The AI Paradox in Financial Crime Prevention: A Data-Rich, Intelligence-Poor Ecosystem

Corresponding author: Aleksandra Kostanecka
Email: aleksandra.kostanecka@studbocconi.it

Abstract

Financial crime networks are becoming increasingly more resilient with the development of new technologies and integration of artificial intelligence (AI) and machine learning (ML). Despite previously existing and newly introduced AML-centered directives, regulations, and agencies working towards jurisdictional and legal coverage regarding the use of AI in financial crime prevention, significant contradictions, gaps, and inconsistencies persist. Through an interdisciplinary qualitative analysis, this paper integrates key insights from financial forensics, behavioural psychology, law, and cybersecurity, to explore how AI works in favor of regulators as well as against them, lowering the barriers to entry for fraud and deceit. The use of academic research and investigative journalism provides bases for comparison across domains in the identification of evident patterns and phenomena. The paper proposes three distinct reforms that reflect the analysis and arguments discussed throughout the research. By synthesising cross-domain insights, this literature provides a holistic view of the dynamics between artificial intelligence, regulators, and perpetrators.

Keywords: financial crime, artificial intelligence, machine learning, money laundering, financial intelligence units, transaction monitoring

Authors and Affiliations: Aleksandra Kostanecka with contributions from Siddhanth Achanta, Stefania Andrei, Chiara Banfi, Sofia Bulycheva, Diana Cernat, Giorgia Dai, Maria Feliziani, Miriam Ferrara, Lilla Felvinczi, Andrada-Ioana Ilie, Zofia Jakub, Alessandra Kozlova, Mariam Labadze, Giovanni Mangiatordi, Hugo Saint Martin, Giacomo Mensi, Federico Moseley, Nidal Oğur, Ayse Ceyda Oral, Nikte Juarez Ozdemir, Evelina Pîntea, Alice Ratti, Ahmet Ozan Savur, Emma Tardivo

June 2026 | Institutional Affiliation: Bocconi University, Milan, Italy

1 Introduction

1.1 Context and Significance

Financial crime networks operate transnationally, crossing the boundaries of jurisdictions, institutions, and agents. These networks have developed into a systemic threat to global economic and political stability, rendering its study and mitigation increasingly indispensable. Between 2% and 5% of global GDP is consumed yearly by money laundering alone, emphasising the sheer scale, severity, and complexity of this criminal phenomenon (UNODC, n.d.).

With the rapid transformation of global processes and operations in response to continuously developing artificial intelligence, a paradox arises. On the one hand, AI has become a support for regulators in pattern identification and process automation, and on the other, it has revolutionised previously low-scale and low-complexity crimes, allowing perpetrators to become continuously more creative and evolved in their deception and fraudulent schemes.

1.2 Research Problem

Despite decades of AML frameworks and the recent rise in AML-specific institutions and agencies, financial crime remains a persistent and evolving challenge, artificial intelligence amplifying this even further. Criminal networks adapt rapidly, exploiting legal loopholes, weak enforcement mechanisms, and hesitation surrounding increasingly sophisticated digital technologies. Not only has AI been used as a predominantly reactive force by regulators but has also been masterfully manipulated by perpetrators in their favor, making identity concealment, online fraud, and financial obscurity a much easier task, requiring increasingly less expertise and technological sophistication. Regulators and institutions continue to struggle with creating appropriate and sufficient legal coverage surrounding artificial intelligence and its ambiguity.

1.3 Aim and Research Question

This paper examines how artificial intelligence supports regulators in anomaly detection, pattern recognition, and transaction analysis to more efficiently and accurately identify financial crime networks. On the other hand, it also explores how artificial intelligence

is used also to help and develop the very criminals regulators are tasked to uncover. Using deepfake technology, automated phishing, and transaction manipulation, AI can be used as a tool to make fraud a more advanced, attainable, creative, and wide-spread phenomenon, lowering the barrier to entry by reducing complexity and required technological expertise. These two contradicting sides are explored through an interdisciplinary lense, focusing on four key perspectives: financial, behavioural, legal, and cyber. The central question, therefore, examines how AI is used both to support and break down criminal enterprises, and tackles the specific 'data-rich, intelligence-poor' paradox that arises from continuous development of technological use.

1.4 Methodology

This paper is based on qualitative and interdisciplinary research, combining a financial, behavioural, legal, and cyber perspective. Since financial crime is not limited to a single domain, an interdisciplinary approach allows the analysis of the effects of artificial intelligence to be holistic and applicable to the real-world environment.

The study relies exclusively on credible, verifiable, and publicly accessible academic and institutional sources. Among these are extracts from the Financial Action Task Force (FATF), Europol, and the United Nations Office on Drugs and Crime (UNODC).

1.5 Contribution to Literature and Paper Structure

The paper's primary contributions are the three reforms proposed, namely the introduction of federated data-sharing models, tiered Suspicious Activity Reporting (SAR) thresholds, and the improvement of coordination and trust in AI-based detection. They take into consideration all arguments presented throughout the paper, aiming to relieve some of the pressure placed on institutions, in particular Financial Intelligence Units (FIUs), and create a more level playing field between regulators and criminals. Each reform is presented with contextual background and application, as well as potential downsides that could arise during implementation, to ensure a balanced and unbiased perspective.

The paper is divided into five main parts. The first discusses how artificial intelligence is used to empower criminals and lower the barriers to entry of fraud and deceptive schemes. It is analysed through the four perspectives mentioned previously, namely financial, behavioural, legal, and cyber. The second part continues this structure, however in the

perspective of regulators, and how artificial intelligence supports pattern recognition, information analysis, and data organisation. The third part aggregates the previous concepts into two phenomena, namely AI acting as a force multiplier and AI acting as a friction amplifier. More precisely, it explores the argument that AI can act as an accelerant for regulators in terms of automation and organisation, however it can also become a significant burden, distracting regulators from actionable intelligence, creating enough noise to mask real threats and anomalies. The fourth part hones into a particular aspect of AI acting as a friction amplifier - the 'data-rich, intelligence-poor' constraint - a phenomenon arising from the mass aggregation and influx of data resulting from the automation of suspicious activity reporting and data collection. The final part of this paper introduces the three key reforms mentioned prior, proposed as a result of the analysis and argumentation throughout this paper, followed by a summarising conclusion.

2 The AI Transformation of Financial Crime

2.1 AI-Enabled Fraud

The introduction of artificial intelligence has significantly transformed the landscape of financial crimes, shifting from human-driven activities to industrialized, systematized operations, thereby removing human constraint on time and expertise. AI usage enables criminals and cyber-criminals to dramatically increase efficiency by operating on a larger scale, at an increased speed, and developing far more personalized frauds than ever before. In this sense, artificial intelligence functions as a force multiplier for illicit activities, despite the absence of a proportional increase in skills and expertise. This transformation is evident in the emergence of recent sophisticated models, such as fraud-as-a-service (FaaS) economics, automated phishing ecosystems, synthetic identities, and deepfake-enabled social engineering. These developments not only reshape the operational structure of financial crimes but also result in increased difficulty in detecting fraudulent activities.

To deeply understand the phenomenon of FaaS, it is important to consider the perspective of professionals specialized in this area. Kennedy Meda, Fraud Prevention Manager & SME for Deseret First Credit Union, has shared the findings of her work on Thomson Reuters, describing Fraud-as-a-Service as a "secretive industry in which cybercriminals offer tools, services, and support to fraudsters in exchange for payment" (Meda, 2025).

In the same article, Meda explains how the support to commit fraud has significantly increased the frequency of cybercrimes. Moreover, despite the efforts by authorities to enhance

prevention and fraud detection through AI and Machine Learning methods, cyber criminals are becoming increasingly more responsive and adaptive.

To deeply understand this reality and its mechanisms, researchers world-wide began investigations. Parida et al. (2025) discussed researchers studying dark web marketplaces through digital ethnography, record analysis, and forensic tools, and identified five main cyber fraud schemes: credential stuffing, Ransomware-as-a-Service, phishing kits, carding, and money mulling services. In addition to this, they detected low entrance barriers in the organizational structure.

Automated phishing ecosystems refer to large-scale fraud systems in which phishing attacks are no longer created and sent manually, but produced, adapted, and distributed through automated tools (Heiding et al., 2024). Traditional phishing previously relied on generic emails and simple fake websites, its success reliant on the sending of large numbers of low-quality messages with the hope that at least a small number of victims would respond (Schmitt & Flechais, 2023). However, with the development of artificial intelligence, phishing has become a lot more sophisticated, targeted, and efficient, generating highly believable messages, which are much harder to distinguish from authentic ones (Schmitt & Flechais, 2023; Czybik et al., 2026).

AI-driven phishing allows criminals to generate very realistic messages, imitating writing styles and personalizing content to adapt scams to different victims and situations at very low cost (Czybik et al., 2026). This means that phishing is now based on very convincing communications that appear legitimate and urgent, making them much harder for victims to identify as fraudulent (Schmitt & Flechais, 2023; Heiding et al., 2024). As this system is also highly automated, it is possible to carry out these attacks on a much larger scale, reaching a wider audience and increasing the speed of phishing campaigns (Heiding et al., 2024; Czybik et al., 2026).

As a result, the introduction of AI has transformed phishing from a mostly manual and low-quality method into a highly efficient, automated, and believable system of deception, leading to more sophisticated attacks, lower detection rates and a growing number of victims (Schmitt & Flechais, 2023; Gallo et al., 2023). This has made automated phishing one of the clearest examples of how AI can influence financial crime, making it less detectable and more widespread (Heiding et al., 2024; Gallo et al., 2023).

Deepfake-enabled social engineering refers to the use of AI-generated audio, video, or images to deceive people into trusting a false identity or situation for fraudulent purposes (Pedersen et al., 2025; Dsouza et al., 2024). Social engineering, more generally, involves

manipulating individuals by exploiting trust, urgency, fear, or authority to influence their behavior. (Naz et al., 2024). In the context of financial crime, this often involves impersonating bankers, company executives, regulators, or other authoritative figures to pressure victims into transferring money, revealing sensitive information or authorizing fraudulent transactions.

Deepfake-enabled social engineering takes this further by making impersonation much more realistic and adaptable. While traditional fraud mostly relied on emails, basic phishing messages, and mock phone calls, AI took scams to another level by allowing criminals to generate realistic videos, clone voices, manipulate existing photos and videos, and produce highly persuasive real-time communications that are difficult to verify (Pedersen et al., 2025). Deepfake technology has clear criminal potential in areas such as CEO fraud and other forms of impersonation-based deception, becoming an increasingly widespread and damaging tool in financial crime. This way, AI does not just elevate older fraud methods, it increases their reach, lowers the cost of producing convincing fake content, and strengthens the criminals' ability to exploit trust, focusing on speed and scale (Dsouza et al., 2024).

AI can also be identified here as a friction amplifier, since rather than reducing uncertainty and making communication clearer, it makes it much harder for both victims and institutions to differentiate between real and fabricated situations, identities, and evidence. This causes verification processes to take longer, makes trust easier to exploit, and complicates fraud detection. The recent scan on AI and deepfakes, conducted by the Financial Action Task Force (FATF), warned that deepfakes can amplify the complexity, scale, and reach of criminal operations by manipulating identities and impersonating authorities, while making the detection and prevention of such content increasingly more difficult (FATF, 2025). As a result, deepfake-enabled social engineering plays a particularly important role in financial crime because it combines the psychological logic of traditional scams with the automation and scalability of AI-generated deception (Dsouza et al., 2024).

2.2 Behavioral Vulnerabilities in AI-Driven Financial Crime

Although not without its advantages, AI can also be used to assist criminals, using tools such as the malware and deepfakes previously discussed in an efficient way to commit fraud. This breeds an abuse of trust, where people tend to be deceived by synthetic content produced by criminal organizations. Moreover, under uncertainty, humans tend to follow advice from sources they perceive as credible, a concept known as authority bias, which criminals exploit

to manipulate decision making. It is further linked to victim cognitive overload, in which AI overwhelms the victim's ability to assess the situation and make rational decisions.

AI is significantly increasing the ability of criminals to exploit trust in digital environments. The United Nations Office on Drugs and Crime (UNODC) highlights how technological developments, along with the creation of a new transnational criminal ecosystem, have created a technology-driven and dynamic underworld that is undergoing a transforming process enhanced by artificial intelligence (UNODC, 2025).

In this context, trust becomes a vulnerability. Individuals are more likely to rely on digital interactions that are a synthetic product of a criminal organization, AI not only making online crimes easier but also more scalable and pervasive.

People's reaction and reliance on expert opinions in undefined situations is greatly influenced by a cognitive tendency called authority bias, which leads humans to defer to advice based on how credible they consider the source to be, with some following it unquestioningly while others remaining more skeptical (Gültekin & Siniksaran, 2025). A study by Klingbeil et al. (2024) revealed that humans uncritically follow AI-generated recommendations.

One example provided by Gültekin and Siniksaran to showcase the reliability of AI generated recommendations was sports betting, since it is a system where statistical inputs and past data carry crucial importance which could be further elevated by the help of AI. However, apart from the seemingly “harmless” usage and assistance of AI, there are matters in which it is imperative to understand whether the reliance on AI advice would cause bias in human decision making.

In their research, Williams-Ceci et al. (2026) show how AI writing assistants could manipulate human cognition by using autocomplete suggestions, with users often being unaware of this fact. This was revealed in their experiment where participants consistently accepted AI-generated advice as sound and well-grounded, which the authors attributed to the fact that AI tends to produce responses that appear accurate and well-suited to the user's needs.

According to the observatory report from the Europol Innovation Lab, the widespread usage of deepfakes poses threats to how people perceive information, since it harms public trust and verified facts. This could ultimately push society towards a state of collective confusion over what is true and trusted, a phenomenon the report refers to as 'reality apathy' (Europol, 2022).

Another important effect of AI-enabled financial crime is victim cognitive overload, which occurs when a person is exposed to too much information, urgency, or emotional pressure at once, making it harder to think carefully and make well-informed, rational decisions (Nasser

et al., 2020). This is especially relevant in financial crime, where scammers exploit how people react in stressful environments, urging them to act fast without taking the time to question whether the situation is real or not (Montañez et al., 2020). AI plays an important role in creating these conditions, making it easier for criminals to imitate voices, fake identities, and produce very convincing messages and deepfakes on a much larger scale (Europol, 2022).

The most dangerous part of this pattern is that the victim is not being deceived during a single isolated time, but is led on for weeks or months, making the fake situation they unknowingly find themselves in more believable and carefully fabricated. By using multiple signals and channels to affect the victim, such as urgent emails, phone calls from so-called authorities, and even fake threats involving third parties, the victim finds themselves under constant pressure and stress, which causes confusion and impedes sensible judgment (Montañez et al., 2020; Nasser et al., 2020). Instead of taking a moment to carefully assess the situation and detect anomalies, the victim is constantly being contacted and pressured, beginning to rely solely on instinct, fear, or trust in what they believe is a legitimate authority figure. Thus, AI helps reinforce the psychological side of fraud, by making scams feel more urgent, believable, and difficult to question (Europol, 2022).

It is therefore possible to conclude that AI not only helps improve the speed and scale of financial crime but also strengthens criminals' ability to manipulate human decision-making. This is a very significant aspect as it shows that AI-enabled fraud is linked to broader behavioral issues such as trust exploitation and authority bias. The technological support itself is only part of the problem, with the wider often depending on how effective AI is in overwhelming a victim's ability to pause, assess the situation, and respond rationally.

2.3 Legal Blind Spots in AI-Enabled Financial Crime

As concerning as social scoring, algorithmic discrimination, or identity theft may appear to a reasonable person, without a comprehensive legal framework to criminalize these practices, they become unprosecutable. Prior to the adoption of the European Union Artificial Intelligence Act (EU AI Act) of June 2024, the regulation of AI practices threatening fundamental human rights remained fragmented, relying on emerging national legal frameworks (Madiega, 2023).

The EU AI Act establishes integrated rules for launching on the market, putting into operation, and the use of artificial intelligence within the boundaries of the Union, while also prohibiting, under Article 5, eight practices posing "unacceptable risks" to human rights and Union values, such as respect for human dignity, freedom, and equality (European Union,

2021). The prohibitions established under the aforementioned article include AI systems that deploy harmful manipulation and deception, exploitation of vulnerabilities, social scoring, individual criminal offense risk assessment and prediction, untargeted scraping to develop facial recognition databases, emotion recognition, biometric categorization, and real-time remote biometric identification (European Union, 2024, Article 5). Additionally, under Article 6 in conjunction with Annexes I and III of the Act, AI systems posing “high risks” to health, safety, and fundamental rights are subject to a set of requirements and obligations. Lastly, AI systems that pose transparency risks are subject to transparency obligations under Article 50. Any system that poses “minimal to no risk” remains unregulated however could be subject to voluntary codes of conduct established by providers and deployers (European Union, 2024, Article 95).

Under the current Regulation, non-compliance with the provisions shall be subject to administrative fines, in accordance with Article 99. Still, attaching criminal liability to AI systems has not yet been part of any major legislative initiatives at the EU level (Sachoulidou, 2024). Therefore, considering the current legal frameworks in place across the Union, AI-related crimes concern the prohibitions, mandatory provisions, and obligations outlined in the aforementioned articles of the EU AI Act.

As AI's role in the financial market is still a rather new development, legislation around criminal acts committed by it is still mostly in the drafting stage. As we have seen in the previous section, the EU was one of the first legal entities that promulgated regulations dealing specifically with AI systems, implementing not only the already cited AI Act, but also other regulations and directives. The Market Abuse Regulation and Market Abuse Directive, for example, create a system of allocation of liability for both administrative and criminal offenses, although they still leave a margin of maneuver for Member States on the measures they apply, as long as those countries consider market manipulation conducted by AI a felony, especially when committed intentionally (Ahmed Abdelaziz, 2025).

Furthermore, the European Commission has issued the "Ethical Guidelines for a Trustworthy AI", which, alongside the Council of Europe draft framework convention on AI and human rights, aims at creating a strategy and a set of principles by which there can be a more ethical and sustainable development of artificial intelligence in conformity with human rights standards (European Commission, 2019).

Nonetheless, the use of AI for criminal purposes still creates doubts in the allocation of liability. As this technology develops and evolves, it has grown more autonomous in its actions, thus challenging the old conception of accountability, which required a certain degree of

intention and consciousness from a human actor that AI systems can stray away from or mask, making it difficult to find an imputable individual (Vuletic & Duic, 2025).

The emergence of AI as a financial crime tool poses important challenges for legal scholars, such as liability gaps, jurisdictional lag and evidentiary difficulties related to synthetic media.

Liability gaps are rooted in the disruption of criminal responsibility by AI systems, which often autonomously perform harmful actions without direct human directions. Causation and foreseeability become increasingly impossible to pinpoint, blurring the boundaries between who hypothetically could be an actus reus (guilty act) or mens rea (guilty mind) of the situation. As current fault-based liability regimes struggle to adequately address the actors and consequences of AI damage, the EU is actively introducing reforms such as the AI Liability Directive, promoting efficient legal mechanisms like a "presumption of causality" and increased access to legal evidence to help victims properly communicate and prove in court the harm caused (Marchant & Allenby, 2017). Criminal liability becomes a difficult assignment as responsibility rests on the shoulders of multiple parties, including both developers and users of complex AI software chains.

Jurisdictional lag corresponds to the mismatch between legal systems of different countries and the internationally transactional nature of AI-enabled crime. Legal enforcement is further complicated by the fact that AI systems may be created in one jurisdiction, hosted in another, and put to fruition globally. Legal fragmentation regarding this issue is caused by different legal frameworks such as the EU, United States, and China addressing the problem of AI-enabled crime in drastically different ways. Particularly, the usage of synthetic media in court proceedings remains a core issue. While the United States criminalizes the perpetration of non-consensual deepfakes, other countries are not as progressive in promoting adequate law packages, which becomes a stimulus for malicious subjects to exploit legal loopholes regarding this subject matter (Roberts et al., 2023).

For the first time in history, the reliability of digital evidence is being challenged in court proceedings as the presence of synthetic media allows for investigative scenarios to be easily falsified, forcing to revolutionize forensic methods and complicating evidence authentication. While important reforms such as the Digital Services Act and the AI Act oblige to comply with obligations of transparency and proper labelling, the "liar's dividend" factor cannot be completely eliminated, forcing courts to deal with defendants that claim the inauthenticity of genuine evidence and vice versa (Artificial Intelligence Act, 2026).

The challenges of assigning legal responsibility, the jurisdictional lags regarding the global nature of AI-enabled crime, and evidentiary challenges come in handy to those committing crimes of large-scale fraud, identity theft, money laundering and other automated financial manipulation. As AI tools produce increasingly convincing phishing schemes, synthetic identities, and deepfakes for impersonation purposes, targeted and efficient attacks are conducted towards both financial institutions and individuals. Autonomously executed fraudulent behaviour further dissipates liability, enhancing jurisdictional problems already present due to the cross-border nature of financial networks and the assumption of criminal activity being human-driven (Chesney & Citron, 2019).

The EU is just recently starting to address these key issues with reforms such as the Anti-Money Laundering directives (EU Directive 2018/843) and the Anti-Money Laundering Authority (AMLA, EU Regulation 2024/1620). However, there is still an important regulatory gap between these newly introduced AML frameworks created for monitoring and addressing suspicious financial transactions, and traditional AI regulation such as the EU AI Act, which serves the purpose of ensuring system safety and transparency. Financial misconduct committed through AI remains not directly criminalized, highlighting contradictions between horizontal AI governance and sector-specific regulations (FATF, 2021).

An exemplifying case is the rise of “CEO fraud” committed through AI and other deepfake scams in the financial sector. AI-generated voice cloning, impersonating senior executives and imitating their voices to a near perfect degree, was heavily used to illegally authorize financial transfers in the past years. These deception schemes manage to reduce the cost of fraud and boost their success and credibility, all while circumnavigating AML directives on suspicious transactions and transparency obligations of the AI Act (Directive (EU) 2018/843). Specialized technical expertise is required to prove a communication was AI-generated in court, further solidifying the role of the “liar’s dividend”. Unfortunately, both financial regulations and AI-specific legislation are not sufficient to properly address this legal enforcement gap, enhancing how current EU approaches are reactive to these problems rather than preventive. Post-factum detection should be consequential to accountability-detecting mechanisms already embedded in AI software to keep on track with increasingly complex and sophisticated AI-enabled financial crime, while financial regulations, criminal law, and AI acts should be complementary (Wall Street Journal, 2019).

2.4 The Cyber Architecture of AI-Enabled Criminal Innovation

The increasing digitalization of financial systems has significantly expanded the scope and complexity of cyber-enabled financial crime. Artificial intelligence plays a central role in this transformation by enhancing the technical capabilities of malicious actors and enabling more efficient and scalable attack strategies. As highlighted by the Financial Action Task Force (FATF), AI represents both a tool for innovation and a powerful enabler of illicit activities, particularly in the context of fraud and money laundering (FATF, 2025). This dual-use nature of AI has contributed to the emergence of increasingly sophisticated cybercrime ecosystems.

One of the most significant impacts of AI on cyber-enabled financial crime is its ability to automate and scale attacks. AI-driven tools allow both low-skilled and highly sophisticated actors to conduct large-scale operations with minimal effort, significantly lowering the barrier to entry for cybercrime (FATF, 2025). Machine learning algorithms can automate phishing campaigns, generate malicious content, and identify vulnerabilities in target systems, enabling attackers to operate at a scale that would otherwise be impossible through manual methods. This automation increases both the volume and efficiency of attacks, allowing criminals to target a larger number of victims simultaneously (Yu et al., 2024). Furthermore, generative AI models enable the automated creation of highly personalized phishing content, significantly increasing the reach and effectiveness of large-scale cyber-attacks, as discussed previously (Falade, 2023).

Beyond enhancing deception, AI enables criminals to directly target the detection systems used by financial institutions. Through adversarial techniques and iterative testing, attackers can identify patterns in transaction monitoring systems and adjust their behavior to remain below risk thresholds. This process, often described as model evasion, transforms financial crime into a feedback-driven optimization problem, where illicit actors continuously refine their strategies based on system responses. As a result, AI-driven detection systems risk becoming reactive rather than preventive, as criminals adapt faster than the updating of regulatory models. (Carletta et al., 2021).

AI is not only improving existing attack methods but also enabling entirely new forms of cyber-enabled financial crime. The development of generative AI and autonomous agents has introduced the possibility of fully automated fraud and money laundering operations. AI systems can generate fake documentation, simulate business transactions, and create complex financial “paper trails” to support fraudulent schemes (FATF, 2025). Additionally, AI-driven agents can execute real-time financial operations, such as routing transactions through multiple

accounts to avoid detection, effectively creating automated laundering pipelines. These developments mark a shift towards more adaptive and self-sustaining cybercrime systems.

The growing role of AI in financial cybercrime is supported by real-world cases that illustrate its impact. For instance, a reported case involved the use of AI-generated deepfake technology to simulate a video conference with senior executives, leading to the fraudulent transfer of approximately $25 million (FATF, 2025). Such cases demonstrate how AI enhances both the credibility and execution of cyber-attacks, making them more difficult to detect and prevent. These examples confirm that AI-driven cybercrime is not a theoretical risk but an already operational and rapidly evolving threat. Recent industry analyses further confirm that AI-enabled fraud, including voice cloning and executive impersonation schemes, is becoming increasingly prevalent and financially damaging for organizations (Deloitte, 2023).

3 AI as a Regulatory and Investigative Tool

3.1 AI in Financial Crime Detection

Financial institutions are spending vast amounts on Know Your Customer (KYC) controls and transaction monitoring mechanisms. However, according to Interpol, the financial industry detects only about 2% of global financial crime flows (Begolli, 2019). This is where AI comes into play, posing as a potential solution that can enhance supervision and prevent significant losses. Artificial intelligence (analytical, generative, and agentic) is an umbrella term for a range of technologies that can comprehend and generate text, interpret images and audios, and learn from data over time (Verhagen et al., 2025). Analytical AI completes analytical tasks quickly and sharpens accuracy and anomaly detection by monitoring large amounts of data. Generative AI learns from patterns, collects and extracts data from documents, and summarizes information about individuals and entities. Lastly, agentic AI (a class of systems designed to act autonomously) can set goals, plan multi-step actions, execute tasks, and adapt based on outcomes (Sloan, 2026).

To explain what the Anti-Money Laundering (AML) Transaction Monitoring Model is and how it works, Schneider and Smalley (2025) furnish an analysis of the instrument and the related phenomenon on the IBM website: “AML (anti-money laundering) transaction monitoring applies AI and machine learning (ML) algorithms, rule-based risk management, suspicious pattern recognition, and advanced analytics to track a customer’s transaction. Furthermore, it also identifies suspicious activity associated with financial crimes (for example, money laundering)”.

In the same article, the authors explain the principal functions of such a model: real time monitoring, scanning, and flagging of suspicious patterns; identification of shady transactions, entities, and activities; and implementation of fintech automation systems, fraud detection tools, and risk assessment.

All these operations are implemented through a set of predetermined rules combined with AI and ML tools to detect new patterns.

From the research by Oztas et al. (2024), it is possible to understand the potential for ML to ease forensic investigation by allowing predictive analysis and improving efficiency through automation, representing a precious resource for regulatory and control authorities.

Specifically, it is fundamental in reducing the number of false positive suspicious activity reports (SARs) filed, as well as expenditure in investigative activity, as highlighted by Oyedokun et al. (2024), since such operations need to be sustainable to continue and develop over time.

Another application of AI in financial regulation lies in the development of network analytics, which shifts the focus from isolated transactions to the relationships between entities. Since traditional anti-money laundering systems rely on rule-based approaches that assess transactions individually, they often fail to detect more complex schemes that involve multiple actors.

Network analytics address this limitation by using a graph-based approach that maps connections between accounts, individuals, and organizations. Empirical research demonstrates that incorporating relational features significantly improves detection performance compared to traditional models, as it captures the context in which transactions occur rather than their standalone characteristics (Wall, 2018; Uddin, 2025). It also enhances operational efficiency by improving risk assessment. By evaluating entities based on their position within a network, such as their proximity to suspicious nodes, AI systems allow compliance functions to prioritize high-risk cases more effectively and reduce false positives (Drake et al., 2025).

However, the effectiveness of this depends heavily on data availability. Fragmented data systems and limited cross-border information sharing restrict the construction of a complete network view, thereby constraining AI systems' overall span of action.

Increased adoption of new AI technology is reshaping financial crime detection. One of the biggest incentives for change is cost reduction, which prior to that increased by 18% in Europe during 2022-2024 (PricewaterhouseCoopers, 2024). AI models that can detect flaws before damages occur offer real time screening and dynamically adjust risk score, directly solving inefficiency problems and report cost savings of 20-30% (Shapiro, 2025). Unlike

traditional monitoring systems, AI platforms develop over time and eliminate manual work, allowing saving on capital expenditure (Ibitola, 2026).

While cutting operating costs, newly implemented technology increases the complexity of regulatory obligations (InnReg, 2025). Alongside this, AI raises the problem of false positives, data privacy, and security issues, and burdens firms with transparency obligations. The Bank of England stated that: “while benefits are expected to grow in all areas, the areas with the largest expected increase over the next three years are operational efficiency, productivity, and cost base.” (Bank of England, 2024, sec. 4.1). Therefore, whether AI will be able to sustain legal burden and optimize revenue generation despite producing false positives and expanding complexity of human workload, remains an open question.

A further advantage of AI in financial regulation is its ability to improve the triage process. This system decides how alerts generated by transaction monitoring systems are prioritized and handled. Traditional triage methods often depend on fixed rules and manual reviews which can lead to inefficiencies and delays, especially with the high number of alerts, many of which are false positives. AI-driven triage systems tackle this issue by ranking alerts based on risk scores from various data points, which include transaction patterns, customer profiles, and contextual information (Drake et al., 2025).

Machine learning models can learn continuously from past investigation results. This helps them better identify high- and low-risk cases over time. Compliance teams can then use their resources more effectively by focusing on the most critical alerts and spending less time on low-risk cases. Additionally, explainable AI techniques improve transparency in decision-making, enabling regulators and compliance officers to understand and justify how prioritization outcomes are reached (Bussmann et al., 2021). Furthermore, advanced AI systems, specifically agentic AI, can automate parts of the triage workflow. For example, they can gather relevant data and suggest next steps, which further boost efficiency. As a result, AI-enabled triage leads to faster decision-making, better detection rates, and a reduced operational burden.

3.2 Human Judgment in Automated Compliance

The uses of artificial intelligence are various and are also being implemented by law enforcement for different purposes. Europol highlights in a report that “AI has the ability to significantly transform policing” (Europol, 2024, p. 12), emphasizing that authorities may be assisted through advanced criminal analytics and tools for the identification of criminals.

However, the Agency underlines how some "ethical considerations of AI use in law enforcement" (Europol, 2024, p. 7) may arise.

In particular, authorities may incur problems of over-reliance. Many policymakers rely on human oversight over AI decision-making processes due to the nature of artificial intelligence; however, scholars have different opinions on how to identify a liable human agent. In particular, the expectation of human oversight may be unrealistic, since humans may not be able to completely understand or control AI-driven processes. (Giannini, 2023).

Relying on AI-driven tools may also lead to investigator deskilling. Europol explains that the implementation of these tools enables "the extraction of actionable insights from vast datasets, improving resource forecasting and operational efficiency" (Europol, 2023, p. 10) and "process unstructured data to provide real-time insights and enhance the ability to tackle urgent situations" (Europol, 2023, p. 10). Law enforcement officials would traditionally conduct these activities, so the risk of deskilling increases. In particular, police officers may become less able to process data independently in complex investigations.

Automation bias is a problem where individuals trust and follow AI more than their own judgement, even when there is a strong presence of other information telling them otherwise. Romeo and Conti (2025) explain that this over reliance creates two types of errors. The first happens when a person follows an incorrect AI recommendation without questioning it. The second happens when a person fails to execute because the AI gave no warning or notification, even though the process was needed. Both errors are made worse by high workloads and task complexity; however, factors such as exposure to past events could help decrease over-trust while not guaranteed.

Alon-Barkat and Busuioc (2023) highlighted that AI systems were originally adopted because they were seen as more neutral and objective than human decision makers, whose judgements are naturally inclined to bias.

3.3 Governing AI in Financial Crime Enforcement

With the progression in the development of AI, it is natural to question whether and how this technology could be used to aid practitioners in incriminating financial crime perpetrators, and for legislators in creating new frameworks to deal with this type of felony. As transactions take on a more digital form, it becomes increasingly harder to follow every new development, as such a powerful instrument like AI could not only take on a surveillance role, being able to monitor different markets and to report violations to authorities, but also help analyse patterns

in criminal behaviours, tracking common elements, and creating strategies to prevent new offences (David & Badi, 2025). Furthermore, using AI guarantees a level of efficiency that other instruments simply cannot match, as it mimics human intelligence and thought processes rather than merely recognising inputs and outputs. By building on its own algorithms to complete tasks such as monitoring transactions or predict new developments, AI systems are also able to operate with a certain grade of autonomy without human intervention, thus reducing the possibility of an oversight or natural human error, even if these machines still have a margin of error based on the prompt given and the accuracy of the data used while training them (Publication, I. A. E. M. E., 2023).

Nonetheless, the use of AI in regulatory procedures still poses some ethical and legal concerns, even if it has already been implemented in various jurisdictions. One example is the allocation of liability for mistakes caused by these systems, since it is a legal "grey area" with no specific allocation of responsibility (neither to the manufacturer or programmer). In addition, the main problem with AI is the risk caused by profiling, as, either by error or by design, it can affect specific groups of individuals and perpetrate harmful biases by over-reporting or pre-emptively blocking their transactions, not only posing a threat to their freedom of economic expression, but also infringing on the fundamental right of non-discrimination (Martin et al., n.d.).

Legal rules play an increasingly important role in shaping how AI is used in financial crime detection. In the European context, AI operates within a combination of legal frameworks such as the GDPR and the AI Act, alongside broader anti-money laundering rules, as discussed previously. These frameworks were not originally designed to work together, which makes their interaction quite difficult in practice. This results in a fragmented regulatory environment, where institutions must deal with multiple layers of compliance at the same time. As recent research shows, this overlap already represents "a huge challenge for the application of international criminal law" (Abel Souto, 2026, p. 21), especially considering how quickly AI technologies are developing.

The GDPR represents a major constraint, as it sets strict limits on how data can be collected and used. Principles such as purpose limitation and data minimisation restrict the ability of institutions to reuse or combine large datasets, even though AI systems depend heavily on large volumes of data to function effectively (European Union, 2016, Article 5). This clearly highlights a conflict between the legal framework's requirement that personal data must be kept to minimal levels and connected to specified purposes, and AI's need for large and flexible data

processing. In practical terms, this could limit the ability of AI models to identify more intricate patterns of financial crime.

The previously mentioned AI Act is another comprehensive framework for AI systems, which imposes stricter requirements on high-risk applications, including those used in financial services (European Union, 2024). The legislation focuses on transparency, risk assessment and human oversight (European Union, 2024, Article 14), ultimately aiming at ensuring that the use of AI technology will be both trustworthy and humancentric. Although such requirements seem entirely reasonable, they do have a negative effect on the adoption of AI technology, as well as on the complexity of the algorithms that are used. Even with the use of AI, the results must often be approved by a human being, thus resulting in less automation. Consequently, although the application of AI could potentially "revolutionize the fight against money laundering", in reality it tends to become more of an obstacle than a boost (Abel Souto, 2026).

Building on the provisions contained in the European Union AI Act (2024), Italy's Law No. 132/2025 marks the first attempt by an EU Member State to design and implement a comprehensive national legal framework regulating the use of AI (Mariuz et al., 2025). Italy's AI Law, which entered into force on 10 October 2025, provides a more detailed, sector-specific approach than the broader EU AI Act. The domestic legislation comprises 28 articles drafted in compliance with the guidelines set forth by the EU Regulation no. 2024/1689, without including any obligations beyond those established by the European framework (Belotti, 2025). Specifically, the law mandates the development of AI systems that are transparent, fair, safe, gender-equality compliant, and sustainable, while emphasizing the need for human oversight to monitor and intervene during AI system operation (Cozzi & Gerbino, 2025).

While this is certainly a positive step towards human-rights-respecting digital tools, Italy's AI Law remains incomplete until more EU member states act to regulate AI domestically. Without this harmonization guided by the EU AI Act, the regulatory landscape remains fragmented and inconsistent.

A very thought-provoking and ongoing topic of legal discussion has been the regulation of AI in war-torn zones, especially as the usage of lethal autonomous weapons has drastically increased, taking the lives of millions of civilians. Scholars around the world have faced increasing difficulty in attributing criminal liability in contexts where chaos and mass destruction are caused by machines that are not humanly regulated.

One of the first attempts to define ethical standards in such a controversial area has been carried out by the United States Department of Defense, which in 2019 published a draft of 5 moral principles that should serve as guidelines in determining legal accountability of

weaponized AI during the kill-chain process (U.S. Department of State, Bureau of Arms Control and Nonproliferation, 2023). The main issues remain the distinction between combatants and non-combatants and the proportionality of civilian damage in relation to military aim, which are fundamental principles of International Humanitarian Law (International Committee of the Red Cross, 2023). Blurring the boundaries of identifying legal subjects responsible for killings has been addressed by philosopher Robert Sparrow of the Australian National University with a paragon to child soldiers, who are deemed causally but not morally responsible for their actions under the UN Geneva Convention (Sparrow, 2016). As weapons become increasingly automated, the moral threshold of entering a war lowers proportionally, since soldiers don't experience battlefield action firsthand and our societies become progressively isolated and feel morally detached from conflict areas. Academics around the world have notably predicted that weaponized AI would dramatically change our political structures, centralizing decision-making power to initiate a war in the hands of very few and jeopardizing democratic control of military interventions (Krishnan, 2016). Legal frameworks around the world have conflicting views on this topic as they reflect the geopolitical interests of multiple parties, with a minority of countries such as Israel, Russia, the United Kingdom, China, and the United States opposing banning regulation of AI in combat zones during the 2023 UN General Assembly summit.

The usage of weaponized AI in conflict zones also intersects with financial crime patterns. The obscurity present in AI systems due to responsibility being divided between state institutions and software developers creates an environment where control of financial flows is weak and corruption in defense procurement persists, with the presence of intermediary shell companies and evasion of export controls being common schemes. A particularly exemplifying case is ISIS, which developed sophisticated drone systems for territorial surveillance and targeted attacks, incorporating AI automation. These were acquired through complex financial networks based on illicit activities such as extortion, illegal trafficking and oil smuggling (Financial Action Task Force, 2015). Despite sanctions, ISIS managed to use front companies and informal value transfer systems to purchase drones and AI software, leveraging financial crime to access military technologies, even while being a non-state entity. The fragmentation of responsibility that complicates legal accountability of AI warfare allows regulatory evasion and illicit financial flows to flourish (SIPRI, 2017).

3.4 Technical Capabilities and Vulnerabilities in AI-Driven Detection Systems

The cybersecurity dimension of regulatory artificial intelligence is increasingly shaped not only by technological capability but also by empirical findings in recent research on fraud detection, adversarial machine learning, and cybersecurity resilience. While AI techniques such as entity resolution, pattern detection, risk scoring, and predictive anomaly detection significantly enhance financial crime detection, academic literature consistently demonstrates that their effectiveness is conditional on data quality, system security, and resistance to adversarial manipulation.

A central finding across the literature is that AI substantially improves the detection of complex and previously unknown fraud patterns. Machine learning and deep learning models outperform traditional rule-based systems by identifying nonlinear relationships and real-time anomalies in large transactional datasets. For instance, recent reviews show that AI-driven systems enhance both speed and accuracy in fraud detection, particularly through predictive analytics and continuous learning from historical data (Adhikari & Hamal, 2024). Similarly, large-scale reviews of fraud detection systems highlight that AI enables scalable risk scoring and behavioral pattern recognition across multiple domains, including banking and insurance, thereby improving early detection capabilities (Zhang et al., 2025).

Entity resolution represents a foundational capability within regulatory AI systems, enabling the identification and linkage of individuals, accounts, and organizations across fragmented datasets. However, its effectiveness is highly dependent on data quality and consistency. In regulatory contexts, inaccurate or incomplete data can propagate systemic errors across interconnected systems, particularly where identifiers vary across institutions or jurisdictions.

Empirical research demonstrates that AI effectiveness is not purely a function of algorithmic sophistication but of system design and governance (Zhang et al., 2025). Even highly advanced models fail in environments characterized by poor data governance or inconsistent data standards. This is particularly relevant for entity resolution systems, where inconsistencies can undermine cross-border intelligence efforts and limit the ability to reconstruct complex financial networks.

Pattern detection is one of the primary strengths of AI-driven financial crime systems, allowing for the identification of non-linear relationships and hidden structures within large datasets. Machine learning models can uncover behavioral patterns that would remain

undetected using traditional rule-based approaches, significantly enhancing the ability to detect fraud schemes.

However, these capabilities are closely tied to cybersecurity risks. Fraud detection models can be deliberately manipulated through adversarial inputs designed to evade detection. Studies show that adversarial techniques applied to transactional data can achieve high evasion rates while remaining difficult to detect (Cartella et al., 2021). As a result, pattern detection systems may become vulnerable to strategic manipulation, reducing their effectiveness in dynamic threat environments.

Risk scoring plays a vital role in using AI for regulatory purposes, by enabling the prioritization of transactions, customers, or entities based on their likelihood of involvement in illicit activity. Research shows that AI enables scalable risk scoring across large datasets, improving early detection and resource allocation within financial institutions (Zhang et al., 2025).

Nevertheless, the reliability of risk scoring models is highly sensitive to data quality issues such as imbalance, limited labeled datasets, and concept drift - where fraud patterns evolve over time (Zhang et al., 2025). These limitations can increase false positives or false negatives, reducing the overall effectiveness of regulatory systems. Furthermore, adversarial strategies allow sophisticated actors to “game” risk scoring models by structuring behavior to remain below detection thresholds, transforming these systems into reactive rather than preventive tools.

Predictive anomaly detection enables regulatory systems to identify unusual or suspicious behavior in real time, supporting a shift from reactive investigation to proactive monitoring. By continuously learning from historical and incoming data, these systems can detect emerging fraud typologies and previously unknown threats.

However, anomaly detection models are particularly vulnerable to cybersecurity risks. Their reliance on continuous updating makes them susceptible to data poisoning and model drift if not properly monitored. Research emphasizes that AI-driven detection systems must be understood as part of a broader cybersecurity ecosystem rather than standalone tools. The integration of adversarial learning techniques has been proposed to improve robustness (Ijiga et al., 2024), but this also increases system complexity and introduces governance challenges related to transparency and accountability.

From a regulatory perspective, these findings highlight the need to integrate cybersecurity considerations directly into AI governance. Securing training data, mitigating adversarial manipulation, and ensuring the resilience of shared intelligence systems are

essential to maintaining the effectiveness of predictive anomaly detection. Without such safeguards, AI systems risk amplifying systemic vulnerabilities rather than mitigating them.

4 Force Multiplier or Friction Amplifier?

4.1 AI as a Force Multiplier: Efficiency, Prevention, and Investigative Precision

The generative and analytical capacities that empower criminal actors are simultaneously being deployed by financial institutions and regulatory bodies to counter financial crime with unprecedented scale and precision. This section examines how AI transforms financial crime prevention across four dimensions: the displacement of rule-based transaction monitoring, deep learning-driven detection performance, KYC automation, and network-based money laundering detection. The evidence base draws on systematic reviews of machine learning applications spanning over a decade of institutional deployment.

4.1.1 The Collapse of Rule-Based Systems and the AI Transition

Rule-based transaction monitoring has proven structurally inadequate for modern financial crime detection. Hernandez Aros et al. (2024), reviewing 104 studies, report that traditional systems generated false positive rates exceeding 95%, a level of noise that paralyzes compliance teams and diverts resources from genuine threats. AI-based architectures address this directly. Yang et al. (2026) identify CNNs, LSTMs, Transformers, GNNs, and hybrid models as the primary detection frameworks across credit card fraud, loan fraud, AML, and cryptocurrency-native attacks. These models adapt to evolving fraud patterns without manual rule updates, capture non-linear behavioral signals, and exploit relational structure between entities, which capabilities rule-based systems cannot replicate. Chen et al. (2025) frame the transition as operationally necessary: with U.S. digital transactions exceeding 2.7 billion in 2023 and organizations losing approximately 5% of annual revenues to fraud globally, AI adoption is not optional.

4.1.2 Deep Learning Performance and Institutional Evidence

The performance gains of AI-based detection are well documented. Chen et al. (2025) report that an LSTM/GRU model trained on 440,000 records from a Danish bank reduced false positive rates by 33.3% while maintaining 98.8% true positive recall, a result they identify as the primary efficiency driver in AML operations. Yang et al. (2026) confirm that Graph Convolutional Networks outperform Logistic Regression, Random Forest, and Gradient

Boosting classifiers in fraud detection, as their relational architecture captures transactional topology invisible to record-level models. Hernandez Aros et al. (2024) add that hybrid supervised-unsupervised architectures detect both known fraud patterns and novel threat typologies simultaneously, a gap purely supervised systems cannot close. In the unstructured data domain, NLP contributes to SAR narrative generation and financial statement auditing by extracting signals from regulatory filings and public records, positioning it as a core tool in next-generation compliance automation.

4.1.3 KYC Transformation, Network Analysis, and Privacy-Preserving Collaboration

AI has restructured KYC and CDD from periodic, document-based procedures into continuously adaptive risk systems. Baah et al. (2025) identify adaptive risk scoring, entity linking, graph-based relationship visualization, and biometric liveness detection as the primary operational advances, the last of which directly counters the synthetic identity attacks described previously. Yang et al. (2026) confirm graph-based models achieve superior AML performance while generating outputs that support both investigator reasoning and regulatory documentation. Privacy-preserving techniques, federated learning, and homomorphic encryption, enable cross-institutional intelligence sharing without direct data disclosure, addressing the data silo problem that has historically limited network-level AML visibility. Chen et al. (2025) identify blockchain-integrated smart contracts as an emerging direction for decentralized fraud detection compliant with GDPR and CCPA.

4.2 AI as a Friction Amplifier: False Positives, Alert Fatigue, and Legal Paralysis

The integration of AI into AML and financial crime compliance has produced not only enhanced detection capability but a set of systemic dysfunctions that threaten to undermine the very objectives these systems serve. False positive cascades, exponential SAR volume growth, the erosion of investigative reasoning, escalating compliance costs, and unresolved liability exposure together constitute the friction amplifier effect. These are not peripheral implementation failures, but structural properties of deploying probabilistic AI models in the high-imbalance, adversarially-contested environment of financial crime.

4.2.1 Dataset Imbalance, False Positives, and the Noise Production Architecture

The core challenge in AI-based fraud detection is class imbalance: fraudulent transactions typically represent less than 1% of total volume. Yang et al. (2026) identify this as

the most persistent issue in the literature, as classifiers trained on imbalanced data systematically favour the majority class and underperform on fraud. At institutional scale, even low per-transaction error rates generate large volumes of false positives. AI substantially improves on the 95% false positive baseline of rule-based systems but does not eliminate noise: Baah et al. (2025) confirm that anomaly detection methods, essential for identifying novel fraud typologies, inherently flag legitimate but statistically unusual transactions. Concept drift compounds this further: as fraud tactics evolve, model decision boundaries progressively misalign, degrading precision and requiring continuous retraining that creates operational gaps. The interpretability deficit of deep learning models adds a final layer, analysts not being able to audit the reasoning behind alerts, forcing investigation of outputs they cannot independently evaluate (Hernandez Aros et al., 2024).

4.2.2 SAR Inflation, Alert Fatigue, and the Erosion of Investigative Reasoning

Kahn (2026) documents how conservatively calibrated AI monitoring systems generate alert volumes that, when fed into automated reporting workflows, produce filings driven by model outputs rather than human judgment. This is legally significant: AML regulatory frameworks require SARs to reflect reasonable grounds for suspicion, a threshold implying genuine analytical reasoning. Kahn (2026) argues AI models operating on probabilistic thresholds do not meet this standard; their narratives, while structurally coherent, derive from the same limited data that triggered the alert, creating an appearance of quality that satisfies surface-level review while failing to generate actionable intelligence. The downstream effect is Financial Intelligence Unit (FIU) congestion: excessive low-value reports reduce the signal-to-noise ratio and delay processing of genuinely significant cases. At the analyst level, sustained exposure to predominantly false alerts produces fatigue - a progressive deterioration of investigative vigilance. Kahn (2026) characterizes this as a structural shift: analysts become algorithm validators rather than independent investigators, eroding the contextual judgment that AI systems cannot replicate. Chen et al. (2025) add a technical circularity: constrained review capacity produces biased training feedback, which miscalibrates models, generating more false positives, and further constraining capacity.

4.2.3 Compliance Cost Escalation, Liability Ambiguity, and the Adversarial Paradox

AI implementation adds to, rather than replaces, legacy costs during transitions, disproportionately burdening smaller institutions and reinforcing the structural advantage of large players. Baah et al. (2025) identify these investment requirements as a formidable

operational challenge for institutions managing regulatory adherence alongside efficiency pressures. Kahn (2026) also identifies a governance gap: when SARs are generated primarily by AI, accountability for erroneous or deficient filings becomes legally ambiguous. SAR submissions carry legal obligations tied to human judgment; displacing that judgment with algorithmic output leaves the accountability chain unresolved and institutions exposed to regulatory risk. Hernandez Aros et al. (2024) note that ethical and organizational dimensions remain underrepresented in the AI fraud detection literature relative to model performance metrics, a maturity gap regulators are only beginning to address. Lastly, Yang et al. (2026) identify the adversarial paradox as a defining systemic risk: the explainability that makes AI detection systems auditable simultaneously makes them exploitable. As detection logic becomes known, criminals engineer transactions to evade learned boundaries, a new dimension of the offense-defence arms race in which transparency is itself a vulnerability. SAR inflation, alert fatigue, cost inequality, liability ambiguity, and adversarial exploitability together form a coherent set of second-order consequences that institutional governance frameworks have not yet adequately resolved.

5 The Data-Rich, Intelligence-Poor Paradox

5.1 The Financial Intelligence Bottleneck

There has been an exponential rise in Suspicious Activity Report (SAR) filings amongst numerous significant financial jurisdictions which stand in sharp contrast compared to stagnant or decreasing prosecution rates. This deviation unveils a core conflict at the heart of AML enforcement. Nearly doubling in the last ten years, FinCEN received more than 3.6 million SARs in 2022, in the United States alone. Similarly, the United Kingdom and Europe show aligned development. Nevertheless, money laundering conviction rates are lagging behind the surge in illicit financial activity; prosecution rates plummeted in the UK during the same period, even as SAR hit an all-time high (Thomson Reuters, 2023).

This gap is not merely administrative. The financial sector suffers from a ‘data-rich, intelligence-poor’ paradox: they produce endless compliance signals but fail to turn them into meaningful, actionable enforcement outcomes. Estimates suggest that fewer than 1% of laundered funds are eventually seized globally, whereas compliance costs borne by institutions continue to rise (Europol, 2014).

The structural factors represent the root of the disconnect. SARs are used by financial institutions to avoid regulatory liabilities, which on its behalf needlessly increases low-signal

reports' count. Investigative agencies face chronic resource constraints, limiting their capacity to triage, cross-reference, and develop cases from raw filings. Generally, money laundering schemes horizon goes beyond one country and/or jurisdiction. Therefore, financial crimes become subject to multiple legal regimes, further slowing down conversion (ACAMS, 2023).

AI-driven detection systems have accelerated SAR generation without proportionally improving downstream intelligence quality. Anomalies, which are well identified by pattern recognition tools, do not serve as evidence. Hence, flagged transactions cannot be bridged to a prosecutable case, solely based on anomalies. Regulatory use of AI can be very beneficial to detect suspicious activities; however, this increases the workload of investigators. These systems require more resources, especially time, and thus widening the conversion gap.

Financial intelligence and compliance institutions, such as UIF and FINTRAC, receive hundreds of thousands of SARs each year, alongside millions of CTR reports concerning large financial transactions. Meanwhile, global AML compliance spending was 274 billion USD in 2022, an increase of 28% (Bank for International Settlements, 2023).

However, an estimated 90-95 percent of SAR-related alerts are considered false positives (Saaradeey et al., 2019), placing significant strain on institutional resources. Every false positive consumes analyst time, energy, and, most importantly, masks real crimes. According to a BPI survey in the United States, across 640,000 SAR reports that have been filed, only 4% of SARs and an average of 0.44% of CTRs warranted follow-up inquiries (Bank Policy Institute, 2018). This results in financial institutions being incentivized to adopt a defensive approach and file as many reports as possible, in fear of receiving regulatory fines that can reach billions of dollars, for example, in the case of TD Bank in 2024 (Financial Crimes Enforcement Network, 2024). The bottleneck this problem creates can be resolved with a quality-based AML approach rather than a quantity-based one.

In today's world, there is continuous production of data and information:
different kinds, formats, and purposes. It is possible that, in some contexts, the data is so extensive that it is like having no data at all: above all, there is an issue when the same piece of information is presented in different ways, making disclosure less interpretable.

Indeed, having uniform regulations and a unique format to disclose economic and financial information could represent a substantial gain in efficiency, both for firms that have to dedicate less time and resources to prepare different reports for authorities with different requirements, and control institutions that would have standardized documents to analyse. For instance, simplifying reporting is one of the aims of the European Commission (ESMA Contributes to Simplification and Burden Reduction, 2025).

For this reason, in the in-depth analysis on Anti Money Laundering Authority in the EU, it is specified how, as a fundamental part of AML/CFT legislative reform, this institution aims to ensure consistent obligations and interpretations, in this framework, across all Member States, complemented by national authorities (European Parliament Think Tank, 2025).

5.2 Behavioral Barriers to Intelligence Sharing

The problem facing modern analytical systems is no longer access to information, but the ability to meaningfully process it. This condition has been conceptualized by the World Health Organization (WHO) as "infodemic" and defined as "too much information including false or misleading information in digital and physical environments during a disease outbreak" (WHO, n.d.). While WHO defined this term in an epidemic context, many have extended its logic to everyday information. Paul Hartog (2017) argued that persistent exposure to excessive information can lead to "information anxiety". This psychological response can be described by feelings of overwhelm, uncertainty, or disorientation. In a 'data-rich, intelligence-poor' environment, the challenge of saturation of information has distinct psychological consequences for analysts operating in Financial Intelligence Units (FIUs). Continuous exposure to high volumes of suspicious transaction reports (STRs), alerts, and raw financial data can lead to cognitive overload, where the human capacity to filter and interpret information becomes significantly strained (Quattash, 2025). Under such conditions, decision-making tends to shift from in-depth reasoning towards shortcuts. This transition increases the risk of missed patterns or superficial assessments, ultimately undermining the quality and reliability of outputs (Hirshleifer et al., 2018).

A major behavioral problem in the 'data-rich, intelligence-poor' environment is defensive reporting culture, where institutions report suspicious activity to protect themselves from risk of regulatory criticism or legal exposure, rather than because the report is likely to produce meaningful intelligence value. Instead of encouraging careful judgment and high-quality reporting, the compliance rules can prompt institutions to report excessively by providing them with incentives, especially when the costs of not producing enough reports are estimated to be higher than costs of over-reporting. United States' SAR guidelines explicitly protect institutions that file suspicious activity reports, even when the case is uncertain, which explains why firms may choose to report defensively even when they are doubtful (FFIEC, 2021).

When large volumes of low-value reports are submitted, FIUs and law-enforcement agencies must spend more time filtering out valuable information, which increases analyst overload and weakens the conversion of raw reporting into actionable intelligence (National Crime Agency, 2024). FATF and the Egmont Group have noted that FIUs require stronger technological and operational capacity precisely because suspicious activity detection and financial intelligence analysis are under pressure from growing data volumes (FATF & Egmont Group, 2021).

From a behavioral perspective, defensive reporting culture reflects institutional risk aversion, weak feedback loops, and a compliance environment which encourages submission of reports for formal protection rather than investigative usefulness, contributing directly to the 'data-rich, intelligence-poor' problem, where more information is produced, but less of it is genuinely useful for detecting, prioritizing, and preventing financial crime (FATF & Egmont Group, 2021).

Most data-rich financial institutions work with an abundance of information, ranging from transaction records and customer profiles to behavioural signals, while simultaneously remaining 'intelligence-poor' in dealing with financial crime and harmful actions committed against them.

One of the key issues is risk aversion in data sharing, as financial institutions operate in strictly regulated environments where the leak of important data can lead to heavy consequences such as legal liability, reputational damage, and other regulatory penalties. In these scenarios of overwhelming potential costs, most companies adhere to loss aversion and prefer to keep a secretive approach of isolation in exchange for potential gains from collaborative intelligence and possibilities of identifying cross-institutional fraud patterns (KANTAR, 2020).

An important consequence of this is the institutional 'ownership' of data, which goes from being a collective resource to merely a proprietary asset. Behavioural experts define this phenomenon as the "endowment effect", implying that institutions overvalue data due to the fact that it is of their possession and are reluctant to share it despite beneficial cross-institutional outcomes. These communication gaps between institutions become vulnerable loopholes that financial criminals can exploit to commit fraud, generate synthetic identities, and use the same tactics across multiple platforms. Personal AI systems that most financial institutions adopt are not sufficient for proper detection due to data constraints and limited scope of knowledge beyond the organisation's experience. Fragmentation of defenders is thus a weakness that works in favor of aggregated and adaptive crime perpetrators (Ganti, 2026).

An additional factor is diffusion of responsibility, as no single financial institution desires to retain accountability for a systemic risk that is a threat to the whole industry. Investments in shared intelligence infrastructure are contingent on the lack of regulations regarding data sharing and thus require collective effort to overcome the current isolation of institutions in their battle against AI-driven crime. Risk exposure can be reduced through institutional cooperation and jointly developed privacy-preserving technologies, while simultaneously reframing the perception of data as a shared defense source against AI threats, rather than a competitive asset. Since AI-driven financial crime thrives on these difficulties of risk sharing and data ownership, it is critical to address these behavioural issues and ensure that institutions become not just data-, but also intelligence-rich (LexisNexis Risk Solutions, 2023).

5.3 Legal Barriers to Financial Intelligence Sharing

A key issue lies in the legal conflict between data protection and anti-money laundering requirements. On the one hand, the GDPR imposes strict limitations on data processing, requiring that personal data must be limited to what is necessary and used only for specific purposes (European Union, 2016). However, in the case of anti-money laundering legislation, financial organizations need to collect and use extensive information to identify illegal transactions. This creates what has been described as a “delicate balancing act” between privacy and financial crime prevention (DigitalEurope, 2020), where institutions must comply with two regulatory ideas that are not fully aligned.

This issue is further complicated by legal fragmentation across jurisdictions. Even though the GDPR was intended to be harmonized, companies that operate in several Member States may find themselves with overlapping laws and uncertainties on jurisdictional applicability (Nolan, 2018). At the same time, anti-money laundering rules were always enacted as directives and there was never aligned consistency in their interpretations and implementations among states. This is partly due to the absence of clarity on the regulatory process and partly because of the inconsistent implementation processes that make data sharing and coordination very difficult. Consequently, data is scattered across organisations in various locations but fails to be compiled into an intelligence system.

While it has become widely accepted that the advancement and deployment of artificial intelligence pose significant societal risks, several large AI companies preserve a relatively rigid approach to practices that can expose the vulnerabilities of their models (Longpre et al., 2024). Due to the complexity of modern AI models, many types of testing are essential to

uncover safety threats, making independent risk assessment, together with testing conducted by external third parties, particularly relevant in the current landscape. These practices are commonly referred to as bug bounty programs, during which companies invite external researchers to identify security weaknesses in their systems, in exchange for monetary rewards, while providing legal protection that guarantees exemption from all legal consequences associated with their work.

Certain AI developers, such as OpenAI and Google, provide monetary rewards and legal protection to external researchers acting in good faith (OpenAI, 2026; Google, n.d.), while others, such as Meta and Anthropic, reserve the discretion to determine whether the action was motivated by good faith and complies with their internal policies (Meta, n.d.; Anthropic, 2025). This introduces legal uncertainty and may discourage research conducted in good faith. Therefore, in the absence of clear and uniform safe harbor guarantees, together with objective determinations of good faith, legitimate research may be deterred, potentially leaving critical vulnerabilities undiscovered.

Oftentimes, data localization requirements pose a constraint when it comes to transforming data into actionable intelligence, by restricting cross-border transfers and processing of data. Such rules are rightly justified by the need for privacy, national sovereignty, and security, although they generate difficulties when it comes to financial crime investigations, which are often transnational. Integrated datasets spanning multiple nations are usually required to detect illegal financial flows by tracking transactions. Localization rules, however, cause these data sets to be fragmented into national silos, making it difficult for authorities and institutions to reconstruct financial networks (Europol, 2017).

Due to this fragmentation, investigations are weakend as investigators are forced to rely on partial data or on cross-border cooperation, which is oftentimes too slow, delaying analysis and hurting the accuracy of findings. Consequently, AI models that require cross-jurisdictional and large data sets are trained on insufficient data, reducing their predictive accuracy. Therefore, although wide amounts of data exist, legislation often prevents their integration into meaningful intelligence, reducing the full potential of AI as well as human solutions to financial crime (Cory, 2017).

5.4 Cybersecurity and the Role of Fragmented Infrastructure

The central cyber problem behind the ‘data-rich, intelligence-poor’ constraint is not that financial institutions and regulators lack data, but that they struggle to convert fragmented

digital information into timely and usable intelligence, as previously reiterated. FIUs, banks, and supervisory authorities now collect vast quantities of suspicious transaction reports, customer due diligence records, payment data, and cross-border transaction information. FATF and the Egmont Group note that operational agencies are already using automation, large datasets, and AI tools, yet the core difficulty remains the same: data arrives from disparate sources and differently structured databases, making rapid prioritisation and analysis difficult (FATF & Egmont Group, 2021). Europol captured this imbalance, reporting that FIUs within the European Union received almost one million reports per year, while only a small share of those were investigated further. In that sense, the AML system is often rich in raw reporting but poor in actionable intelligence (Europol, 2017).

This gap is partly a cyber-infrastructure problem. Legacy systems, incompatible reporting formats, weak API standardisation, and inconsistent access rules across jurisdictions prevent investigators from identifying links and trends between accounts, companies, and transactions that may belong to the same criminal network (Allegrezza, 2022; CPMI, 2024). The European Parliament's study on the proposed AML Authority argues that reform is needed to improve coordination and information sharing among FIUs, including through common technical standards, common suspicious transaction report templates, and stronger management of FIU.net (Allegrezza, 2022). At the level of payment infrastructure, the BIS makes a similar point: APIs are increasingly central to the electronic exchange of payment data, but without harmonisation, they risk reinforcing fragmentation rather than solving it (CPMI, 2024). Poor interoperability therefore limits not only speed, but also the ability to build a coherent cross-border picture of financial crime (Allegrezza, 2022; CPMI, 2024).

Artificial intelligence can help, but mainly as a support layer rather than a complete solution. FATF and the Egmont Group observe that machine learning can partially automate risk analysis over large volumes of unstructured data and help identify emerging risks that do not fit already known profiles (FATF & Egmont Group, 2021). This is particularly important in AML because suspicious behaviour is often relational: laundering schemes are hidden not in one payment, but in patterns across people, firms, and accounts (Weber et al., 2018; Eddin et al., 2021). Eddin et al. show that a graph-based triage model tested on a real banking dataset reduced false positives by 80% while still detecting more than 90% of true positives (Eddin et al., 2021). Weber et al. similarly argue that graph learning is promising because financial data is massive, dense, and dynamic (Weber et al., 2018). At the same time, the literature still stresses that AML systems must remain interpretable and explainable if investigators are to trust and operationalise their outputs (Bellomarini et al., 2020).

For that reason, the cyber-oriented answer to financial crime is not simply ‘more AI’. More important reforms may lie in the architecture around AI: interoperable databases, standardised APIs, shared reporting templates, privacy-compliant information-sharing mechanisms, and better entity resolution through common identifiers (FATF, 2022; FSB, 2024; Allegrezza, 2022). FATF has argued that stronger information sharing can be designed consistently with data-protection rules, while the FSB’s recent work on the Legal Entity Identifier shows how standardised identifiers can support automated reconciliation and validation in cross-border payments (FATF, 2022; FSB, 2024). AI can improve triage and pattern recognition, but it becomes genuinely effective only when embedded in an infrastructure that allows data to move, match, and be interpreted across institutions and jurisdictions (FATF, 2022; CPMI, 2024; FSB, 2024). The real cyber challenge, then, is transforming fragmented technical systems into a common intelligence architecture (FATF & Egmont Group, 2021; Allegrezza, 2022).

6 Proposed Reforms

AI has the potential to change the financial crime detection landscape and assist public and financial institutions in their endeavours to limit financial losses. However, the current results of AI implementation are largely reactive, rather than preventive. In practice, AI relies on fragmented and incomplete institutional datasets, generates alerts only after suspicious behaviour has occurred, sends false positive signals, and highlights mostly low-priority cases to be further investigated by specialists (FATF, 2022a). This reactive model is decisively inadequate, since nowadays, criminals operate across institutions and borders, while most monitoring remains isolated and fails to point out suspicious network patterns that matter.

Therefore, there is a need for a preventive intelligence architecture that would enable AI to act earlier, across institutions, and with privacy protection built in. Instead of relying on isolated monitoring (accomplished by separated bodies), AI should have access to a larger network of data (i.e. payments, transactions, sources of funding, actors), while employing a privacy-enhancing technology to protect the identities of those whose information is collected (BIS, 2024a).

AI will not become fully preventive simply by improving algorithms inside existing compliance silos. Structural modernization is required because anti-financial crime systems suffer from fragmented data, inconsistent policies, privacy constraints, and weak interoperability between institutions and authorities (BIS, 2024b). In the current environment,

AI is asked to predict risk without having access to the relational and cross-institutional information that generates it.

6.1 Federated Data-Sharing Models

The first reform proposed is the development of federated data-sharing models. Federated models allow institutions to collaborate on risk analysis without centralizing all raw customer data in one place (Abadi et al., 2024). Instead of sharing raw data, governments can build their own models and share just the model parameters, while a central server aggregates all of them. (Kilian Sprenkamp et al., 2024). These models are highly important for detecting money laundering schemes, which are typically distributed across banks, intermediaries, and jurisdictions. If institutions collaborate through federal architectures, AI can learn broader suspicious typologies and network signals, while each participant retains data control (Yang et al., 2023). The debate is thus moved away from "should institutions share data?" to "how can they share data safely and productively?" However, the reform has some practical limitations. Federated systems are technically difficult to be implemented in real financial environments because institutions hold highly heterogeneous data, by having access to different customer populations, transaction types, and risk distributions (FATF, 2022b). A second limitation is the governance complexity required. Even if raw data never leaves each institution, participants still need common definitions of suspicious behavior and trusted procedures for investigating the outputs. There are also important data-privacy concerns (BIS, 2024c). Research increasingly suggests that even current model updates may still leak sensitive information under certain attack conditions.

6.2 Tiered SAR Thresholds

A second reform is the introduction of tiered SAR thresholds. One weakness of the current reactive AI is that it generates excessive volumes of false positives, consuming investigator's time and reducing the informational value of all the suspicious activity reports previously generated (Federal Financial Institutions Examination Council [FFIEC], 2021a). A tiered SAR model can respond to this by differentiating between levels of suspicion: cases that raise limited concern should be kept under closer monitoring, those that appear more suspicious should lead to internal escalation and deeper checks, while only the most serious cases must be forwarded to a formal SAR filing (Geiger & Wu, 2024). The advantage of tiering is the efficient

allocation of investigative attention and resources. In preventive architecture, AI mission should not be to produce as many alerts as possible, but to help institutions intervene earlier, reduce background distractions, and highlight the most probative cases for action. Among limitations, the most important is under-reporting. Allowing institutions full discretion to handle lower or medium risk cases internally, rather than reporting them immediately, may result in some genuinely suspicious activities not being communicated to the competent authorities in a timely manner (FFIEC, 2021b). A second drawback is the possible institutional inconsistency. Tiered thresholds may look proficient in theory, but in practice institutions may define “low”, “medium”, and “high” risk cases in different ways, which can lead to uneven reporting quality. Regarding data privacy concerns, a tiered SAR framework indeed involves the fact that institutions would keep information for longer in their internal monitoring systems before deciding whether to fill a report, but this does not possess a serious threat for data leakage.

6.3 Improving Coordination and Trust in AI-based Detection

To improve AI-based resolution to financial fraud, preventive intelligence architecture can be used to address the structural limitations that prevent AI outputs from being interpreted and shared across institutions, including transnational institutions. One of the main constraints is the lack of standardized APIs, which limit data exchanges between financial institutions, FIUs, and supervisory authorities. Unless institutions adopt common data formats and standardized communication protocols, standardized APIs, data, and information will remain fragmented, further reducing the speed and scalability of AI-driven analysis. With compatible communication systems, real-time data flows would be much more efficient, and the integration of AI tools would be easier to sustain across different jurisdictions (European Court of Auditors, 2021; FATF, 2021).

Alongside the optimization of AI operability, the improvement of AI model explainability would be beneficial to increase trust and action related to insights generated by AI. Many AI systems rely on models whose decision-making process is not transparent or understandable to humans, named “black box” models. These models only produce an output after being given an input, without showcasing the process of getting there. Thus, they produce alerts without explaining the underlying drivers of risk. Thus, usability of these otherwise effective models is severely limited by the lack of transparency, making them untrustworthy to investigators and regulators. If these systems were explainable, institutions and regulators

would be able to understand the factors behind suspicious activity, improving defenses and investigations of financial crime, as well as information sharing (BIS, 2024).

Lastly, these improvements may also be enhanced by greater cross-border supervisory harmonisation. Financial crime happens across multiple borders, though enforcement is mostly handled at the national level, creating a data and information gap. As money is moved across borders by criminals, national governments, FIUs, and international bodies do not always coordinate their investigations, decreasing their efficiency. Usually, the cause of this is the difference in regulations, data access, and supervisory practices, causing information to be more difficult to share in an efficient way, limiting the effectiveness of investigations. Improving coordination between these institutions is, therefore, essential. This could consist of aligning regulatory standards, simplifying data-sharing procedures, and strengthening cooperation mechanisms across countries.

The role of international organisations, such as the Anti-Money Laundering Authority (AMLA), becoming more prevalent could facilitate better communication between national bodies as well as provide consistent and unbiased oversight. Thus, by reducing fragmentation and increasing the efficiency of communication between institutions, cross-border illegal activity and patterns could be more rapidly and accurately detected by authorities, making the use of AI systems in financial crime detection more relevant and instrumental (Europol, 2022; European Court of Auditors, 2021; FATF, 2021).

7 Conclusion

This paper set out to understand how artificial intelligence changes the environment in which regulators and perpetrators operate by analysing both its positive and negative impacts on regulators, as well as the benefits it provides to criminals. Across all four perspectives, namely financial, behavioural, legal, and cyber, the findings present balanced arguments for both parties. However, it is evident that criminals are currently better-equipped and more agile to utilise the benefits of AI to their highest potential, regulators being confined by more rigid and traditional restrictions. Although regulators and institutions benefit from better pattern recognition, anomaly detection, and data organisation, this advantage is unfortunately compromised and dulled by the lowering of barriers to entry for fraud and deception schemes that perpetrators can use at lower cost and complexity.

It is evident that, with greater inter-agency collaboration and filtration of data, regulators can become more flexible and adapt quicker to this dynamic technological

environment. However, at this current premature stage, the noise and over-crowding of data from automated transaction monitoring significantly limits the benefits that can be extracted from the advancements in AI. Until these concerns are remedied and individuals become more confident and trusting with the presence and use of AI, its benefits will be noticeably capped.

In summary, this paper introduces a balanced and unbiased perspective on the effects and implications of artificial intelligence and machine learning on both the regulator and perpetrator in financial crime networks. Through the analysis of the two parties and contextual environment, the introduction of federated data-sharing models, tiered SAR thresholds, and the improvement of coordination and trust in AI-based detection is proposed in order to reduce the limits that AI imposes on regulators, so that institutions are better equipped to withstand pressures and developments of perpetrators.

8 Bibliography

Abadi, M., et al. (2024). *Starlit: Privacy-preserving federated learning to enhance financial fraud detection*. arXiv. https://arxiv.org/abs/2401.10765

Abdelaziz, D. A. (2025). *Criminal liability for the misuse and crimes committed by AI: A comparative analysis of legislation and international conventions*. Journal of Infrastructure, Policy and Development, 9, Article 10722. https://doi.org/10.24294/jipd10722

Abel Souto, M. (2026). *Challenges of artificial intelligence and money laundering in the application of international criminal law*. Journal of Illicit Trade, Financial Crime, and Compliance.

ACAMS. (2023). *Money laundering reporting failures rarely prosecuted in Britain*. https://www.acams.org/en/news/money-laundering-reporting-failures-rarelyprosecuted-in-britain

Adhikari, S., & Hamal, K. (2024). *Artificial intelligence in fraud detection: Revolutionizing financial security systems*. International Journal of Science and Research Archive. https://ijsra.net/sites/default/files/fulltext_pdf/IJSRA-2024-1860.pdf

Agnew, H. (2020). Jim Chanos: "*We are in the golden age of fraud*." Financial Times. https://www.ft.com/content/ccb46309-bba4-4fb7-b3fa-ecb17ea0e9cf?syn-25a6b1a6=1

Ahmad Naser Eddin, et al. (2021). *Anti-money laundering alert optimization using machine learning with graphs*.

Allegrezza, S. (2022). *The proposed Anti-Money Laundering Authority, FIU cooperation, powers and exchanges of information: A critical assessment*. European Parliament.

Alon-Barkat, S., & Busuioc, M. (2023). *Human–AI interactions in public sector decision making: "Automation bias" and "selective adherence" to algorithmic advice*. Journal of Public Administration Research and Theory, 33(1), 153–169. https://academic.oup.com/jpart/article/33/1/153/6524536

Anthropic. (2025). *Responsible disclosure policy*. www.anthropic.com/responsible-disclosure-policy

Artificial Intelligence Act. (2026). Article 50: *Transparency obligations for providers and deployers of certain AI systems*. https://artificialintelligenceact.eu/article/50/

Baah, E. K., et al. (2025). *Role of AI in preventing financial crime: A comprehensive analytical review*. Scientific African. https://doi.org/10.1016/j.sciaf.2025.e02764

Bank for International Settlements. (2023). *The power of data, technology and collaboration to combat money laundering across institutions and borders*. https://www.bis.org/publ/othp66.pdf

Bank for International Settlements. (2024). *Project Aurora: The power of data, technology and collaboration to combat money laundering*. BIS Innovation Hub. https://www.bis.org/publ/othp66.pdf

Bank for International Settlements. (2024). *Project Mandala: Streamlining cross-border transaction compliance*. BIS Innovation Hub. https://www.bis.org/publ/othp87.pdf

Bank for International Settlements. (2024). *Regulating AI in the financial sector: Recent developments and main challenges*. https://www.bis.org/fsi/publ/insights63.pdf

Bank of England. (2024). *Artificial intelligence in UK financial services*. https://www.bankofengland.co.uk/report/2024/artificial-intelligence-in-uk-financialservices

Bank Policy Institute. (2018). *Getting to effectiveness: Report on U.S. financial institution resources devoted to BSA/AML and sanctions compliance*. https://bpi.com/getting-to-effectiveness-report-on-u-s-financial-institution-resourcesdevoted-to-bsa-aml-sanctions-compliance/

Begolli, G. (2019). *Revealing the true cost of financial crime, focus on Europe*. Knowledge International Journal, 31(5), 1463–1468. https://doi.org/10.35120/kij31051463b

Bellomarini, L., Laurenza, E., & Sallinger, E. (2020). *Rule-based anti-money laundering in financial intelligence units: Experience and vision*.

Belotti, S. (2025). Italian Law No. 132/2025: *The first domestic law in the European Union on the use of artificial intelligence: Insights*. Squire Patton Boggs. https://www.squirepattonboggs.com/insights/publications/italian-law-no-132-2025the-first-domestic-law-in-the-european-union-on-the-use-of-artificial-intelligence/

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2020). *Explainable AI in fintech risk management*. Frontiers in Artificial Intelligence, 3, Article 26. https://doi.org/10.3389/frai.2020.00026

Cartella, F., Anunciacao, O., Funabiki, Y., Yamaguchi, D., Akishita, T., & Elshocht, O. (2021). *Adversarial attacks for tabular data: Application to fraud detection and imbalanced data*. arXiv. https://arxiv.org/abs/2101.08030

Chen, Y., Zhao, C., Xu, Y., Nie, C., & Zhang, Y. (2025). *Year-over-year developments in financial fraud detection via deep learning: A systematic literature review*. arXiv. https://arxiv.org/abs/2502.00201

Chesney, R., & Citron, D. (2019). *Deepfakes and the liar's dividend*. https://www.deccanherald.com/opinion/deepfakes-and-the-liars-dividend-3497282

Colucci, D., Franco, C., & Valori, V. (2021). *Endowment effects at different time scenarios: The role of ownership and possession*. https://www.ec.unipi.it/documents/Ricerca/papers/2021-279.pdf

Committee on Payments and Market Infrastructures & Bank for International Settlements. (2024). *Promoting the harmonisation of application programming interfaces to enhance cross-border payments*.

Cory, N. (2017). *Cross-border data flows: Where are the barriers?* Information Technology and Innovation Foundation. https://itif.org/publications/2017/05/01/crossborder-data-flows-where-are-barriers-and-what-do-they-cost/

Cozzi, F., & Gerbino, A. (2025). *Italy becomes the first in Europe to set national AI rules*. McDermott Will & Emery. https://www.mcdermottlaw.com/insights/italy-becomes-the-first-in-europe-to-setnational-ai-rules/

Czybik, S., Kouam, A. J., Heubl, P., Nold, J. M., & Rieck, K. (2026). *A largescale study of personalized phishing using large language models*. https://mlsec.tuberlin.de/docs/2026-sec.pdf

David, H., & Badi, S. (2025). *Harnessing artificial intelligence for regulatory compliance: Transparent systems to combat financial crimes and ensure governance*. https://doi.org/10.13140/RG.2.2.29633.06241

Deloitte. (2023). *Deepfake fraud: The emerging threat of synthetic media in financial crime*. Deloitte Insights. https://www2.deloitte.com

DigitalEurope. (2020). *Almost two years of GDPR: Celebrating and improving the application of Europe's data protection framework.* https://www.digitaleurope.org/resources/

Directive (EU) 2018/843. (2018). *Directive (EU) 2018/843 of the European Parliament and of the Council*. https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32018L0843

Drake, F., Forsström, D., & Granfeldt, P. S. (2025). *How AI is reshaping the future of transaction monitoring*. EY. https://www.ey.com/en_se/insights/financialservices/how-ai-is-reshaping-the-future-of-transaction-monitoring

Dsouza, D. S., Hajjar, A. E., & Jahankhani, H. (2024). *Deepfakes in social engineering attacks. In Cybersecurity, privacy and freedom protection in the connected world*. Springer. https://link.springer.com/content/pdf/10.1007/978-3-03164045-2_8

ESMA. (2025). *ESMA contributes to simplification and burden reduction*. https://www.esma.europa.eu/press-news/esma-news/esma-contributessimplification-and-burden-reduction

European Commission. (2019). *Ethics guidelines for trustworthy AI. Shaping Europe's Digital Future*. https://digital-strategy.ec.europa.eu/en/library/ethicsguidelines-trustworthy-ai

European Court of Auditors. (2021). *EU efforts to fight money laundering in the banking sector are fragmented and insufficiently implemented*. https://www.eca.europa.eu/en/publications/sr21_13

European Parliament. (2025). *The future of anti-money laundering in the European Union*. https://www.europarl.europa.eu/thinktank/en/document/ECTI_IDA(2025)773721

European Union. (2016). *Regulation (EU) 2016/679: General Data Protection Regulation*. https://gdpr.eu

European Union. (2018). *Directive (EU) 2018/843 of the European Parliament and of the Council of 30 May 2018 amending Directive (EU) 2015/849 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, and amending Directives 2009/138/EC and 2013/36/EU*. Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018L0843

European Union. (2021). *Aims and values*. https://europeanunion.europa.eu/principles-countries-history/principles-and-values/aims-andvalues_en

European Union. (2024). *Regulation (EU) 2024/1689: Artificial Intelligence*

Act. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Europol. (2014). *Global anti-money laundering framework: Europol report reveals poor success rate and offers ways to improve*. https://www.europol.europa.eu/media-press/newsroom/news/global-anti-moneylaundering-framework-%E2%80%93-europol-report-reveals-poor-success-rate-andoffers-ways-to-improv

Europol. (2017). *From suspicion to action: Converting financial intelligence into greater operational impact*. https://www.europol.europa.eu/publicationsevents/publications/suspicion-to-action-converting-financial-intelligence-greateroperational-impact

Europol. (2022). *EU serious and organised crime threat assessment (SOCTA).* https://www.europol.europa.eu/publications-events/main-reports/soctareport

Europol. (2022). *Facing reality? Law enforcement and the challenge of deepfakes. Publications Office of the European Union*. https://www.europol.europa.eu/publications-events/publications/facing-reality-lawenforcement-and-challenge-of-deepfakes

Europol. (2023). *The second quantum revolution: The impact of quantum computing and quantum technologies on law enforcement*. Publications Office of the European Union.

Falade, P. V. (2023). *Decoding the threat landscape: ChatGPT, FraudGPT, and WormGPT in social engineering attacks*. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 9(5), 185–198. https://doi.org/10.32628/CSEIT2390533

FATF & Egmont Group of Financial Intelligence Units. (2021). *Digital transformation of AML/CFT for operational agencies*. https://www.fatfgafi.org/en/publications/Digitaltransformation/Digital-transformation-aml-cft.html

Federal Financial Institutions Examination Council. (2021). *Assessing compliance with BSA regulatory requirements: Suspicious activity reporting*. https://bsaaml.ffiec.gov/docs/manual/06_AssessingComplianceWithBSARegulatoryRequirements/04.pdf

Financial Action Task Force. (2015). *Financing of ISIL*. https://www.fatfgafi.org/content/dam/fatf-gafi/reports/Financing-of-the-terrorist-organisation-ISIL.pdf

Financial Action Task Force. (2021). *Opportunities and challenges of new technologies for AML/CFT*. https://www.fatf-gafi.org/content/dam/fatfgafi/guidance/Opportunities-Challenges-of-New-Technologies-for-AML-CFT.pdf

Financial Action Task Force. (2021). *Stocktake on data pooling, collaborative analytics and data protection*. https://www.fatf-gafi.org/en/publications/Digitaltransformation/Data-pooling-collaborative-analyticsdata-protection.html

Financial Action Task Force. (2022). *Partnering in the fight against financial crime: Data protection, technology and private sector information sharing*. FATF. https://www.fatf-gafi.org/content/dam/fatf-gafi/guidance/Partnering-int-the-fightagainst-financial-crime.pdf

Financial Action Task Force. (2025). *Artificial intelligence and deepfakes: Impacts on money laundering, terrorist financing and proliferation financing: Horizon scan*. FATF.

Financial Action Task Force. (2025). *Horizon scan: AI and deepfakes*. https://www.fatf-gafi.org/en/publications/Methodsandtrends/horizon-scan-aideepfake.html

Financial Crimes Enforcement Network. (2024). *FinCEN assesses record $1.3 billion penalty against TD Bank*. https://www.fincen.gov/news/newsreleases/fincen-assesses-record-13-billion-penalty-against-td-bank

Financial Stability Board. (2024). *Implementation of the Legal Entity Identifier: Progress report*.

Gallo, L., Gentile, D., Ruggiero, S., Botta, A., & Ventre, G. (2023). *The human factor in phishing: Collecting and analyzing user behavior when reading emails*. Computers & Security. https://doi.org/10.1016/j.cose.2023.103671

Ganti, A. (2026). *Understanding the endowment effect: Causes, examples and impacts*. Investopedia. https://www.investopedia.com/terms/e/endowmenteffect.asp

Geiger, M., & Wu, E. (2024). *Transaction monitoring in anti-money laundering: A qualitative analysis of organizational challenges*. Computers & Security. https://www.sciencedirect.com/science/article/pii/S0167739X24002607

Giannini, A. (2023). *Criminal behavior and accountability of artificial intelligence systems* [Doctoral dissertation, Maastricht University & University of Florence]. https://flore.unifi.it/bitstream/2158/1345544/1/A.%20Giannini-Criminal%20Behavior%20and%20Accountability%20of%20Artificial%20Intelligence%20Systems.pdf

Google. (n.d.). *Google and Alphabet vulnerability reward program (VRP) rules*. Google Bug Hunters. https://bughunters.google.com/about/rules/googlefriends/google-and-alphabet-vulnerability-reward-program-vrp-rules

Gültekin, D. G., & Siniksaran, E. (2025). *Overconfidence, gender, and authority attribution in algorithmic versus human advice: A cognitive perspective on compliance*. Acta Psychologica. https://www.sciencedirect.com/science/article/pii/S0001691825009813

Hartog, P. (2017). A *generation of information anxiety: Refinements and recommendations*. The Christian Librarian, 60(1). https://digitalcommons.georgefox.edu/tcl/vol60/iss1/8

Heiding, F., Lermen, S., Kao, A., Schneier, B., & Vishwanath, A. (2024). *Evaluating large language models' capability to launch fully automated spear phishing campaigns: Validated on human subjects*. arXiv. https://doi.org/10.48550/arXiv.2412.00586

Hernandez Aros, L., Bustamante Molano, L. X., Gutierrez-Portela, F., Moreno Hernandez, J. J., & Rodríguez Barrero, M. S. (2024). *Financial fraud detection through the application of machine learning techniques: A literature review*. Humanities and Social Sciences Communications, 11, Article 1130. https://doi.org/10.1057/s41599-024-03606-0

Hirshleifer, D., Levi, Y., Lourie, B., & Teoh, S. H. (2018). *Decision fatigue and heuristic analyst forecasts* (NBER Working Paper No. 24293). National Bureau of Economic Research. https://doi.org/10.3386/w24293

Ibitola, J. (2026). *How AI is transforming AML compliance for the future*. Flagright. https://www.flagright.com/post/ai-and-the-future-of-aml-compliance

InnReg. (2025). *AI in financial services: Use cases and regulatory compliance*. https://www.innreg.com/blog/ai-in-financial-services

International Association of Engineering and Management Education. (2023). *Leveraging artificial intelligence for countering financial crimes*. IAEME Publication.

International Committee of the Red Cross. (2023, July 20). *Rules of war*. https://www.icrc.org/en/document/FAQ-rules-of-war-ihl

Ijiga, O. M., Idoko, I. P., Ebiega, G. I., Olajide, F. I., Olatunde, T. I., & Ukaegbu, C. (2024). *Harnessing adversarial machine learning for advanced threat detection: AI-driven strategies in cybersecurity risk assessment and fraud prevention*. Open Access Research Journal of Science and Technology, 11(1), 001–004 https://doi.org/10.53022/oarjst.2024.11.1.0060

Kahn, F. (2026, April). *Why AI-driven SAR explosion is quietly breaking financial intelligence*. FinCrime Central. https://fincrimecentral.com/ai-saroverproduction-risk-fiu-effectiveness/

Kantar, Department for Digital, Culture, Media & Sport. (2020). *Motivations for and barriers to data sharing*. https://assets.publishing.service.gov.uk/media/5ef4b5e6e90e075c5cc9d7e9/_Kantar_research_publication.pdf

Krishnan, A. (2016). *Killer robots: Legality and ethicality of autonomous weapons*. Routledge. https://books.google.it/books?id=UNkFDAAAQBAJ

Gazzetta Ufficiale della Repubblica Italiana. (2025). *Law No. 132/2025*. https://www.gazzettaufficiale.it/

LexisNexis Risk Solutions. (2023). *Collaboration can help fight fraud*. https://risk.lexisnexis.com/global/en/insights-resources/article/collaboration-can-helpfight-fraud

Longpre, S., Kapoor, S., Klyman, K., Ramaswami, A., Bommasani, R., Narayanan, A., Liang, P., & Henderson, P. (2024). *A safe harbor for AI evaluation and red teaming*. Knight First Amendment Institute. https://knightcolumbia.org/blog/asafe-harbor-for-ai-evaluation-and-red-teaming

Madiega, T. (2023, February). *Artificial intelligence liability directive (PE 739.342)*. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_EN.pdf

Marchant, G. E., & Allenby, B. (2017). *Soft law: New tools for governing emerging technologies*. https://www.researchgate.net/publication/313788116_Soft_law_New_tools_for_governing_emerging_technologies

Mariuz, G., Pacitti, A., & Albanese, A. (2025). *Italy's AI law: The good, the bad, and the actual substance*. Hogan Lovells. https://www.hoganlovells.com/en/publications/italys-ai-law-the-good-the-badand-theactual-substance

Martin, A., Conca, S. D., Kosta, E., & Roussos, M. (n.d.). *Chapter 10: How technology and law interact to combat financial crime*.

Meda, K. (2025, February 21). *Fraud-as-a-service: Creating a new breed of fraudsters*. Thomson Reuters Institute. https://www.thomsonreuters.com/enus/posts/corporates/faas-new-fraudsters/

Meta. (n.d.). *Program terms. Meta Bug Bounty*. https://bugbounty.meta.com/terms/

Montañez, R., Golob, E., & Xu, S. (2020). *Human cognition through the lens of social engineering cyberattacks*. Frontiers in Psychology. https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2020.01755/ful l

Nasser, G., Morrison, B. W., Bayl-Smith, P., Taib, R., Gayed, M., & Wiggins, M. W. (2020). *The role of cue utilization and cognitive load in the recognition of phishing emails*. https://pmc.ncbi.nlm.nih.gov/articles/PMC7931973/

Naz, A., Sarwar, M., Kaleem, M., Mushtaq, M. A., & Rashid, S. (2024). *A comprehensive survey on social engineering attacks, countermeasures, and mitigation on social media platforms*. International Journal of Advanced and Applied Sciences.https://www.science-gate.com/IJAAS/Articles/2024/2024-1104/1021833ijaas202404016.pdf

Nolan, K. (2018). GDPR: *Harmonization or fragmentation?* Berkeley Technology Law Journal. https://btlj.org/2018/01/gdpr-harmonization-orfragmentation-applicable-law-problems-in-eu-data-protection-law/

OpenAI. (2026). *Bug bounty program*. Bugcrowd. https://bugcrowd.com/openai

Oyedokun, O., Ewim, S. E., & Oyeyemi, O. P. (2024). *A comprehensive review of machine learning applications in AML transaction monitoring*. International Journal of Engineering Research and Development, 20(11), 173–143.

Oztas, B., Cetinkaya, D., Adedoyin, F., Budka, M., Aksu, G., & Dogan, H. (2024). *Transaction monitoring in anti-money laundering: A qualitative analysis and points of view from industry*. Future Generation Computer Systems, 159, 161–171.

Parida, S., Neeraj, Rawat, H., & Agrawal, R. (2025). *Decoding cyber fraud schemes and the rise of "fraud-as-a-service."* In 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN) (pp. 538-542). IEEE. https://doi.org/10.1109/CICTN64563.2025.10932616

Pedersen, K. T., Pepke, L., Stærmose, T., Papaioannou, M., Choudhary, G., & Dragoni, N. (2025). *Deepfake-driven social engineering: Threats, detection techniques, and defensive strategies in corporate environments*. Journal of Cybersecurity and Privacy, 5(2),

Article 18. https://www.mdpi.com/2624-800X/5/2/18 PricewaterhouseCoopers. (2024). *EMEA AML Survey 2024*. PwC. https://www.pwc.lu/en/financial-crime/anti-money-laundering/aml-survey-2024lux.html

Quattash, M. S. (2025). *The impact of cognitive load on decision-making efficiency*. Global Council for Behavioral Science. https://gc-bs.org/p/24-impactcognitive-load-7b49/

Roberts, H., et al. (2023). *Global AI governance: Divergent regulatory models. In Global AI governance*. Springer. https://link.springer.com/chapter/10.1007/978-3-031-41566-1_4

Romeo, G., & Conti, D. (2025). *Exploring automation bias in human–AI collaboration: A review and implications for explainable AI*. AI & Society. https://link.springer.com/article/10.1007/s00146-025-02422-7

Saaradeey, S., Ghosh, D., Ray, R., Ganesan, S., & Rajagopalan, R. (2019). *Disrupting status quo in AML compliance*. Oracle. https://www.oracle.com/a/ocom/docs/industries/financial-services/fs-disruptingstatus-quo-aml-complaince-wp.pdf

Sachoulidou, A. (2024). *AI systems and criminal liability: A call for action*. Oslo Law Review, 11(1), 1–10. https://doi.org/10.18261/olr.11.1.3

Schmitt, M., & Flechais, I. (2023). *Digital deception: Generative artificial intelligence in social engineering and phishing*. arXiv. https://arxiv.org/pdf/2310.13715

Schneider, J., & Smalley, I. (2025). *What is AML transaction monitoring?* IBM. https://www.ibm.com/think/topics/aml-transaction-monitoring

Shapiro, D. (2025). *How AI is transforming financial crime detection in 2025: From customer due diligence to transaction monitoring*. ThetaRay. https://thetaray.com/how-ai-is-transforming-financial-crime-detection-in-2025-fromcustomer-due-diligence-to-transaction-monitoring/

Sloan, M. (2026). *Agentic AI, explained*. MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/agentic-ai-explained

Sparrow, R. (2016). *Robots and respect: Assessing the case against autonomous weapon systems*. Ethics & International Affairs, 30(1), 93–116. https://doi.org/10.1017/S0892679415000647

Sprenkamp, K., Delgado Fernández, J., Eckhardt, S., & Zavolokina, L. (2024). *Overcoming intergovernmental data sharing challenges with federated learning*. Data & Policy, 6. https://doi.org/10.1017/dap.2024.19

Stockholm International Peace Research Institute. (2017). *Mapping the development of autonomy in weapon systems*. https://www.sipri.org/sites/default/files/2017-11/siprireport_mapping_the_development_of_autonomy_in_weapon_systems_1117 _1.pdf

The Wall Street Journal. (2019). *Fraudsters used AI to mimic CEO voice in unusual cybercrime case*. https://www.wsj.com/articles/fraudsters-use-ai-to-mimicceos-voice-in-unusual-cybercrime-case-11567157402

Thomson Reuters. (2023). *Special report: Suspicious activity reports*. https://www.thomsonreuters.com/en-us/posts/investigation-fraud-and-risk/specialreport-suspicious-activity-reports/

Uddin, N. (2025). *Role of AI in preventing financial crime: A comprehensive analytical review*. https://www.sciencedirect.com/science/article/pii/S2949791425000764

UK National Crime Agency. (2024). *SARs annual report*. https://www.nationalcrimeagency.gov.uk/who-we-are/publications/747-sars-annualreport-2024/file

United Nations Office on Drugs and Crime. (n.d.). Money laundering overview. https://www.unodc.org/unodc/en/money-laundering/overview.html

United Nations Office on Drugs and Crime. (2025). *Emerging threats: The intersection of criminal and technological innovation in the use of automation and artificial intelligence in the cybercrime landscape of Southeast Asia* (Cybercrime Technical Brief Series 2).

U.S. Department of State, Bureau of Arms Control and Nonproliferation. (2023). *Political declaration on responsible military use of artificial intelligence and autonomy*. https://www.state.gov/political-declaration-on-responsible-military-use-ofartificial-intelligence-and-autonomy-2/

Verhagen, A., Luget, A., Conjeaud, O., & Stergiou, V. (2025, August 7). *How agentic AI can change the way banks fight financial crime*. McKinsey & Company. https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/how-agenticai-can-change-the-way-banks-fight-financial-crime

Vuletic, I., & Duic, D. (2025). *Adapting to new realities: Financial crimes and emerging AI technology in global and EU perspective law*. Balkan Social Science Review, 25, 67–91. https://doi.org/10.46763/BSSR25252567v

Wall, L. D. (2018). *Some financial regulatory implications of artificial intelligence*. https://www.sciencedirect.com/science/article/abs/pii/S0148619517302618

Weber, M., et al. (2018). *Scalable graph learning for anti-money laundering: A first look*.

Williams-Ceci, S., Jakesch, M., Bhat, A., Kadoma, K., Zalmanson, L., & Naaman, M. (2026). *Biased AI writing assistants shift users' attitudes on societal issues*. Science Advances, 12, Article eadw5578. https://pmc.ncbi.nlm.nih.gov/articles/PMC12978215/

World Health Organization. (n.d.). *Infodemic*. https://www.who.int/healthtopics/infodemic#tab=tab_1

Yang, H., Shukur, Z., & Sahran, S. (2026). *A review of artificial intelligence for financial fraud detection*. Applied Sciences, 16(4), Article 1931. https://doi.org/10.3390/app16041931

Yang, X., et al. (2023). *Privacy-preserving financial anomaly detection via federated learning*. arXiv. https://arxiv.org/abs/2310.04546

Yu, J., Yu, Y., Wang, X., Lin, Y., Yang, M., Qiao, Y., & Wang, F.-Y. (2024). *The shadow of fraud: The emerging danger of AI-powered social engineering and its possible cure*. arXiv. https://arxiv.org/abs/2407.15912

Zhang, Y., Zhao, C., Xu, Y., & Nie, C. (2025). *A review of artificial intelligence for financial fraud detection*. Applied Sciences. https://www.mdpi.com/2076-3417/16/4/1931