. The broad themes used for the content analysis followed the workshop structure: (1) the socio-technical context and (2) technical approaches of the supervisory authorities, (3) the AML-CFT legal requirements, (4) supervisors' questions on AI, and (5) ideas for designing AI justifications and explanations. Based on the open codes gathered for each of these five overarching themes, we used axial coding to establish links between the concepts and refine them [22]. The first author, who was also the interviewer and note-taker for the non-recorded workshops, carried out the thematic and axial coding for 5 workshops—three fully transcribed and two partially-transcribed using notes. Another author analysed the audio transcripts of a workshop and applied open thematic coding separately. The two authors then discussed all the codes they had created and refined them on a Miro board[6].

3.2 Empirical legal research

As agents of regulation, supervisors' goals are embedded in the legal requirements they enforce. During the workshops, we observed that not having a full grasp of the various legal themes to which the participants were referring prevented us from capturing their motivations to ask for specific justifications. Therefore, we complemented the scenario-based eliciting approach with qualitative empirical legal research [115]. We believe that combining needs elicitation with a legal analysis is key to fully understanding regulator needs. In fact, the legal field is also keen on qualitative approaches, using interviews and legal document analyses, with methods similar to those used in the social sciences. Webley [115] points out that "many common law practitioners are unaware that they undertake qualitative empirical legal research on a regular basis". We conducted this legal approach in parallel to the analysis of the workshops.

3.2.1 AI Compliance Assessment. Our methodology was adapted to address our research question, as recommended by [115]. It was carried out by the first author, who does not have a legal background, but the methodology and findings were discussed multiple times with another author with extensive experience in legal practice and research. We began by a doctrinal research as described by McConville [78], which consists in seeking what the

[6] https://miro.com/app/dashboard/

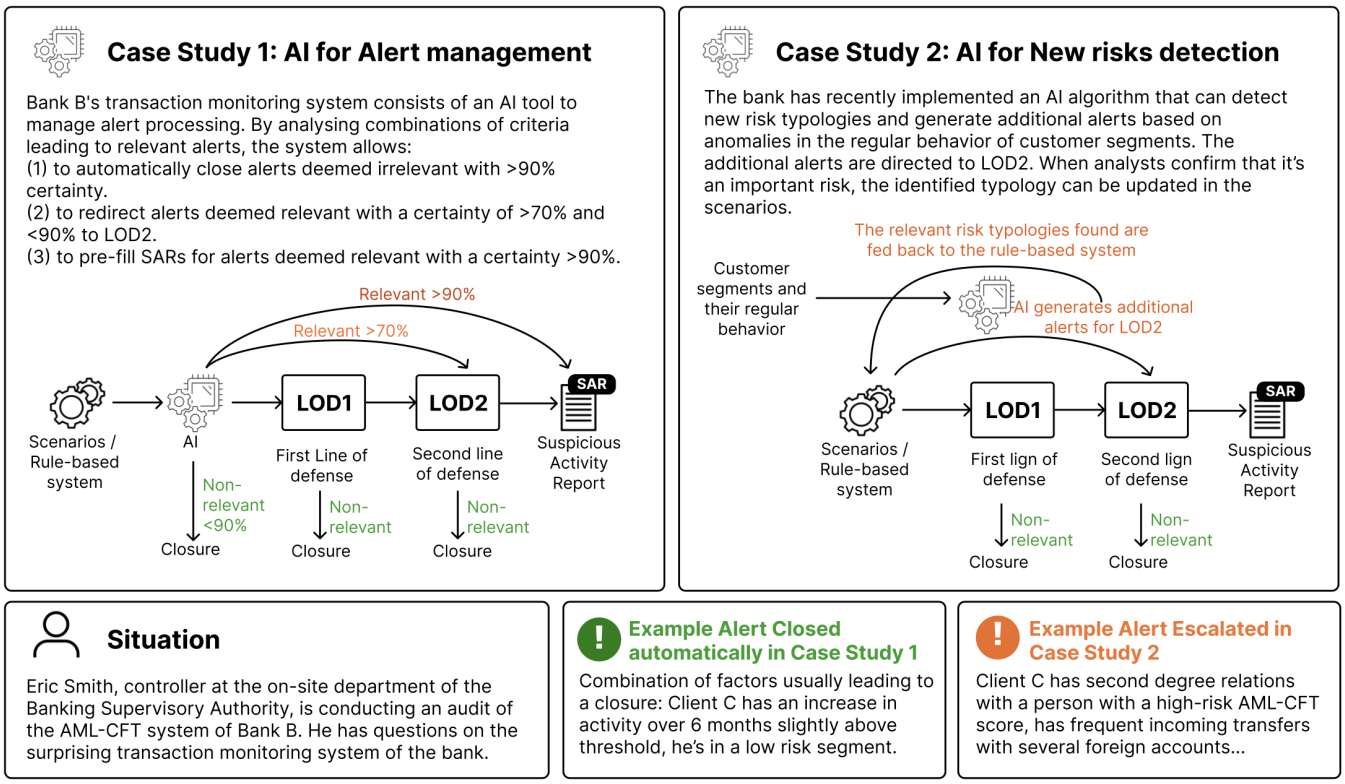

Figure 1: Scenarios used during the workshops with supervisors, with a description of the two use cases of AI in AML-CFT, and two examples of alerts that were generated or closed by the AI-enhanced systems. Only one of these case studies was presented in each workshop.

law is in a particular area. We thus examined regulatory sanction cases on AML-CFT, the relevant articles of the French Monetary Code, and other useful legal documents on the advice of a lawyer from the French Banking Supervisory Authority. The data collected we used for this legal approach is detailed in Table 1. We narrowed our focus on AML-CFT and internal control requirements, as these are the requirements that banks are evaluated against during AML-CFT supervisory audits. We identified the main legal themes and specified their meaning, first using open coding on five regulatory sanction cases, because they reflect how supervisors' interpret and structure AML-CFT laws. We then refined the themes with the rest of the data collected. We used the scenarios we defined in Section 3.1.3 to assess how AI opacity impacts each identified theme. Finally we conducted feedback interviews. In short, our method follows these six steps:

(1) Identify the applicable laws in AML-CFT and define the scope of the research through *"doctrinal research"*.
(2) Define the main themes in the applicable laws, building on the format of the legal documents and invoked themes in the workshops.
(3) Specify the meaning of the requirements in each theme, drawing on the supervisors' perspective and legal documents, such as case law, which inform on how the law is commonly interpreted.
(4) Define scenarios featuring AI systems in AML-CFT.
(5) Consider how the opacity of these systems conflicts with each sub-theme identified, which can also be formulated as goals for which the supervisors seek transparency.
(6) Obtain feedback on our analysis from AML-CFT experts during interviews.

3.2.2 Feedback interviews. Because step 5 of the above methodology can be somewhat subjective and potentially inaccurate due to the lack of expertise of the first author in AML-CFT law, we conducted two interviews to elicit feedback and corrections from experts. The two participants were solicited upon advice from internal contacts at the French supervisory authority, given their unique expertise in both AI and law. One of them was a lawyer and the other an on-site inspector with extensive background in AI. Our pre-interview included a presentation of the research, confidentiality risk mitigation measures, and request to record interviews. We began by asking participants two general questions: what do they see as the key challenges in assessing AI's compliance with AML-CFT requirements, and how does the opacity of AI make compliance with AML-CFT requirements difficult. We then presented them an initial version of Table 3 in the Appendix and asked for feedback. Interviews were used to both correct and complement our prior analyses. Interviews were recorded, transcribed, and two authors

analyzed and coded them according to the process described in Section 3.1.4.

Table 1: Data used for the empirical legal research

Type	Document
Regulatory sanction cases	• Sanction Commission Decision 2022-04 against BMW Finance • Sanction Commission Decision 2022-02 against Financière des paiements électroniques • Sanction Commission Decision 2022-01 against Axa Banque • Sanction Commission Decision 2021-05 of 1 December 2022 against Caisse régionale de Crédit agricole mutuel du Languedoc • Sanction Commission Decision 2021-01 of 1 March 2022 against W-HA
Law, orders	• AML-CFT: Articles L561-1 to L564-2 of the French Monetary and Financial Code [72] • Internal control: French Monetary and Financial Code, Articles L511-55, L522-6, L522-14 and L526-27, Order of November 3rd, 2014 [71].
Soft law	• Joint ACPR and Tracfin guidelines on reporting obligations to TRACFIN • Thematic review: Automated systems for monitoring of AML-CFT transactions
Interviews	• 5 Workshops with 13 supervisors/controllers • 2 Interviews with 2 AI/AML-CFT supervisors

4 FINDINGS

The results presented in this section are structured around three axes, each aimed at improving our understanding of a user group that is under-represented in the literature: regulators, more specifically, supervisors in AML-CFT. The three axes correspond to our research questions: understanding the supervisors' socio-technical context (RQ1), understanding the regulatory goals of supervisors in AMl-CFT (RQ2), and articulating the supervisors' needs for AI justifications and explanations (RQ3).

4.1 Socio-techno-legal context and auditing approaches of supervisors in AML-CFT (RQ1)

Figure 2 provides an overview of the workshop findings and the socio-techno-legal context of supervisors.

4.1.1 How are supervisory audits organized in practice? (socio-organizational context). The French Banking Supervisory Authority carries out two types of inspections: document-based control and on-site.

The document-based control unit's mission is to **assess the maturity of the AML-CFT system of each regulated entity in France (around 1,300)**. This control is based on numerous records, including an AML questionnaire that banks report annually and exchange with the regulated entities. They then notify the banks of their observations. This unit can also suggest on-site inspections, as one participant notes: "*when we see a lot of deficiencies, we will inform the on-site inspection and propose that the establishment be included in the investigation programme*".

The role of on-site inspections is to **confirm the true state of a bank's declarations concerning their system for AML-CFT and to assess their effectiveness**. Inspectors will challenge a bank's system, observe how employees work, compare declarative practices with what actually occurs, exchange information with bank practitioners, and perform IT extractions to identify any major deficiencies within the allotted time for inspection, *i.e.* a few months. One participant emphasised the importance of the iterative process when communicating with banks which helps prevent misunderstandings. Around 40 on-site investigations take place annually [2]. Following the findings of an on-site inspection, a sanctions committee may then be called upon to decide whether a penalty should be imposed. Figure 3 details the anti-money laundering and terrorist financing controls for the French supervisor.

It is worth noting that the large majority of controllers have a legal background with expertise in financial crime analysis. Many participants, therefore, expressed unease with complex statistical tools such as AI. For example, some participants said "*our IT skills are a little limited*" (P3) and expressed their lack of computer science knowledge to deal with the particularities of machine learning models. One of these participants, however, was aware of unsupervised and supervised learning and many participants with little familiarity with AI were able to generally describe the functioning of the AI-based systems they had seen in banks. Moreover, on-site missions include at least one computer scientist to support non-tech controllers. One participant stated "*When you need to go into details, you need to have knowledge, experience or even ideas of what to do. Their [the banking actors'] job and ours is evolving, we'll have to speak both the financial crime and python languages"* (P11).

4.1.2 How do supervisors describe the legal context in AML-CFT, specifically transaction monitoring? Section 2.5.2 provided an objective review of the legal context. Below we give a brief impression of participants' perspectives on these regulations. Supervisors described the AML-CFT regulation as "*prolix*" (P1) and "*subtle, with high expectations and not much room for error*" (P11). Another participant added that "*every system, even the best, does not detect everything*, confirming that a **small margin for errors** is left in transaction monitoring given there is an obligation of implementing the best means and not an obligation of results. Just as there exists a small margin for error for data quality—roughly below 5%—they expect AI tools to also make errors. Supervisor tolerance is qualitative, and depends on error severity and systematicity. It was also noted the regulation does not stipulate a requirement to automate tools. It is instead the size of the regulated entity and its volume of transactions that will drive an implementation of automated "scenarios" and ultimately, AI. One participant noted that "*[Banks] are fairly up to speed with regulation, they will end up on AI one day or another.*"

4.1.3 What are the approaches of supervisors to audit the automated AML-CFT systems in banks (technical context)? Participants emphasized that there is no single approach to auditing; all audits adapt to their context. We identified, however, some common approaches to auditing. Investigations or document-based assessments usually start by examining the risk classification of banks—"*everything flows from the risk classification*". Banks must produce this document, which identifies the money laundering and terrorist financing risks related to the bank's activities, size, customers, etc. Supervisors can then identify gaps in the identified risks, in the risks covered by scenarios, and other automated tools. Then, during controls, supervisors assess the quality and compliance of

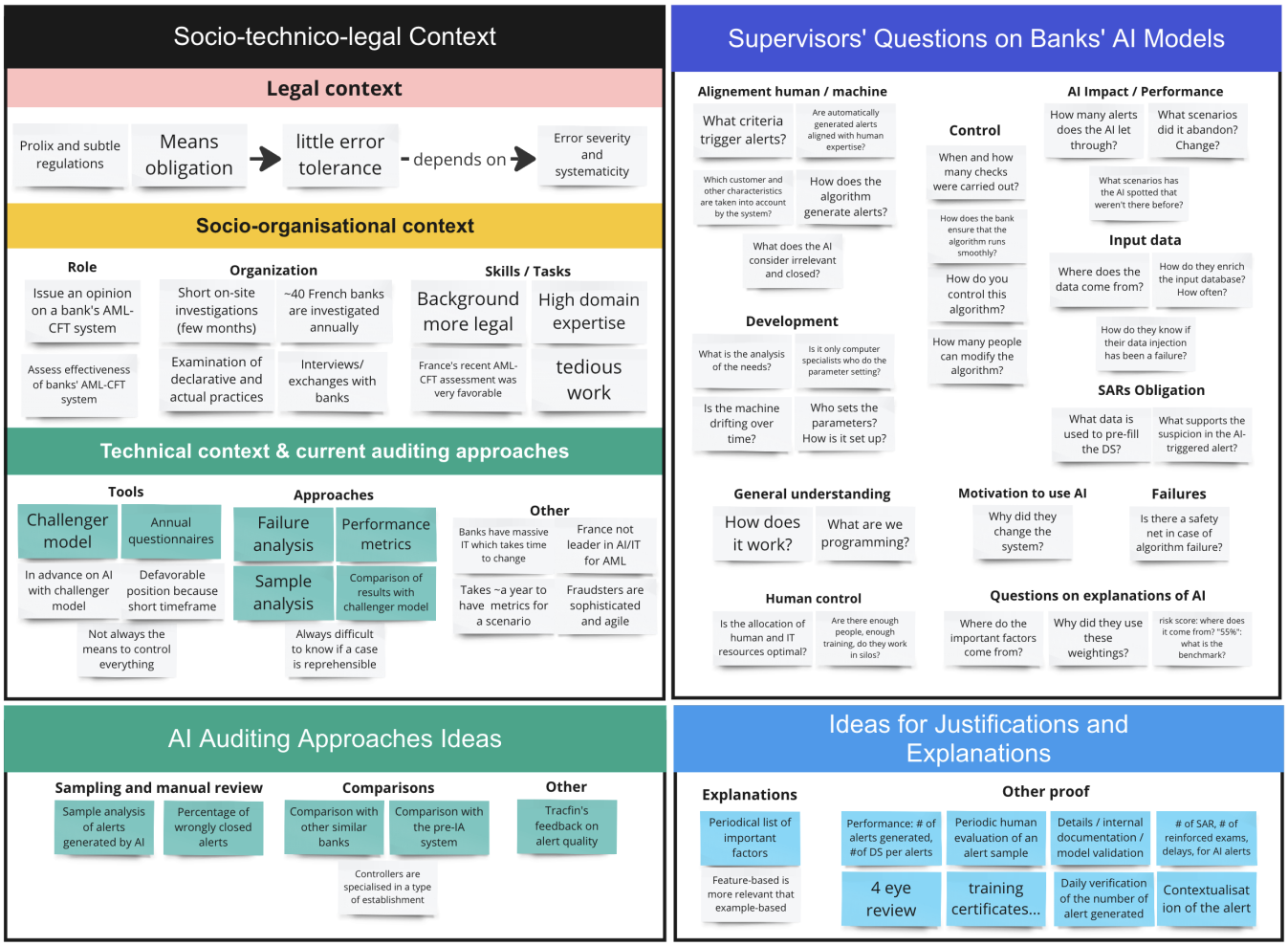

Figure 2: Summary of the workshops, with socio-techno-legal context of supervisors, supervisors' questions on AI, AI auditing approaches ideas, and ideas for justifications and explanations.

two aspects of the bank's AML-CFT systems: processes and results. Approaches to evaluate results may pinpoint failures in the process and vice versa. Audit strategies of AML-CFT frameworks can be broadly summarized in three approaches: "global", "global to local" and "local towards global".

Global approaches consist in looking at metrics characterising the efficiency of AML-CFT devices. These metrics include, for example, the number of alerts generated, the number of reinforced examinations, and the number of SARs. Supervisors interpret these metrics in relation to the bank's characteristics; as a participant notes, "*We'll see if they're consistent with the establishment's activity*" (P3). It takes some time, however, for these measures to reflect the value of a new tool: "*as long as the scenario hasn't really run for a year, we won't have very interesting statistics*" (P4).

Furthermore, a **"global to local approach"** enables controllers to find cases to investigate. The French supervisory authority recently developed an AI-based tool, "LUCIA", to support controllers in sampling cases and comparing them with the bank's results [62]. Participants highlighted time-saving and novel offerings of this tool: "*It makes it possible to review, I don't know, thousands of operations, whereas as an on-site controller we can see a panel of about fifty operations.*" (P8). P1 reported that the work of controllers is often very tedious and stressed the need for tools like LUCIA, *"so that we are in a position, not to anticipate anything, but to react to regulations and perhaps to detect loopholes more easily."* (P1). P7 summarized that the main goal of SupTech tools is to "*enrich the control by giving possibilities or ideas that the analysts would not have had or that they would not have had the means to look at.*" (P7).

Local approaches involve examining specific cases or part of the AML-CFT framework to see if there are any crude errors in reasoning. Examining local cases can also give conclusions about the results. **The "local towards global" approach aims at drawing conclusions on the system from ad-hoc observations**. Supervisors draw on a thread of errors observed in specific cases to trace systematic errors in the system. This is enabled by "failure analyses" or "sample analyses" which consist of examining cases either

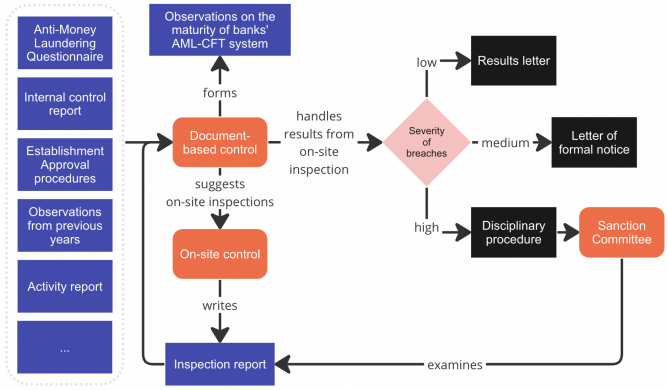

Figure 3: Flow diagram of the supervisor's control procedures in AML-CFT

brought to the attention of supervisors by TRACFIN or another public authority, or drawn from a sampling strategy. Supervisors ask "*should the system have detected [the errors]? Was it within its scope? Was it within its objectives and why didn't it detect them, what went wrong?* " (P14).

Overall, **the superposition of different methods** for auditing and detecting financial crime in banks, whether AI-based or not, improves the efficiency and robustness of the frameworks: "*We know that there will be illegal operations that go undetected. We can't detect everything, but there's an obligation to try and detect as much as possible, and if we start relying solely on AI, well, we're bound to miss things. But we'll miss less if we superimpose different methods*" (P14).

4.2 What provisions in AML-CFT laws does AI opacity conflict with? (RQ2)

This section presents the results of our compliance assessment, the methodology of which was presented in Section 3.2. The paragraphs below present a regulatory goal (RG) with which AI opacity can conflict. Table 3 in the Appendix also provides a summary of this analysis.

4.2.1 Verifying risk adaptation (RG1). As part of compliance requirements, supervisory authorities verify the adequacy and completeness of a bank's operation monitoring system in relation to its risk classification[7]. Much of this assessment is based on a qualitative understanding of the reasoning and criteria used by the system to generate alerts. This enables controllers to verify that important characteristics of the business relationship are considered (e.g., income), or that the thresholds are relevant based on business expertise. The opacity and complexity of AI led some participants to fear that this assessment would become difficult: ***"We're going to end up with this like chickens with a knife*** *and we won't know exactly why it generated this alert...we won't be able to assess the adaptation to the risk."* (P4)

4.2.2 Verifying the bank's ability to perform constant and careful examination (RG2). Supervisors also have to verify that transaction monitoring systems detect inconsistencies with up-to-date customer knowledge and fulfill the bank's obligations of carrying out "careful examinations" of operations[8]. Supervisors typically use performance metrics and a "local to global approach" to evaluate this. As AI algorithms are opaque, however, supervisors may not be able to establish if an ad-hoc error in detecting financial crime is linked to a broader issue in the system. Moreover, clarifying how AI systems adjust to input updates might be needed to comply to constant vigilance obligations.

4.2.3 Verifying the bank's ability to perform "enhanced vigilance", to produce quality Suspicious Activity Reports, and to update their risk classification (RG3). Financial institutions also have the obligation to increase surveillance with regard to complex or risky transactions and to submit high-quality SARs to TRACFIN. As one participant said: "*All alerts must be duly substantiated and analysed*" (P10). This implies that sufficient explanations be given on why a scoring algorithm (as in the first scenario) considers an operation as risky and why an alert was generated by an algorithm (as in the second scenario), so that human analysts can write high-quality SARs: "***We need to be able to understand the criteria that generate a risk.*** *It's a question of auditability. Actually, before that, it's a question of a human analyst's ability to understand what to look at*" (P14).

4.2.4 Verifying that banks can detect incidents and have control over the purpose and operation of any device used (RG4). Internal control obligations require banks to: be able to detect incidents; control the operation of their devices, notably over time; demonstrate control over the purpose of their system, particularly when it is provided by a third party; and plan for safety nets in case of failures [9]. However, AI opacity can prevent banks from correctly **anticipating failures**— "*If you don't know what behaviour is expected, you can't say that there's been a malfunction*" (P10)—or detecting instabilities like drift. The inscrutability of algorithms can also create **dependencies on AI**: "*there is a risk of dependence on AI if the criteria are not understood*" (P7).

4.2.5 Verifying the correct allocation of material and human resources (RG5). AML-CFT laws also require banks to put in place the tools and human resources needed to monitor operations[10]. Case law indicates that it is a question of striking a balance between human and automated tools. AI transparency will be needed to **show how human expertise and AI systems are balanced and complementary**. Many participants insisted that human expertise cannot be replaced in many instances: "*there is a human expertise that cannot be replaced, particularly in advising banks on signs of radicalisation...*" (P1). For that reason, the auto-filling of SARs by AI, if not verified and substantiated by a human, as presented in scenario 1, was seen as problematic. Moreover, explainability can have a major role in enabling transitions between machine and human analysts and to ensure timely processing of the alerts, as P10 noted: "*There may be an impact of explainability on processing*

[7] c.f. Article R. 561-12-1 of the French Monetary Code (CMF) and Decision against AXA Banque of the 15/02/23

[8] c.f. Article L561-6 of the CMF

[9] C.f Article R561-38-4 of the CMF, Order of November 3, 2014

[10] c.f. Article R561-38 CMF

times". Indeed, SARs should be filed without delay so that TRACFIN can bring cases to court as quickly as possible.

4.2.6 Understanding the motivation for AI use (RG6). Some participants, during the semi-structured workshops, were also questioned on whether banks needed to justify the use of AI. Most participants claimed that while it is not legally required, it could help better understand the implemented transaction monitoring system. One participant explained: "***I'd use motivate rather than justify, in other words, the Bank is free to use AI**. On the other hand, it must always be able to motivate, to explain why such change in its system.*" (P7).

4.3 Supervisors' needs for model justifiability in AML-CFT (RQ3)

The summary of the workshops presented in Figure 2 shows the questions that supervisors asked about the AI systems described in the scenarios. Based on the supervisors' regulatory objectives described above and their questions about AI, we formulate supervisor needs for justifiability below.

4.3.1 Understand the basics (N1). Supervisors who are primarily lawyers require high-level explanations or machine-learning training to answer their questions like "How does it work?" and "What are we programming exactly [in machine learning programs]?". They want be able to autonomously use a "Challenger" model, the supervisor's AI model, to assess the banks' systems. As noted by participant P11, **controllers "*have to be able to understand the purpose and operation of the SupTech tools that their IT team implements*"**. Their profession will evolve towards hybrid profiles that are both legal and technical. However, the current challenger model developed by the Supervisor, LUCIA, is designed as a support tool for in-depth analyses. One participant explained: "*Paradoxically, the stakes may not be so high because you get to the stage where you're digging into the details anyway, and then you abstract from the surveillance system.*" (P10).

4.3.2 Demonstrate legitimacy (N2). With LUCIA, supervisors are in an advanced position where AI is challenging traditional rule-based systems. The errors found during this process also highlight the added-value of AI, one participant noted. However, participants from the bank have stressed the **need to be on a level playing field, according to the legitimacy principle of due process rights of regulated companies** ("equality of arms") [90]. For that purpose, they would like to understand the data or methodology used by the supervisor, especially data they do not have access to. Banking professionals also wanted to know if the challenger model was using sensitive data, or if it was discriminatory in any way, as they are entities subject to privacy regulations[11]. Nevertheless, a supervisor pointed out that they are rather at a disadvantage when it comes to finding undetected financial crime, which fuels their need for AI tools: "*the tight time-frame [for investigations four months][12], we need to start everything from scratch each time, the data, everything...*" (P14). Supervisors have implemented question-answering sessions for banks on this issue.

4.3.3 Measure global efficiency (N3). The global approaches described in Section 4.1.3 to measure the AML-CFT framework performance are likely to remain valid for any system, AI or not. One participant indicated that "*Even before AI, the black box phenomenon already existed.*"(P14). In particular, the current sampling strategy by the supervisory authorities is still suited to assess AI-enhanced AML-CFT systems. "*For us, **the most practical and realistic way of checking that this [the system] is not absurd is not to look at the parameterisation.** Because it's difficult to understand the effects of a parameter when it interacts with other parameters. It's a question of seeing in situ **how it behaves in reality when faced with examples that we have selected ourselves.***" (P14). A participant indicated three main approaches envisaged for evaluating global performance of AI-enhanced AMl-CFT systems: (1) compare efficiency with the pre-AI system, potentially comparing performances with similar establishments; (2) analysis of the "failures" reported to the supervisory authority; (3) comparison of the banks' results with the results obtained using a challenger model on sampled cases. The sampling approach was mentioned in all the workshops with supervisors.

P1 and P2 also brainstormed about "*simple, basic*" indicators to measure efficiency, using, for example, the ratio of suspicious transaction reports to turnover "*or something similar*", refined for relevant clusters of similar establishments, potentially made with AI. Aggregated statistics of **this indicator could also be shared with financial institutions to encourage improvement**: "*If we give them the average, they set themselves a performance target which is, I don't know, like, 20% above average.*" (P2)

Another group of participants felt more dismayed by the increasing opacity and complexity of AI systems. They argued for another approach to measure efficiency that relies more on financial intelligence units: "*The standard controller will be completely helpless. We'll have to change the way we monitor, we'll have to work more with the financial intelligence unit, TRACFIN, which will then be the only one able to give an opinion on the alerts*".

4.3.4 Establish reprehensibility (N4). Despite implementing sampling strategies, having a closer look into the AI system inner workings might be necessary to establish the reprehensibility of the errors detected. Understanding why a suspicious transaction was not detected might help conclude on the systematicity, and therefore the reprehensibility of the problem. This requires a contrastive explanation, focusing on the negative which answers questions such as "**why did the system behave in this way (letting the fishy transaction go) and not in this other way (flagging the transaction)?**". One participant described: "*It's the question of how you go from analysing individual declarative failings to making structural observations about the structural failings of the system"* (P10). Banks also need to implement such explanations when implementing anomaly detection AI systems, as in Scenario 2. In this case, the unsupervised algorithm may encounter a risk typology not covered by the bank's traditional system. The bank then has to understand why this risk was not detected and, if necessary, update the risk classification.

[11]The participants from the bank were concerned that LUCIA would use insights coming from comparisons with other banks or sensitive data, but this is not the case. The AI-based supervisory tool only relies on the data provided by the inspected bank [62].

[12]Which is already longer than in some other countries where investigations are sometimes carried out quickly (a few days), the participant noted.

Need	Description and related regulatory goal	Model / XAI Developer	Design ideas for explanations and justifications
N1: General comprehension	Understand how the challenger model works to extract relevant and representative case samples. Have a general understanding of how the bank's algorithm works (RG6).	Challenger and Bank model / Supervisor and Banks	High-level and global explanations, practice using the model and training, description of AI's role.
N2: Ensure legitimacy and efficiency of challenger model	Monitor performance of the challenger model and make banks appreciate the overall workings of the challenger model.	Challenger model / Supervisor	Global explanation, specific question-answering with banks.
N3: Measure efficiency	Measure the performance of the algorithm, not only in absolute terms but also more concretely in a relative way. Linked to (RG1), (RG2), (RG3).	Bank's model / Bank and Supervisor	Performance metrics: delays, number of SARs, number of reinforced examinations, sampling analysis, Tracfin's feedback on alert quality.
N4: Establish the reprehensibility of sampled error cases	Understand why a bank's algorithm **did not** detect a suspicious case, so as to understand if it was an isolated event or part of a bigger pattern: is the error systematic, reprehensible? Linked to (RG1), (RG2), (RG3).	Bank's model / Supervisor	Local feature importance, Conterfactual explanations.
N5: Verify correct use of explainability	Ensure that banking analysts have a clear understanding of the alerts they are required to handle, so that they can produce high-quality analyses. Linked (RQ3), (RQ4), (RG5).	Bank's model / Bank	Justifications that explanations for analysts are present and efficient, alert contextualisation.
N6: Verify human alignment of decision criteria	Verify that the criteria used by AI to generate or escalate alerts are consistent with the risk exposure and aligned with human expertise. Linked to (RG1), (RG6)	Bank's model / Bank	Feature combination used for a few cases with justifications of the weights (divide the full list of features into groups for readability).
N7: Verify model control by the bank	Ensure that the bank's model does not drift over time, that there is no bias. Linked to (RG4).	Bank's model / Bank	Justify the existence and relevance of tests: Periodically draw up a list of important factors, periodic human evaluation of an alert sample.

Table 2: Summary of supervisors' needs for model justifiability, corresponding description, model concerned and developer of justifications/explanations, and justification and explanation design ideas that emerged during the workshops.

4.3.5 Verify and challenge banks' AI understanding (N5, N6, N7). As noted in Section 4.2.3, supervisors may need to examine a bank's explanatory practices to **ensure that analysts are able to understand alerts** and justify their suspicious nature (N6). To that end, justifications based on local feature importance explanations, which would be implemented by banks, have been preferred by participants: "***The feature importance explanation is more interesting than the example-based one**, which is quite limited eventually*" (P7). Bank participants said they were currently testing an explanation based on Shapley values [70]. The contextualisation with graphs networks has also been appreciated by some participants. In the advent where graph neural networks would be used, we can also imagine that graph visualisation will be highly recommended by supervisors, as is the case for digital asset service providers using blockchain, one participant commented. Views regarding uncertainty estimators were divided. One participant mentioned that: "*It is important to know whether the connections made are coincidental or not*" (P14). However, some participants warned against the confirmation bias it can trigger: "*all these very precise indicators create a push-button risk: as soon as there's a lot of red, bang! [the alert is escalated]*" (P9). Bank participants also confirmed they saw investigators fall into this bias when testing explanations.

Supervisors also want to **verify the human alignment of the decision criteria** used by AI systems (N6). Even though the need for explanations of supervisors is more global, they may look for ad-hoc examples of local explanations: "*We're more interested in the global [...] We'll ask them for local, but local examples for specific cases.*" (P7). Supervisors will not only be interested in the explanation, but more importantly in the justification of why or how developers have validated these feature weights: "*The weight has to be less than... OK a priori, but why?*"(P6) ; "***It can be a relatively aggregated explanation**, i.e. we're not trying to go into the details of the calculation, but to identify the main steps" (P8).*

Finally, supervisors also need justifications that **banks control what their AI system is doing** (N7):"*It's the idea that it creates a dependency on the AI and that the day the AI changes or is hacked, we don't notice the change because we don't know what was at the origin?*" Feature-based importance was seen as useful to that goal: "*With the feature importance explanation, we'll be able to assess: are we in agreement with all these factors?*" (P7). Another participant mentioned that justifications, such as the **daily number of alerts generated and periodic human verification of a sample of alerts** could be effective measures to prevent drift. **Documentation was also seen as crucial** for N7 and N6: *"Documentation is*

super-important to check that they master their tools" (P9). Certifications from third parties, however, elicited more cautious responses. Some supervisor participants argued that, if certification was to become the norm for AI models, it would put regulators in the difficult position of having to adjust the scope of their audits. Other participants from the AML department of the supervisory authority said they would ignore this third party accreditation as it infringes upon their role.

5 DISCUSSION

In this section, we discuss the importance of relying on accurate information about AI systems to justify compliance, explanation limits and alternative approaches like tests and challenger models.

5.1 The role of explanations for justifications

In this paper, we saw that regulators mainly seek *justifications* from regulatees, *i.e.* argumentative demonstrations that their AI systems comply with certain legal requirements. Justification is therefore a critical element in the process of enforcing regulations, *i.e.* for auditability and more broadly for accountability [51]. Just like explanation, justification is a process [82]. One participant mentioned the importance of exchanging with regulatees. Another mentioned that "*justifications are meant to be challenged*" (P11).

[48, 49, 52] argued that explanations are not sufficient to justify a decision. Further, Hildebrandt [52] added "we must not allow the discourse of explainability to stand in the way of the question whether a decision is legally justified, which requires a specific type of legal reasons" [49, 52]. Additionally, Henin and Le Métayer [49] precise that "justifications are complete only if they establish a continuous link between the high-level objectives of the [AI system] (the applicable norms, for example non-discrimination, reduction of recidivism rate, or compliance with a given legal requirement) and its implementation". The authors also stress that justifications are "extrinsinc" in the sense that they refer to external norms such as legal requirements.

However, we argue that acceptable justifications about AI systems should also take into account descriptive, intrinsic, and accurate information about the "implementation" of AI models, to establish this "continuous link". Just like explanations may not always be sufficient to ensure the legitimacy of AI systems , information about an AI system's objectives, design choices, or performance may not always be sufficient to justify the proper implementation of AI models. Furthermore, justifications are intended to be challenged and if they do not rely on factual information about algorithms, there is a risk that the question of the legitimacy of an AI system becomes subjective and arbitrary. In their paper about algorithmic audits, Koshyiama *et al.* [56] argued that, without explainability, a decision cannot be duly contested. Explanations may therefore be insufficient, but are necessary, to provide descriptive, accurate and faithful information about the behavior of an algorithm on which to develop a justification.

The list of needs described in Section 4 illustrate why regulators may need justifications from banks in AML-CFT, whether those rely on explainability or on other kinds of proof such as documentation and tests. In AML-CFT, regulators not only assess results but also processes. Therefore, looking at explanations of the inner workings of AI systems, even high-level ones [10, 25], may become necessary not only for banks, but also for supervisors. The needs N1, N2 and N4 in Section 4 reflect this.

5.2 Considering the limits of explanations

However, current XAI techniques may fall short of regulators' expectations to provide accurate and faithful information about AI system's inner workings. As outlined in [47, 48], the fidelity, robustness, and truthfulness of explainability can be limited by the fact that the many features used by complex algorithms are highly correlated. This is a well-studied and strong limitation of feature-based explanations, which make it difficult to comply with legal requirements to indicate the most important factors in a decision [48, 97]. This goes back to the question of the reliance of AI systems on correlations rather than causal relationships. This can be an issue for measuring model performance as well [48].

Another issue with explanations is that they can be misinterpreted by their users due to the technical language they usually use. Ronan *et al.* call it the "transparency fallacy" when explanations are not effectively understood. We saw this in the reaction of some of the participants in this study who were unsettled by the precise weightings given by the feature importance explanations. Moreover, as demonstrated by Gerlings *et al.* [42] and highlighted by some participants, investigators must have access to sufficient information other than explanations, specifically risk scores, or they will fall into confirmation bias. Supervisors will therefore need to verify that the context in which explanations are presented to investigators, or supervisors themselves, takes into account this bias and mitigates it.

Given their mostly legal background, regulators may also be too quick to accept these explanations as trustworthy. Moreover, the argumentative process of transforming explanations into justifications could be used to the advantage of regulated entities to conceal technical inaccuracies. For example, Zhou and Joachim [118] investigate the concept of "malicious justification". They develop a malicious explanation system that replaces the discriminatory factors (*i.e.* race) used by a biased decision model with non-discriminatory factors to defend the decision. Further, they demonstrate that it is almost impossible even for auditors, who have access to all the decisions, to uncover the deception. The authors also highlight that current explanations do not provide answers to questions like: "what factors caused the model to predict X instead of Y?". Yet, as highlighted in Section 4.3.4, supervisors are likely to need such contrastive explanations to establish reprehensibility of failure cases (N4). As a result, regulators may be in a difficult position to evaluate the adequacy of explainable methods developed by banks and may have to develop their own "explainability challenger" toolkit.

5.3 Supporting model performance measurement and testing

To address the limits of explainability to audit AI systems, specifically regarding fairness, Zhou and Joachim [118] suggest that system-wide metrics are more useful. This was overall supported by the supervisors interviewed in this study. In fact, system-wide evaluation is a pillar in the auditing approaches implemented by the

AML-CFT supervisor. This is reflected in the role of the document-based unit: assessing the maturity of banks' AML-CFT systems, and in the new challenger model developed for investigations. Supervisors are therefore more likely to continue on that "global" or "local to global" path, *c.f.* Section 4.1.

In the field of AML-CFT, however, current metrics to evaluate the effectiveness of systems are limited, notably because banks and supervisors do not know the ground truth regarding alerts, *i.e.* whether a suspicious case was actually money laundering or not. Instead, they have to rely on proxies such as number of suspicious activity reports. The supervisor may have more feedback on the ground truth through the financial intelligence unit, but perhaps not to the point that they can calculate the precision of the system, *i.e.* true positives reported to the sum of true positives and false positives. AI's entry in the industry could represent an opportunity for the supervisor to get closer to the financial intelligence unit, as one participant noted.

The consolidation and disclosure of aggregated data such as precision on the performance of AI models from different banks could be useful for the regulated entities' self-assessment and research purposes. In healthcare, the disclosure of a database of AI-based medical technologies with regulatory approvals enabled researchers to point out some AI weaknesses [80]. Further, such initiatives can help respect the due process rights of regulated entities (N2), while striking a balance with advancing the fight against financial crime.

However, this approach does not inform on the false negatives of AML-CFT systems. Challenger models such as LUCIA can do this to some extent by identifying some crimes that have fallen through the cracks. However, they cannot fully measure the true proportion of crime that has not been detected. This calls for relative comparisons instead of absolute ones, such as comparing banks' practices or pre-AI systems as outlined by participants.

Lastly, to verify processes in addition to results, supervisors in this study have proposed some testing and human oversight mechanisms. More advanced testing methods however will have to be developed to prevent risks specific to AI such as drift, discrimination, and over-reliance on AI. Certifications of the model development were seen as overlapping with supervisors' role. Discussions between certification providers and supervisors might be beneficial to talk about best practices, such as standard models for documentation [41, 83], or mathematical proofs that a code is correct, when applicable [49].

In summary, future work could investigate the design of:

- contrastive explanations to help supervisors establish reprehensibility of failure cases (N4),
- meaningful sectorial, system-wide, metrics and databases to compare the efficiency of AI-enhanced systems in relation to each other or to pre-AI systems (N3),
- meaningful tests for AI to support supervisors in verifying correct use of XAI (N5), human alignment of decision criteria (N6) and model drift control (N7).

6 LIMITATIONS

As the scenario-based elicitation task came fairly early in supervisors' thinking about the use and audit of AI, their responses may not include in-depth considerations on the issue. The purpose of this paper was to articulate the needs of supervisors at a time when the use of AI in AML-CFT and investigations into AI-enhanced systems are in their infancy. We recognise that their needs may evolve as AI audits in AML-CFT develop and new regulatory and case law guidance is issued. Moreover, our research results rely on the specific scenarios and artefacts we presented to participants. This may limit the scope and generalisablity of the results. Specifically, we investigated two use cases of AI, which are considered as the most common and promising in the literature, but other AI applications exist [17]. We also limited the number of conceptual explanations and justifications to six to not overwhelm the participants and to respect their time as volunteers. Other explanations could be considered in future explorations with regulators. Further, we described in the methodology section that two workshops were not recorded due to participants' concerns; we are aware that this limits the analysis and findings from those workshops. However, we were able to conduct a recorded interview with one of the participants in an unrecorded workshop, which enabled us to study the views of this person more closely. Finally, as the first author who conducted the legal approach has no legal training, the method remains fairly straightforward, but we did put in place quality controls with another author, who has a legal background, and two AML-CFT experts. We hope this study demonstrates the feasibility and suitability of such an approach for HCI practitioners.

7 CONCLUSION

This paper examines a socio-techno-legal supervision system in a highly-regulated industry, taking the example of the anti-money laundering and countering terrorism financing domain (AML-CFT) in France. We draw on 6 workshops with supervisors and bank practitioners to outline the auditing approaches of AML-CFT supervisors. We then outline AML-CFT compliance requirements which raise clear issues with AI opacity, and draw up a list of seven model justifiability needs for the supervisors, integrating explainability aspects. In particular, we find that supervisors primarily need to measure the performance of the AI-enhanced AML-CFT system. However, supervisors may need contrastive AI explanations to establish the reprehensibility of sampled failure cases, to verify and challenge banks' correct understanding of the AI, and to demonstrate the legitimacy of their challenger model. These needs are intricately linked to the regulations that supervisors enforce, hence the need for a dual interview-based and legal approach. We also present explanations as having a role of "trial evidence" for justifications. We hope that this work will inform future research to design AI justifications for regulators.

ACKNOWLEDGMENTS

The views expressed in this article are exclusively those of the authors and the participants of this study in their personal capacity. They cannot be taken as the views or policies of the ACPR (Autorité de Contrôle Prudentiel et de Résolution) or of the Crédit Agricole. This research is sponsored by the Agence Nationale de la Recherche (ANR)—grant n°ANR-20-CHIA-0023-01—and the XAI4AML Research Chair. We would like to thank all the participants in this study for their time and invaluable insights and Olivier Fliche, Christine Saidani and Julien Uri for their enlightening comments.

REFERENCES

[1] Raghad Al-Shabandar, Gaye Lightbody, Fiona Browne, Jun Liu, Haiying Wang, and Huiru Zheng. 2019. The Application of Artificial Intelligence in Financial Compliance Management. In *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM 2019)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3358331.3358339

[2] Autorité de Contrôle Prudentiel et de Résolution. 2023. *Annual Report of the ACPR 2022.* Technical Report. ACPR, Bank of France. https://acpr.banque-france.fr/sites/default/files/medias/documents/20230524_rapport_annuel_colb_2022.pdf

[3] Autorité de Contrôle Prudentiel et de Résolution. 2023. *Thematic review on automated systems for monitoring AML/CFT transactions.* Technical Report. ACPR, Bank of France. p.13–14 pages. https://acpr.banque-france.fr/dispositifs-automatises-de-surveillance-des-operations-en-matiere-de-lcb-ft

[4] Valérie Beaudouin, Isabelle Bloch, David Bounie, Stéphan Clémençon, Florence d'Alché Buc, James Eagan, Winston Maxwell, Pavlo Mozharovskyi, and Jayneel Parekh. 2020. Flexible and Context-Specific AI Explainability: A Multidisciplinary Approach. http://arxiv.org/abs/2003.07703 arXiv:2003.07703 [cs].

[5] Luigi Bellomarini, Eleonora Laurenza, and Emanuel Sallinger. 2020. Rule-based Anti-Money Laundering in Financial Intelligence Units: Experience and Vision. In *Proceedings of the 14th International Rule Challenge, 4th Doctoral Consortium, and 6th Industry Track @ RuleML+RR 2020.* CEUR Workshop Proceedings, Oslo, Norway, 12.

[6] Mariette Bengtsson. 2016. How to plan and perform a qualitative study using content analysis. *NursingPlus Open* 2 (Jan. 2016), 8–14. https://doi.org/10.1016/j.npls.2016.01.001

[7] Astrid Bertrand, Rafik Belloum, James R. Eagan, and Winston Maxwell. 2022. How Cognitive Biases Affect XAI-assisted Decision-making: A Systematic Review. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society (AIES '22)*. Association for Computing Machinery, New York, NY, USA, 78–91. https://doi.org/10.1145/3514094.3534164

[8] Astrid Bertrand, James R. Eagan, and Winston Maxwell. 2023. Questioning the ability of feature-based explanations to empower non-experts in robo-advised financial decision-making. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 943–958. https://doi.org/10.1145/3593013.3594053

[9] Astrid Bertrand, Winston Maxwell, and Xavier Vamparys. 2021. Do AI-based anti-money laundering (AML) systems violate European fundamental rights? *International Data Privacy Law* 11, 3 (Aug. 2021), 276–293. https://doi.org/10.1093/idpl/ipab010

[10] Adrien Bibal, Michael Lognoul, Alexandre de Streel, and Benoît Frénay. 2021. Legal requirements on explainability in machine learning. *Artificial Intelligence and Law* 29, 2 (June 2021), 149–169. https://doi.org/10.1007/s10506-020-09270-4

[11] Douglas Blakey. 2022. AI in anti money laundering. https://www.retailbankerinternational.com/comment/ai-money-laundering/

[12] Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, Financial Crimes Enforcement Network, National Credit Union Administration, and Office of the Comptroller of the Currency. 2018. *Joint Statement on Innovative Efforts to Combat Money Laundering and Terrorist Financing.* Technical Report. Federal Reserve Board. https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20181203a1.pdf

[13] Ana Isabel Canhoto. 2020. Leveraging machine learning in the global fight against money laundering and terrorism financing: An affordances perspective. *Journal of Business Research* 131 (Oct. 2020), 441–452. https://doi.org/10.1016/j.jbusres.2020.10.012

[14] John M. Carroll. 1997. Chapter 17 - Scenario-Based Design. In *Handbook of Human-Computer Interaction (Second Edition)*, Marting G. Helander, Thomas K. Landauer, and Prasad V. Prabhu (Eds.). North-Holland, Amsterdam, 383–406. https://doi.org/10.1016/B978-044481862-1.50083-2

[15] Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. 2019. Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics* 8, 8 (Aug. 2019), 832. https://doi.org/10.3390/electronics8080832 Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

[16] Larissa Chazette and Kurt Schneider. 2020. Explainability as a non-functional requirement: challenges and recommendations. *Requirements Engineering* 25, 4 (Dec. 2020), 493–514. https://doi.org/10.1007/s00766-020-00333-1

[17] Zhiyuan Chen, Le Dinh Van Khoa, Ee Na Teoh, Amril Nazir, Ettikan Kandasamy Karuppiah, and Kim Sim Lam. 2018. Machine learning techniques for anti-money laundering (AML) solutions in suspicious transaction detection: a review. *Knowledge and Information Systems* 57, 2 (Nov. 2018), 245–285. https://doi.org/10.1007/s10115-017-1144-z

[18] Furui Cheng, Dongyu Liu, Fan Du, Yanna Lin, Alexandra Zytek, Haomin Li, Huamin Qu, and Kalyan Veeramachaneni. 2022. VBridge: Connecting the Dots Between Features and Data to Explain Healthcare Models. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (Jan. 2022), 378–388. https://doi.org/10.1109/TVCG.2021.3114836 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[19] Douglas Cirqueira, Dietmar Nedbal, Markus Helfert, and Marija Bezbradica. 2020. Scenario-Based Requirements Elicitation for User-Centric Explainable AI. In *Machine Learning and Knowledge Extraction (Lecture Notes in Computer Science)*, Andreas Holzinger, Peter Kieseberg, A Min Tjoa, and Edgar Weippl (Eds.). Springer International Publishing, Cham, 321–341.

[20] Roberto Confalonieri, Ludovik Coba, Benedikt Wagner, and Tarek R. Besold. 2021. A historical perspective of explainable Artificial Intelligence. *WIREs Data Mining and Knowledge Discovery* 11, 1 (2021), e1391. https://doi.org/10.1002/widm.1391 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1391.

[21] Conseil d'Orientation pour la lutte contre le blanchiment et le financement du terrorisme. 2023. *Annual Report 2022.* Technical Report. COLB. https://acpr.banque-france.fr/sites/default/files/medias/documents/20230524_rapport_annuel_colb_2022.pdf

[22] Juliet Corbin and Anselm Strauss. 2014. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory.* SAGE Publications, Inc. 2455 Teller Road Thousand Oaks, California 91320. Google-Books-ID: hZ6kBQAAQBAJ.

[23] Valdemar Danry, Pat Pataranutaporn, Yaoli Mao, and Pattie Maes. 2023. Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Questions Improve Human Logical Discernment Accuracy over Causal AI explanations. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3544548.3580672

[24] Finale Doshi-Velez and Mason A. Kortz. 2017. Accountability of AI Under the Law: The Role of Explanation. *Berkman Klein Center for Internet & Society working paper* Berkman Klein Center Working Group on Explanation and the Law (2017), 17. https://dash.harvard.edu/handle/1/34372584 Accepted: 2017-11-21T16:33:48Z Publisher: Berkman Klein Center for Internet & Society.

[25] Laurent Dupont, Olivier Fliche, and Su Yang. 2020. *Governance of Artificial Intelligence in Finance.* Discussion document. ACPR.

[26] Upol Ehsan, Q. Vera Liao, Michael Muller, Mark O. Riedl, and Justin D. Weisz. 2021. Expanding Explainability: Towards Social Transparency in AI systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–19. https://doi.org/10.1145/3411764.3445188

[27] Malin Eiband, Daniel Buschek, and Heinrich Hussmann. 2021. How to Support Users in Understanding Intelligent Systems? Structuring the Discussion. In *26th International Conference on Intelligent User Interfaces (IUI '21)*. Association for Computing Machinery, New York, NY, USA, 120–132. https://doi.org/10.1145/3397481.3450694

[28] European Banking Authority. 2016. *Guidelines on risk based supervision.* Technical Report. EBA. https://www.eba.europa.eu/regulation-and-policy/anti-money-laundering-and-e-money/guidelines-on-risk-based-supervision

[29] European Commission. 2021. Proposal for a Regulation of the European Parliament and of the Council laying down Harmonised Rules on Artificial Intelligence and amending certain Union Legislative Acts. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021PC0206

[30] European Parliament and Council. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). http://data.europa.eu/eli/reg/2016/679/oj/eng Legislative Body: EP, CONSIL.

[31] European Parliament and Council. 2022. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) (Text with EEA relevance). http://data.europa.eu/eli/reg/2022/2065/oj/eng Legislative Body: EP, CONSIL.

[32] Gregory Falco, Ben Shneiderman, Julia Badger, Ryan Carrier, Anton Dahbura, David Danks, Martin Eling, Alwyn Goodloe, Jerry Gupta, Christopher Hart, Marina Jirotka, Henric Johnson, Cara LaPointe, Ashley J. Llorens, Alan K. Mackworth, Carsten Maple, Sigurður Emil Pálsson, Frank Pasquale, Alan Winfield, and Zee Kin Yeong. 2021. Governing AI safety through independent audits. *Nature Machine Intelligence* 3, 7 (July 2021), 566–571. https://doi.org/10.1038/s42256-021-00370-7 Number: 7 Publisher: Nature Publishing Group.

[33] Massimo Felici, Theofrastos Koulouris, and Siani Pearson. 2013. Accountability for Data Governance in Cloud Ecosystems. In *2013 IEEE 5th International Conference on Cloud Computing Technology and Science*, Vol. 2. IEEE, Bristol, UK, 327–332. https://doi.org/10.1109/CloudCom.2013.157

[34] Financial Action Task Force. 2007. *Guidance on the risk-based approach to combating money-laundering and terrorist financing.* Technical Report. FATF. 47 pages.

[35] Financial Action Task Force. 2014. *Risk-Based Approach for the Banking Sector.* Technical Report. FATF. 50 pages. https://www.fatf-gafi.org/en/publications/Fatfrecommendations/Risk-based-approach-banking-sector.html

[36] Financial Conduct Authority. 2019. *Machine learning in UK financial services.* Technical Report. FCA, Bank of England. https://www.bankofengland.co.uk/-/media/boe/files/report/2019/machine-learning-in-uk-financial-services.pdf

[37] Financial Conduct Authority. 2022. *Artificial Intelligence and Machine Learning*. Technical Report DP-5-22. FCA, Bank of England. https://www.bankofengland.co.uk/-/media/boe/files/prudential-regulation/publication/2022/dp5-22--artificial-intelligence-and-machine-learning.pdf

[38] Financial Stability Board. 2017. *Artificial intelligence and machine learning in financial services*. Technical Report. FSB. https://www.fsb.org/wp-content/uploads/P011117.pdf

[39] Sebastian Fritz-Morgenthal, Bernhard Hein, and Jochen Papenbrock. 2022. Financial Risk Management and Explainable, Trustworthy, Responsible AI. *Frontiers in Artificial Intelligence* 5 (2022), 14. https://www.frontiersin.org/articles/10.3389/frai.2022.779799

[40] Bill Gaver and Heather Martin. 2000. Alternatives: exploring information appliances through conceptual design proposals. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems (CHI '00)*. Association for Computing Machinery, New York, NY, USA, 209–216. https://doi.org/10.1145/332040.332433

[41] Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, and Kate Crawford. 2021. Datasheets for Datasets. https://doi.org/10.48550/arXiv.1803.09010 arXiv:1803.09010 [cs].

[42] Julie Gerlings and Ioanna Constantiou. 2022. Machine Learning in Transaction Monitoring: The Prospect of xAI. http://arxiv.org/abs/2210.07648 arXiv:2210.07648 [cs].

[43] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal. 2018. Explaining Explanations: An Overview of Interpretability of Machine Learning. In *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, Turin, Italy, 80–89. https://doi.org/10.1109/DSAA.2018.00018

[44] Maartje M. A. de Graaf and Bertram F. Malle. 2017. How People Explain Action (and Autonomous Intelligent Systems Should Too). In *2017 AAAI Fall Symposium Series*. AAAI, Arlington, Virginia, 8. https://www.aaai.org/ocs/index.php/FSS/FSS17/paper/view/16009

[45] Rob Gruppetta. 2017. Using artificial intelligence to keep criminal funds out of the financial system. https://www.fca.org.uk/news/speeches/using-artificial-intelligence-keep-criminal-funds-out-financial-system

[46] Abhishek Gupta, Dwijendra Nath Dwivedi, and Jigar Shah. 2023. *Artificial Intelligence Applications in Banking and Financial Services: Anti Money Laundering and Compliance*. Springer Nature, Singapore. https://doi.org/10.1007/978-981-99-2571-1

[47] Ronan Hamon, Henrik Junklewitz, and Ignacio Sanchez. 2020. *Robustness and Explainability of Artificial Intelligence*. JRC Technical Report EUR 30040 EN. European Commission Joint Research Center.

[48] Ronan Hamon, Henrik Junklewitz, Ignacio Sanchez, Gianclaudio Malgieri, and Paul De Hert. 2022. Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making. *IEEE Computational Intelligence Magazine* 17, 1 (Feb. 2022), 72–85. https://doi.org/10.1109/MCI.2021.3129960 Conference Name: IEEE Computational Intelligence Magazine.

[49] Clément Henin and Daniel Le Métayer. 2022. Beyond explainability: justifiability and contestability of algorithmic decision systems. *AI & SOCIETY* 37, 4 (Dec. 2022), 1397–1410. https://doi.org/10.1007/s00146-021-01251-8

[50] Christian Herzog. 2022. On the risk of confusing interpretability with explicability. *AI and Ethics* 2, 1 (Feb. 2022), 219–225. https://doi.org/10.1007/s43681-021-00121-9

[51] High-Level Expert Group on AI (HLEG). 2019. *Ethics guidelines for trustworthy AI | Shaping Europe's digital future*. Technical Report. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[52] Mireille Hildebrandt. 2019. Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning. *Theoretical Inquiries in Law* 20, 1 (Jan. 2019), 83–121. https://doi.org/10.1515/til-2019-0004 Publisher: De Gruyter.

[53] Marit Hoegen, Hilko van Rooijen, and Maarten Rijssenbeek. 2023. Three fundamental changes to the Dutch AML system. https://www2.deloitte.com/nl/nl/pages/finance/articles/three-fundamental-changes-to-the-dutch-aml-system.html

[54] M. Jullum, A. Løland, R.B. Huseby, G. Ånonsen, and J. Lorentzen. 2020. Detecting money laundering transactions with machine learning. *Journal of Money Laundering Control* 23, 1 (2020), 173–186. https://doi.org/10.1108/JMLC-07-2019-0055

[55] Sunnie S. Y. Kim, Elizabeth Anne Watkins, Olga Russakovsky, Ruth Fong, and Andrés Monroy-Hernández. 2023. "Help Me Help the AI": Understanding How Explainability Can Support Human-AI Interaction. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3544548.3581001

[56] Adriano Koshiyama, Emre Kazim, Philip Treleaven, Pete Rai, Lukasz Szpruch, Giles Pavey, Ghazi Ahamat, Franziska Leutner, Randy Goebel, Andrew Knight, Janet Adams, Christina Hitrova, Jeremy Barnett, Parashkev Nachev, David Barber, Tomas Chamorro-Premuzic, Konstantin Klemmer, Miro Gregorovic, Shakeel Khan, and Elizabeth Lomas. 2021. Towards Algorithm Auditing: A Survey on Managing Legal, Ethical and Technological Risks of AI, ML and Associated Algorithms. *SSRN Electronic Journal* (2021), 31. https://doi.org/10.2139/ssrn.3778998

[57] Luisa Kruse, Nico Wunderlich, and Roman Beck. 2019. Artificial Intelligence for the Financial Services Industry: What Challenges Organizations to Succeed. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*. ScholarSpace, Hawaii, 10. http://hdl.handle.net/10125/60075

[58] Ouren Kuiper, Martin van den Berg, Joost van der Burgt, and Stefan Leijnen. 2021. Exploring explainable AI in the financial sector: Perspectives of banks and supervisory authorities. In *Artificial Intelligence and Machine Learning: 33rd Benelux Conference on Artificial Intelligence*. Springer, Esch-sur-Alzette, Luxembourg, 105–119.

[59] E. Kurshan and H. Shen. 2021. Graph Computing for Financial Crime and Fraud Detection: Trends, Challenges and Outlook. https://doi.org/10.48550/arXiv.2103.03227 arXiv:2103.03227 [cs].

[60] Dattatray Vishnu Kute, Biswajeet Pradhan, Nagesh Shukla, and Abdullah Alamri. 2021. Deep Learning and Explainable Artificial Intelligence Techniques Applied for Detecting Money Laundering–A Critical Review. *IEEE Access* 9 (2021), 82300–82317. https://doi.org/10.1109/ACCESS.2021.3086230 Conference Name: IEEE Access.

[61] Nevine Makram Labib, Mohammed Abo Rizka, and Amr Ehab Muhammed Shokry. 2020. Survey of Machine Learning Approaches of Anti-money Laundering Techniques to Counter Terrorism Finance. In *Internet of Things—Applications and Future (Lecture Notes in Networks and Systems)*, Atef Zaki Ghalwash, Nashaat El Khameesy, Dalia A. Magdi, and Amit Joshi (Eds.). Springer, Singapore, 73–87.

[62] Matthias Laporte. 2021. ACPR Conference, p.85, "LUCIA": a SupTech tool to support the fight against money laundering and terrorism financing. https://acpr.banque-france.fr/sites/default/files/media/2022/11/15/20211126_presentations_des_intervenants_de_la_matinee.pdf

[63] Michael Levi and Peter Reuter. 2006. Money Laundering. *Crime and Justice* 34 (Jan. 2006), 289–375. https://doi.org/10.1086/501508 Publisher: The University of Chicago Press.

[64] Q. Vera Liao, Daniel Gruen, and Sarah Miller. 2020. Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376590

[65] Q. Vera Liao, Hariharan Subramonyam, Jennifer Wang, and Jennifer Wortman Vaughan. 2023. Designerly Understanding: Information Needs for Model Transparency to Support Design Ideation for AI-Powered User Experience. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–21. https://doi.org/10.1145/3544548.3580652

[66] Q. Vera Liao and Kush R. Varshney. 2022. Human-Centered Explainable AI (XAI): From Algorithms to User Experiences. https://doi.org/10.48550/arXiv.2110.10790 arXiv:2110.10790 [cs].

[67] Brian Y. Lim and Anind K. Dey. 2009. Assessing demand for intelligibility in context-aware applications. In *Proceedings of the 11th international conference on Ubiquitous computing (UbiComp '09)*. Association for Computing Machinery, New York, NY, USA, 195–204. https://doi.org/10.1145/1620545.1620576

[68] Tania Lombrozo. 2006. The structure and function of explanations. *Trends in Cognitive Sciences* 10, 10 (Oct. 2006), 464–470. https://doi.org/10.1016/j.tics.2006.08.004

[69] Joana Lorenz, Maria Inês Silva, David Aparício, João Tiago Ascensão, and Pedro Bizarro. 2021. Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity. In *Proceedings of the First ACM International Conference on AI in Finance (ICAIF '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. https://doi.org/10.1145/3383455.3422549

[70] Scott M. Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 4768–4777.

[71] Légifrance. 2023. Arrêté du 3 novembre 2014 relatif au contrôle interne des entreprises du secteur de la banque, des services de paiement et des services d'investissement soumises au contrôle de l'Autorité de contrôle prudentiel et de résolution. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000029700770

[72] Légifrance. 2023. Chapitre Ier : Obligations relatives à la lutte contre le blanchiment des capitaux et le financement du terrorisme (Articles L561-1 à L561-50). https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000006072026/LEGISCTA000006154830/

[73] Wendy E. Mackay and Anne-Laure Fayard. 1997. HCI, natural science and design: a framework for triangulation across disciplines. In *Proceedings of the 2nd conference on Designing interactive systems: processes, practices, methods, and techniques (DIS '97)*. Association for Computing Machinery, New York, NY, USA, 223–234. https://doi.org/10.1145/263552.263612

[74] Nicholas Maltbie, Nan Niu, Matthew Van Doren, and Reese Johnson. 2021. XAI tools in the public sector: a case study on predicting combined sewer overflows. In *Proceedings of the 29th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE*

2021). Association for Computing Machinery, New York, NY, USA, 1032–1044. https://doi.org/10.1145/3468264.3468547

[75] Aniek F. Markus, Jan A. Kors, and Peter R. Rijnbeek. 2021. The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies. *Journal of Biomedical Informatics* 113 (Jan. 2021), 103655. https://doi.org/10.1016/j.jbi.2020.103655

[76] Winston Maxwell and Bruno Dumas. 2023. *Meaningful XAI based on user-centric design methodology: Combining legal and human-computer interaction (HCI) approaches to achieve meaningful algorithmic explainability.* Technical Report. CERRE.

[77] Elizabeth McCaul. 2022. Technology is neither good nor bad, but humans make it so. https://www.bankingsupervision.europa.eu/press/speeches/date/2022/html/ssm.sp220713~73f22a486e.en.html

[78] Mike McConville and Wing Hong Chui. 2017. *Research Methods for Law* (2nd edition ed.). Edinburgh University Press, JSTOR. https://www.jstor.org/stable/10.3366/j.ctt1g0b16n

[79] Jessie McWaters and Matthew Blake. 2019. *Navigating Uncharted Waters: A Roadmap to Responsible Innovation with AI in Financial Services. Part of the Future of Financial Services Series. World Economic Forum.* Technical Report. World Economic Forum. https://www3.weforum.org/docs/WEF_Navigating_Uncharted_Waters_Report.pdf

[80] Bertalan Meskó and Eric J. Topol. 2023. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *npj Digital Medicine* 6, 1 (July 2023), 1–6. https://doi.org/10.1038/s41746-023-00873-0 Number: 1 Publisher: Nature Publishing Group.

[81] Danaë Metaxa, Joon Sung Park, Ronald E. Robertson, Karrie Karahalios, Christo Wilson, Jeff Hancock, and Christian Sandvig. 2021. Auditing Algorithms: Understanding Algorithmic Systems from the Outside In. *Foundations and Trends® in Human–Computer Interaction* 14, 4 (2021), 272–344. https://doi.org/10.1561/1100000083

[82] Tim Miller. 2019. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence* 267 (Feb. 2019), 1–38. https://doi.org/10.1016/j.artint.2018.07.007

[83] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model Cards for Model Reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 220–229. https://doi.org/10.1145/3287560.3287596

[84] Brent Mittelstadt. 2019. Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence* 1, 11 (Nov. 2019), 501–507. https://doi.org/10.1038/s42256-019-0114-4 Number: 11 Publisher: Nature Publishing Group.

[85] Sina Mohseni, Niloofar Zarei, and Eric D. Ragan. 2021. A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. *ACM Transactions on Interactive Intelligent Systems* 11, 3-4 (Sept. 2021), 24:1–24:45. https://doi.org/10.1145/3387166

[86] David L. Morgan. 1996. Focus Groups. *Annual Review of Sociology* 22, 1 (1996), 129–152. https://doi.org/10.1146/annurev.soc.22.1.129 _eprint: https://doi.org/10.1146/annurev.soc.22.1.129.

[87] Jakob Mökander, Jonas Schuett, Hannah Rose Kirk, and Luciano Floridi. 2023. Auditing large language models: a three-layered approach. *AI and Ethics* 3, 2 (May 2023), 31. https://doi.org/10.1007/s43681-023-00289-2

[88] Luca Nannini, Agathe Balayn, and Adam Leon Smith. 2023. Explainability in AI Policies: A Critical Review of Communications, Reports, Regulations, and Standards in the EU, US, and UK. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 1198–1212. https://doi.org/10.1145/3593013.3594074

[89] E. W. T. Ngai, Yong Hu, Y. H. Wong, Yijun Chen, and Xin Sun. 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems* 50, 3 (Feb. 2011), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

[90] OECD. 2021. *OECD Business and Finance Outlook 2021: AI in Business and Finance, Chapter 5: The use of SupTech to enhance market supervision and integrity.* OECD. https://doi.org/10.1787/ba682899-en

[91] Erik Overrein. 2020. How machine learning can dramatically reduce financial institutions' cost of compliance. https://www.bearingpoint.com/en-no/insights-events/insights/machine-learning-is-the-key-to-efficient-and-effective-aml/

[92] Cecilia Panigutti, Andrea Beretta, Daniele Fadda, Fosca Giannotti, Dino Pedreschi, Alan Perotti, and Salvatore Rinzivillo. 2023. Co-design of Human-centered, Explainable AI for Clinical Decision Support. *ACM Transactions on Interactive Intelligent Systems* 13, 4 (2023), 21:1–21:35. https://doi.org/10.1145/3587271

[93] Cecilia Panigutti, Ronan Hamon, Isabelle Hupont, David Fernandez Llorca, Delia Fano Yela, Henrik Junklewitz, Salvatore Scalzo, Gabriele Mazzini, Ignacio Sanchez, Josep Soler Garrido, and Emilia Gomez. 2023. The role of explainable AI in the context of the AI Act. In *2023 ACM Conference on Fairness, Accountability, and Transparency*. ACM, Chicago IL USA, 1139–1150. https://doi.org/10.1145/3593013.3594069

[94] Inioluwa Deborah Raji and Joy Buolamwini. 2019. Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*. ACM, Honolulu HI USA, 429–435. https://doi.org/10.1145/3306618.3314244

[95] Inioluwa Deborah Raji, Andrew Smart, Rebecca N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 33–44. https://doi.org/10.1145/3351095.3372873

[96] Mary Beth Rosson and John M. Carroll. 2009. Scenario-based design. In *Human-computer Interaction* (1st edition ed.). CRC Press, Boca Raton, 20. https://doi.org/10.1201/9781420088892-14 Pages: 161-180 Publication Title: Human-Computer Interaction.

[97] Antoinette Rouvroy. 2013. The end(s) of critique: Data behaviourism versus due process. In *Privacy Due Process and the Computational Turn: The Philosophy of Law Meets the Philosophy of Technology*. Taylor & Francis, 143–167. https://doi.org/10.4324/9780203427644

[98] Christian Sandvig, Kevin Hamilton, K. Karahalios, and Cédric Langbort. 2014. Auditing Algorithms : Research Methods for Detecting Discrimination on Internet Platforms. In *Preconference at the 64th Annual Meeting of the International Communication Association*. University of Michigan, Seattle, WA, USA, 23.

[99] Donghee Shin. 2021. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *International Journal of Human-Computer Studies* 146 (Feb. 2021), 102551. https://doi.org/10.1016/j.ijhcs.2020.102551

[100] Radish Singh, Miguel Fernandes, Nick Lim, and Eric Ang. 2018. *The case for artificial intelligence in combating money laundering and terrorist financing*. Technical Report. Deloitte. https://www2.deloitte.com/mm/en/pages/financial-advisory/articles/the-case-for-artificial-intelligence-in-combating-money-laundering-and-terrorist-financing.html

[101] Dominic S.B. Soh and Nonna Martinov-Bennie. 2011. The internal audit function: Perceptions of internal audit roles, effectiveness and evaluation. *Managerial Auditing Journal* 26, 7 (Jan. 2011), 605–622. https://doi.org/10.1108/02686901111151332 Publisher: Emerald Group Publishing Limited.

[102] Jiao Sun, Q. Vera Liao, Michael Muller, Mayank Agarwal, Stephanie Houde, Kartik Talamadupula, and Justin D. Weisz. 2022. Investigating Explainability of Generative AI for Code through Scenario-based Design. In *27th International Conference on Intelligent User Interfaces (IUI '22)*. Association for Computing Machinery, New York, NY, USA, 212–228. https://doi.org/10.1145/3490099.3511119

[103] Harini Suresh, Steven R. Gomez, Kevin K. Nam, and Arvind Satyanarayan. 2021. Beyond Expertise and Roles: A Framework to Characterize the Stakeholders of Interpretable Machine Learning and their Needs. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–16. https://doi.org/10.1145/3411764.3445088

[104] The Federal Reserve Board of Governors in Washington DC. 2011. The Fed - Supervisory Letter SR 11-7 on guidance on Model Risk Management. https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm

[105] Adeline Toader. 2019. Auditability of AI Systems – Brake or Acceleration to Innovation? https://doi.org/10.2139/ssrn.3526222

[106] Dylan Tokar. 2023. Google Cloud Launches Anti-Money-Laundering Tool for Banks, Betting on the Power of AI. *Wall Street Journal* (June 2023). https://www.wsj.com/articles/google-cloud-launches-anti-money-laundering-tool-for-banks-betting-on-the-power-of-ai-2512ccce

[107] Richard Tomsett, Dave Braines, Dan Harborne, Alun Preece, and Supriyo Chakraborty. 2018. Interpretable to Whom? A Role-based Model for Analyzing Interpretable Machine Learning Systems. In *2018 ICML Workshop on Human Interpretability in Machine Learning*. arXiv, Stockholm, Sweden, 7. http://arxiv.org/abs/1806.07552 arXiv: 1806.07552.

[108] Trade and Industry Appeals Tribunal. 2022. Bunq vs. DNB, ECLI:NL:CBB:2022:707, 21/323 and 21/1108. https://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:CBB:2022:707 Soort: Uitspraak.

[109] Jon Truby, Rafael Brown, and Andrew Dahdal. 2020. Banking on AI: mandating a proactive approach to AI regulation in the financial sector. *Law and Financial Markets Review* 14, 2 (April 2020), 110–120. https://doi.org/10.1080/17521440.2020.1760454 Publisher: Routledge _eprint: https://doi.org/10.1080/17521440.2020.1760454.

[110] Chun-Hua Tsai, Yue You, Xinning Gui, Yubo Kou, and John M. Carroll. 2021. Exploring and Promoting Diagnostic Transparency and Explainability in Online Symptom Checkers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–17. https://doi.org/10.1145/3411764.3445101

[111] UNODC. 2011. *Estimating illicit financial flows resulting from drug trafficking and other transnational organized crimes*. Discussion paper. United Nations. https://www.unodc.org/documents/data-and-analysis/Studies/Illicit_financial_flows_2011_web.pdf

[112] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300831

[113] Mark Weber, Jie Chen, Toyotaro Suzumura, Aldo Pareja, Tengfei Ma, Hiroki Kanezashi, Tim Kaler, Charles E. Leiserson, and Tao B. Schardl. 2018. Scalable Graph Learning for Anti-Money Laundering: A First Look. http://arxiv.org/abs/1812.00076 arXiv:1812.00076 [cs].

[114] Patrick Weber, K. Valerie Carl, and Oliver Hinz. 2023. Applications of Explainable Artificial Intelligence in Finance—a systematic review of Finance, Information Systems, and Computer Science literature. *Management Review Quarterly* 73, 1 (Feb. 2023), 41. https://doi.org/10.1007/s11301-023-00320-0

[115] Lisa Webley. 2010. Qualitative Approaches to Empirical Legal Research. In *The Oxford Handbook of Empirical Legal Research*, Peter Cane and Herbert M. Kritzer (Eds.). Oxford University Press, Oxford, 0. https://doi.org/10.1093/oxfordhb/9780199542475.013.0039

[116] Christine T. Wolf. 2019. Explainability scenarios: towards scenario-based XAI design. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, Marina del Ray California, 252–257. https://doi.org/10.1145/3301275.3302317

[117] Yvette D. Clarke. 2023. Algorithmic Accountability Act of 2023. , 553 pages. https://www.govinfo.gov/app/details/BILLS-118hr5628ih Call Number: Y 1.6:, Y 1.4/6: Committee: Committee on Energy and Commerce Publisher: U.S. Government Publishing Office Source: DGPO.

[118] Joyce Zhou and Thorsten Joachims. 2023. How to Explain and Justify Almost Any Decision: Potential Pitfalls for Accountability in AI Decision-Making. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23)*. Association for Computing Machinery, New York, NY, USA, 12–21. https://doi.org/10.1145/3593013.3593972

APPENDIX

AML-CFT Theme	Legal reference	Is AI opacity a problem? For which model?	Why?
Customer knowledge and constant vigilance over business relationships	French Monetary Code (CMF) Articles L.561-4-1 to L. 561-14-2	No	The update of customer and beneficial owner databases is not made with AI in the use cases we are considering.
Risk classification	CMF Article L. 561-4-1	Yes for NT	Banks need to understand the new typologies of risk detected by the AI to update their risk classification.
Calibration / allocation of material and human resources	CMF Article R. 561-38	Yes for RS	Assessing the suitability of AI for prioritizing alerts
Constant vigilance	CMF Article L. 561-6	Yes for NT	Justifications might be needed on the training frequency.
Careful examination: Ability to detect inconsistencies/anomalies	CMF Article L. 561-6	Yes for NT	The relevance of a model can be justified with performance statistics, but understanding why an anomaly was not detected is important for both supervisors and banks.
Processing alerts in a timely manner	Sanction Decision BMW Finance 16/06/23	Yes for NT and SR	AI opacity can make reviews longer
Adaptation / completeness of the system in relation to the risk classification	CMF Article R. 561-12-1, Sanction Decision Axa Banque 15/02/23	Yes for NT	The alignment between human and machine on important parameters should be demonstrated
Enhanced vigilance: ability to analyze risky alerts	CMF Article L. 561-10-2	Yes for SR	We need to be able to understand the criteria that generate a risky alert.
SAR obligation: ability to produce high-quality SAR when relevant	CMF Article L. 561-15	Yes for SR and NT	We need to be able to understand the criteria that generate a risky alert.
Internal control: incident detection; stability over time; mastering of the system (from external service provider); safety net in case of failure	CMF Article R561-38-4, Order of November 3, 2014	Yes for SR and NT	Have to be able to anticipate the model's behavior to anticipate plausible incidents; Have to demonstrate AI behavior does not drift; Have to be able to demonstrate the control of your system.

Table 3: Summary of the compliance assessment to determine the points in the AML-CFT legislation with which AI opacity interferes. The assessment was made for the two AI use cases presented in Figure 3.1.3. They are denoted here with the acronyms "SR" for "Risk Scoring" (Scenario 1), and "NT" for "New typologies (scenario 2).

Participant ID	Role	Years in profession	Familiarity with AI (on a 7 points Likert scale)	Workshop and Interview ID	Recorded
P1	Supervisor, document-based control	>10	2	W1	Yes
P2	Supervisor, on-site control	>10	3	W1	Yes
P3	Supervisor, document-based control	Between 1 and 3	3	W2	Yes
P4	Supervisor, document-based control	Between 4 and 10	3	W2	Yes
P5	Supervisor, document-based control	Between 1 and 3	3	W2	Yes
P6	Supervisor, document-based control	Less than a year	3	W3	Yes
P7	Supervisor, document-based control	Between 4 and 10	5	W3	Yes
P8	Supervisor, document-based control	Between 4 and 10	3	W3	Yes
P9	Supervisor, on-site control	Between 1 and 3	7	W4	No
P10	Supervisor, on-site control	Between 4 and 10	7	W4, I1	No, Yes
P11	Supervisor, on-site control	Between 4 and 10	1	W5	Yes
P12	Supervisor, on-site control	Between 4 and 10	3	W5	Yes
P13	Supervisor, on-site control	Between 4 and 10	3	W5	Yes
P14	Supervisor, AML-CFT policy	>10	6	I2	Yes
P15	Bank, Head of AML-CFT compliance	>10	3	W6	No
P16	Bank, Head of data science	Between 4 and 10	7	W6	No
P17	Bank, AML-CFT Compliance Officer	Between 4 and 10	1	W6	No
P18	Bank, AML-CFT Compliance Officer	Between 4 and 10	3	W6	No
P19	Bank, Data scientist	Between 1 and 3	7	W6	No
P20	Bank, Data scientist	Between 1 and 3	7	W6	No

Table 4: Description of role, experience, familiarity with AI of participants in the study.

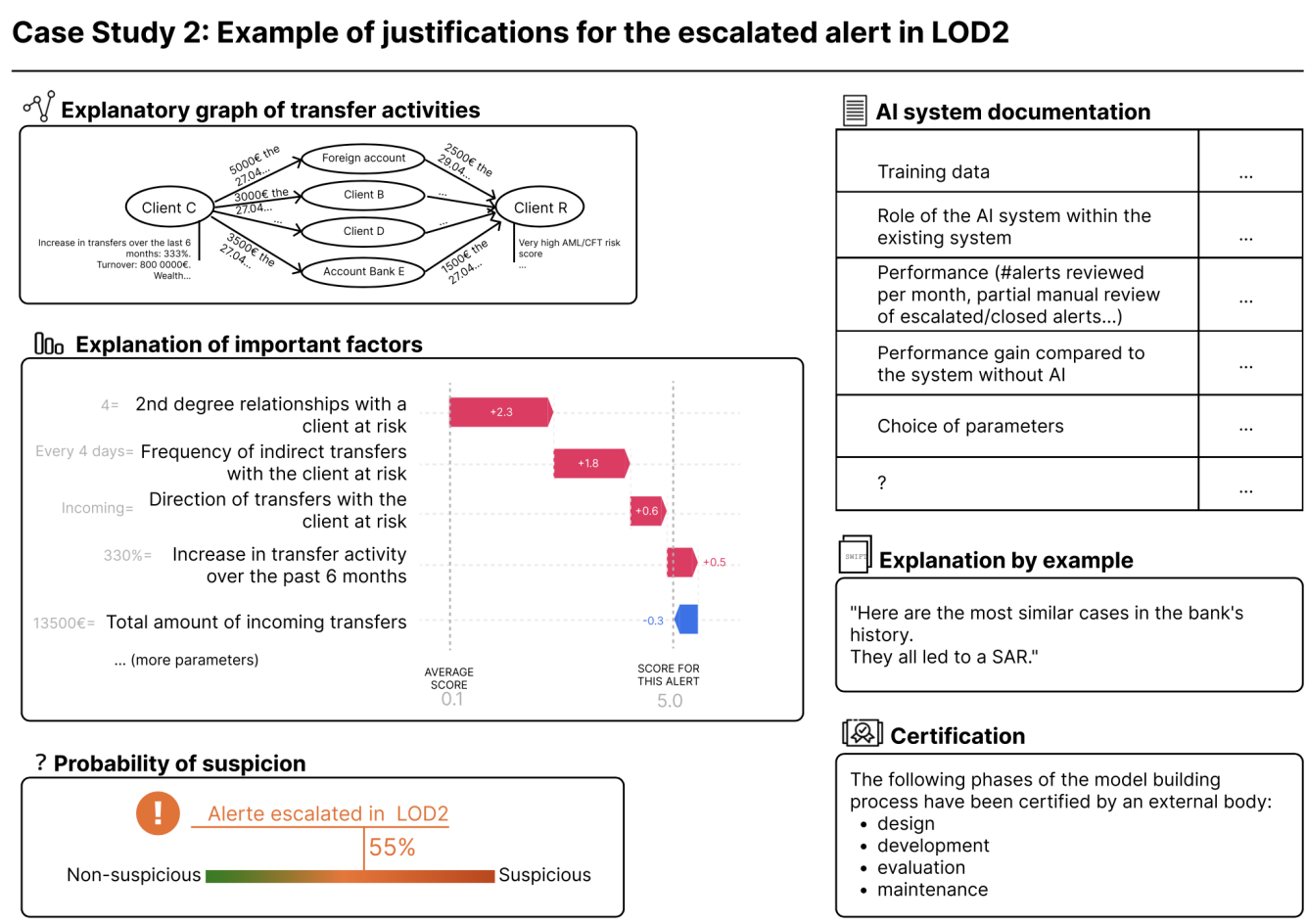

Figure 4: Conceptual justifications shown for the scenario 2 and its example alert. Conceptual justifications for the scenario 1 followed the same format.