Research Papers

May 2026

Algorithms of Evasion: The Rise of AI-Enabled Proliferation Financing

Aaron Arnold

RUSI

Disclaimer

The content in this publication is provided for general information only. It is not intended to amount to advice on which you should rely. You must obtain professional or specialist advice before taking, or refraining from, any action based on the content in this publication.

The views expressed in this publication are those of the authors, and do not necessarily reflect the views of RUSI or any other institution.

To the fullest extent permitted by law, RUSI shall not be liable for any loss or damage of any nature whether foreseeable or unforeseeable (including, without limitation, in defamation) arising from or in connection with the reproduction, reliance on or use of the publication or any of the information contained in the publication by you or any third party. References to RUSI include its directors, trustees and employees.

RUSI Research Papers, May 2026

ISSN 2977-960X

Publications Team

Editorial

Director of Publications: Alice Trouncer

Managing Editor: Sarah Hudson

Assistant Editor: Sophie Boulter

Design

Graphic Designer: Lisa Westthorp

Research Editorial

Head of Research Governance and Editorial: Elias Forneris

Cover image: Wanan / Adobe Stock

Get in touch

www.rusi.org

enquiries@rusi.org

+44 (0)207 747 2600

The Royal United Services Institute for Defence and Security

61 Whitehall, London

SW1A 2ET

United Kingdom

Follow us on

Contents

Methodology

Research for this paper was conducted from June 2025 to November 2025, with funding from the US Department of State. The paper was submitted on 12 February 2026 and approved for copy-editing on 13 February 2026.

The research in this paper is primarily based on semi-structured interviews with experts in the AI field, informal discussions with selected stakeholders (such as those in intergovernmental organisations), data collection from official reports (for example, the UN Panel of Expert Reports), and a review of literature from academic, non-governmental and private sector organisations (for example, consulting firms' trend reports).

This paper makes several assessments and predictions based on current use cases within the AI field. These assessments and predictions assume an upward trajectory in terms of AI adoption and usage.

While we strive for accuracy and quality, please note that the information provided may not be entirely error-free or up to date. RUSI does not assume any responsibility or liability for the use or interpretation of this content.

Key Terms

Adversarial AI: The use of AI techniques that attempt to exploit social, political, economic or IT systems for illicit purposes.

Agentic AI network: A system comprising multiple AI agents that have the capability to act autonomously.

Application Programming Interface (API): A set of programming functions that allow access to remote services, such as querying or updating databases.

AI: The simulation of human intelligence by computers, which involves pattern recognition and complex reasoning.

Decentralised finance (DeFi): A financial system that operates without central intermediaries like banks or brokers, using blockchain technology and smart contracts to enable peer-to-peer transactions and services.

Deepfake: Synthetic media in which a person's likeness is replicated using AI to produce a photo, video or audio.

Dual-use goods: Items, software and technology that can be used for both civilian and military applications.

Financial Action Task Force (FATF): An intergovernmental organisation, founded in 1989, responsible for developing policies to combat global money laundering. In 2001, its mandate expanded to include terrorist financing, and in 2008, it included proliferation financing.

Generative AI: AI that is capable of creating new content, such as images, text, video or audio.

Large Language Model (LLM): A type of AI model that can recognise and generate text, among other tasks. LLMs are trained using large corpuses of text.

Machine learning (ML): A type of AI that can apply previously learned patterns to new data for tasks such as object or pattern recognition.

Malware (malicious software): Software intentionally designed to cause damage to a computer, server, client or computer network.

Model Context Protocol (MCP): An open source standard for AI models to connect with external data, tools and systems.

Panel of Experts: A group of specialists or authorities in a particular field convened to advise on, investigate or report on a specific issue. In the context of international sanctions, a UN Panel of Experts often monitors the implementation of sanctions regimes.

Phishing: A type of cyber attack that specifically relies on social engineering to gain access to the victims' data.

Proliferation financing (PF): The use of funds or financial services to acquire, develop, or otherwise deal in WMD or related materials (such as in dual-use goods or technologies).

Sanctions: Penalties imposed by one country on another, or by international bodies, often to achieve specific foreign policy or national security objectives. These can include restrictions on trade, financial transactions, travel or other activities.

Synthetic identity: A fabricated identity that combines real and fictitious information, often created using a mix of stolen personal data and made-up details, such as an address or phone number. Commonly used for fraudulent activities, such as opening fraudulent accounts, obtaining loans or engaging in other financial crimes.

Virtual asset: A digital representation of value that can be digitally traded or transferred and used for payment or investment purposes. This includes cryptocurrencies but can also encompass other digital assets, like non-fungible tokens (NFTs), that are blockchain enabled. In this context, virtual assets do not include digital payment processors, such as PayPal or Venmo.

Virtual asset service provider (VASP): Any natural or legal person that conducts, for or on behalf of another, activities such as exchanging, transferring, safekeeping or managing virtual assets like cryptocurrencies. Defined by the FATF, these entities – including crypto exchanges and wallet providers – must comply with Anti-Money Laundering (AML) and Counter-Terrorist Financing (CTF) regulations.

Acknowledgements

The author would like to thank the sponsor for their support of this paper. Also, a special thanks to Togzhan Kassanova and Bryan Early for their review and helpful comments. Last but not least, thank you Wojciech Pawlus (maestro of PF) for all your efforts.

Executive Summary

While AI promises to usher in an era of significant economic revitalisation and technological innovation across nearly all sectors, the potential for malicious use cases is evident and growing. Already, AI is reshaping the future of proliferation financing (PF) and sanctions evasion, presenting novel challenges to global enforcement and implementation. States like North Korea and Iran are now developing and deploying AI models to aid with sanctions evasion activities. The current trajectory suggests that the use of AI for malicious purposes is quickly evolving from AI-assisted models to AI-enabled ones: in other words, using AI to supplement sanctions evasion activities versus using it to fully direct them.

AI has the potential to radically increase the scale of PF activities, like sanctions evasion, to levels that overwhelm current PF and sanctions evasion detection and enforcement capabilities. This paper focuses on how proliferators like North Korea currently deploy AI models, how these activities are evolving in scale and sophistication, the legal and regulatory challenges to mitigating risks from malicious AI, and what to do about them. Furthermore, the paper highlights the risks that the current asymmetry between enforcement and criminal activity poses, and how a new frontier in regulatory arbitrage around AI may be emerging. In this new regulatory frontier, sophisticated actors leverage gaps in the patchwork quilt of global rules to ensure effective sanctions evasion.

The paper makes the following key findings:

- Generative AI (GenAI) can mass-produce high-quality fraudulent documents, including driver's licenses, bank statements and vessel registrations, at scale and with contextual accuracy. The use of GenAI threatens to overwhelm current detection processes, which are largely manual.
- AI can automate the administrative minutia of managing extensive shell company networks. It can architect multi-layered, cross-jurisdictional ownership chains involving synthetic identities that are mathematically optimised to hide beneficial owners.
- AI-powered systems can analyse blockchain patterns in real time to dynamically adjust cryptocurrency mixing strategies, effectively evading detection tools.

- Proliferators are beginning to use adversarial AI to probe for weaknesses in defensive screening systems, systematically mapping thresholds to engineer successful evasion. Emerging agentic AI networks (autonomous agents), for example, can manage complex evasion schemes without direct human intervention, rendering traditional disruption strategies that target human nodes increasingly obsolete.

Given these current and emerging challenges, the paper makes the following key recommendations:

Key Recommendations to Policymakers

- National authorities should explore safe harbour provisions for black-box, AI-driven counterproliferation financing (CPF) models (namely, where the AI decision-making logic is not known) and establish a National Security Data Enclave to pool pseudonymised transaction data. Regulators should also establish computational know-your-client checks (compute-KYC) and cloud liability standards for infrastructure-as-a-service providers, and pilot a mandatory adversarial red-teaming certification.
- International organisations and institutions should develop model legislative frameworks for the malicious use of AI and know-your-API (KYA) standards for AI agents accessing financial systems. The FATF should update Recommendations 1 and 15 to explicitly recommend that jurisdictions assess their vulnerability to autonomous AI agents.
- Private sector institutions, including financial institutions, should upgrade CPF KYC procedures to account for deepfake capabilities, deploy defensive AI to audit trade documents for semantic inconsistencies, and adopt dynamic compliance modelling using behaviour-based analytics.

Introduction

Tools such as generative AI (GenAI), which can produce sophisticated fraudulent identification documents, for example, have helped North Korea perpetrate phishing attacks against Western companies. Over the next three to five years, governments and private sector institutions will need to rapidly adapt identification and mitigation protocols as adversaries move from AI-assisted to AI-enabled sanctions evasion and PF. AI-assisted sanctions evasion involves AI models that aid discrete sanctions evasion activities, whereas AI-enabled sanctions evasion involves AI models that act autonomously over entire sanctions evasion schemes. Furthermore, international institutions, such as the Financial Action Task Force (FATF), should continue to seek out and adopt recommendations to contend with AI and its expected evolution and impact over the subsequent five to 10 years.

According to a recent UN report, 'the global AI market will soar from $189 billion in 2023 to $4.8 trillion by 2033'.[1] As governments, industry and consumers integrate AI more into social, economic and political systems, the costs related to cybersecurity are also expected to increase. One research firm predicts that corporate spending on cybersecurity will hit nearly $240 billion by the end of 2026.[2] Ultimately, this systematic integration of AI will lower the barriers of entry for a range of criminal activity, including sanctions evasion and PF.

Importantly, AI is not necessarily changing PF and sanctions evasion typologies but instead increasing their efficiency and effectiveness. In other words, while actors will still require the use of third-party intermediaries, secrecy jurisdictions and opaque financial structures to obfuscate their PF activities, AI – especially in the short term – will significantly increase the speed and scale at which such actors can operate. For instance, AI can quickly analyse datasets to identify weaknesses in sanctions frameworks, predict enforcement actions and even generate circuitous routes for

1. UN Trade and Development (UNCTAD), 'AI Market Projected to Hit $4.8 Trillion by 2033, Emerging as Dominant Frontier Technology', 7 April 2025, <https://unctad.org/news/ai-market-projected-hit-48-trillion-2033-emerging-dominant-frontier-technology>, accessed 1 November 2025.
2. *The Economist*, 'How AI-Powered Hackers Are Stealing Billions', 19 August 2025.

obscuring financial flows.[3] This could include the automated creation of shell companies, the sophisticated manipulation of trade data, or the use of AI-driven tools to obscure the origin and destination of funds.

Furthermore, AI's ability to learn and adapt autonomously means that evasion strategies can become increasingly dynamic and resilient to countermeasures. As financial institutions and law enforcement agencies deploy their own AI-powered detection systems, illicit actors can use AI to develop countermeasures that evolve in real time, engaging in an AI-driven 'arms race' of detection and evasion. This constant adaptation will make it significantly harder for authorities to keep pace, potentially rendering traditional investigative methods obsolete or severely hampering them.

This paper begins with an overview of historical PF typologies and methodologies, and explains how North Korea and Iran currently employ AI to assist with sanctions evasion and PF activities. The second chapter then develops a high-level topology of AI-enabled activities that could significantly undermine current PF implementation practices. The third chapter looks ahead to next-generation AI threats that pose new challenges to states and industry when it comes to implementing sanctions obligations and preventing PF. The fourth chapter provides an overview of current technical and governance challenges, as well as recommendations for both public and private sectors to address and mitigate the emerging threats associated with AI.

3. Companies already employ AI in a range of business tasks, from optimising shipping routes and supply chains, to ensuring legal and regulatory compliance. The same AI tools can be adapted for illicit purposes. For an example of the use of AI in corporate legal and regulatory compliance, see Jesus M Olivera, 'Enhancing Regulatory Compliance in the AI Age by Grounding Documents with Generative AI', 18 November 2025, <https://www.ibm.com/think/insights/enhancing-regulatory-compliance-ai-age>, accessed 30 March 2026.

Background: A Primer to North Korea and Iran's Proliferation Financing Activities

Economic statecraft, like sanctions and export controls, has emerged as the preferred policy instrument to address a wide swathe of international security concerns. With varying degrees of success, countries like the US have enacted sanctions against WMD proliferators to deny access to international finance, trade and commerce, and certain technologies.

International efforts to curb North Korea and Iran's nuclear and ballistic missile ambitions include subjecting both countries to sweeping UN sanctions, which have ranged from asset freezes and travel restrictions to embargoes. North Korea, for example, has been under UN sanctions since October 2006, when the country conducted its first nuclear test.[4] In late August 2025, the E3 (France, Germany and the UK) formally initiated the 'snapback' mechanism of the Joint Comprehensive Plan of Action (JCPOA) – often referred to as the Iran Nuclear Deal – signalling their intent to reinstate UN sanctions against Iran, which were lifted in 2015.[5]

Despite both international and unilateral sanctions, both countries continue to exploit global financial and commercial systems to both generate revenue and acquire restricted goods and technologies to support respective WMD and ballistic missile programmes. The following sub-section outlines how both state and non-state actors

4. UN Security Council Resolution 1718, October 2006, S/RES/1718.
5. According to the E3, Iran 'exceeded JCPOA limits on enriched uranium, heavy water, and centrifuges, restricted the International Atomic Energy Agency's ability to conduct JCPOA verification and monitoring activities, and has abandoned the implementation and the ratification process of the Additional Protocol to its Comprehensive Safeguards Agreement'. See Foreign, Commonwealth & Development Office, 'E3 Joint Statement on Iran: Initiation of the Snapback Process', 28 August 2025, <https://www.gov.uk/government/news/e3-joint-statement-on-iran-initiation-of-the-snapback-process>, accessed 1 November 2025.

exploit global financial and commercial systems to evade sanctions and export controls in more detail, focusing on how they have evolved.

Common Proliferation Financing Typologies

'Proliferation financing', while lacking a widely agreed definition, broadly 'refers to the use of funds or financial services to acquire, develop or otherwise deal in WMD or related materials'.[6] Two key typologies describe the totality of PF: revenue-generating activities and the acquisition of proliferation-sensitive goods and technologies. In each case, both state and non-state actors deploy a range of methods to gain access to global financial and commercial systems while obfuscating their identity and/or end users. From 2009 to 2024, for example, the UN Panel of Experts on North Korea – an eight-member body responsible for investigating and reporting on North Korea's sanctions violations – documented hundreds of cases whereby North Korea successfully evaded international sanctions to either generate illicit revenue or acquire proliferation-sensitive goods and technologies to support its WMD programme.[7][8] These activities constitute PF.

At its core, PF is the business of hiding, and thus requires some level of obfuscation, whether it is hiding the movement of goods, beneficial ownership information, transactions or true end-users. In its June 2025 report, 'Complex Proliferation Financing and Sanctions Evasion Schemes', FATF outlines several common methods that sanctions evaders tend to share. These include the use of intermediary parties (such as front and shell companies), circuitous payment routing through third-party countries, obfuscation of beneficial ownership information, and the use of virtual currencies and other technologies.[9] It is important to note, however, that these obfuscation typologies are not unique to PF and are frequently employed in other types of financial crime, including money laundering, terrorist financing and tax avoidance, among others.

6. Aaron Arnold and Daniel Salisbury, 'Guide to Conducting a National Proliferation Financing Risk Assessment', RUSI, 21 November 2024, <https://www.rusi.org/explore-our-research/publications/special-resources/guide-conducting-national-proliferation-financing-risk-assessment-2024>, accessed 1 November 2025.
7. In March 2024, Russia vetoed a resolution to extend the UN Panel of Experts' mandate, effectively abolishing the only monitoring mechanism. See UN, 'World News in Brief: Russia Vetoes DPR Korea Sanctions Resolution, Children Under Fire in Sudan, Drought Plagues Malawi', 28 March 2024, <https://news.un.org/en/story/2024/03/1148121>, accessed 28 October 2025.
8. For the most recent Panel of Experts report, see UN Security Council, 'Final Report Pursuant to UNSCR 2680 (2023)', March 2024, <https://documents.un.org/doc/undoc/gen/n24/032/68/pdf/n2403268.pdf>, accessed 28 October 2025.
9. Financial Action Task Force (FATF), 'Complex Proliferation Financing and Sanctions Evasion Schemes', June 2025, <https://www.fatf-gafi.org/content/dam/fatf-gafi/reports/Complex-PF-Sanctions-Evasions-Schemes.pdf>, accessed 1 November 2025.

One notable characteristic of PF and sanctions evasion is the ability of state and non-state actors to adapt to new challenges, rules and regulations, and technologies. In the case of North Korea, for example, the Panel of Experts reports paint a striking portrait of a country fully capable of adapting to external changes by adopting new technologies and methods. As early as 2016, North Korea began using virtual currencies to launder proceeds from ransomware attacks. According to a 2019 Panel of Experts report, the country received payment in Bitcoin from the global WannaCry ransomware attacks, which affected more than 200,000 computers in 150 countries.[10] North Korea later laundered the Bitcoin through a series of virtual currency exchanges and different virtual assets to further obfuscate its origin. The country has since stolen billions of dollars' worth of virtual currencies – mostly from hacking virtual asset service providers.[11]

Similarly, starting in 2018, Iran also turned to virtual currencies to evade US sanctions and generate revenue. Unlike North Korea, however, Iran's exploits have focused less on the theft and laundering of virtual currencies than on the use of virtual currencies to facilitate trade outside of traditional financial channels.[12] In September 2025, the US Department of Justice sought a civil forfeiture action against a virtual currency wallet that an Iranian company, which manufactures navigation modules for Iran's military drone programme, used to avoid transactional scrutiny.[13] In this particular case, the Iranian company used Tether – a cryptocurrency pegged to the value of the US dollar – to circumvent traditional payment processors.

Ultimately, both countries' adoption of cryptocurrency to directly or indirectly support their WMD ambitions demonstrates a keen ability to employ novel methods, driven by the need to maintain access to global financial and commercial systems. Equally important is that it shows how the adoption and use of new technologies enable these states to outpace enforcement efforts. While North Korea has used cryptocurrency since 2016, for example, significant enforcement efforts (and global awareness) took several years to catch up.

10. UN Security Council, 'Midterm Report of the Panel of Experts Submitted Pursuant to Resolution 2464 (2019)', 30 August 2019, <https://documents.un.org/doc/undoc/gen/n19/243/04/pdf/n1924304.pdf>, accessed 28 October 2025. (The author of this report was a member of the Panel of Experts from 2019 to 2021.)
11. Multilateral Sanctions Monitoring Team, 'The DPRK's Violation and Evasion of UN Sanctions through Cyber and Information Technology Worker Activities', <https://msmt.info/Publications/detail/MSMT%20Report/4221>, accessed 30 October 2025.
12. Angus Berwick and Tom Wilson, 'Crypto Exchange Binance Helped Iranian Firms Trade $8 Billion despite Sanctions', *Reuters*, 4 November 2022; US Department of the Treasury, 'Treasury Designates Iran-Based Financial Facilitators of Malicious Cyber Activity and for the First Time Identifies Associated Digital Currency Addresses', press release, 8 February 2025, <https://home.treasury.gov/news/press-releases/sm556>, accessed 3 March 2026.
13. US Department of Justice, 'United States Seeks Civil Forfeiture of Cryptocurrency Associated with Iranian National Mohammad Abedini', press release, 11 September 2025, <https://www.justice.gov/usao-ma/pr/united-states-seeks-civil-forfeiture-cryptocurrency-associated-iranian-national-mohammad>, accessed 28 October 2025.

North Korea and Iran, as well as other proliferating states like Russia, China and Pakistan, will continue to require access to global financing, commerce and shipping to generate revenue and acquire proliferation-sensitive goods and technologies. Over the next five years, AI will become an increasingly prominent feature in sanctions evasion and PF activities.

The next chapter outlines how countries, primarily North Korea and Iran, currently employ AI technology to evade sanctions. It also describes over-the-horizon AI use cases (namely, emerging scenarios that are theoretically feasibly but not yet implemented).

An Emerging Typology of AI-Enabled Proliferation Financing

A recent assessment of serious and organised crime by Europol suggests that AI is fundamentally reducing barriers to crime, noting that 'the scale of online fraud, driven by advancements in automation and AI, has reached an unprecedented magnitude and is projected to continue growing'.[14] A business consultancy predicts that generative AI 'could enable fraud losses to reach $40 billion in the United States by 2027, from $12.3 billion in 2023, a compound annual growth rate of 32%'.[15]

Similarly, a 2024 report by the US Department of Homeland Security predicts that AI may 'transform some crimes so thoroughly that society will need to contend with them as new crimes'.[16] Much of this is driven by increased accessibility of GenAI models and tools – that is, tools that have the capability to generate new content (such as text, images and even programmatic code) based on the data they were trained on.

Currently, criminal use of AI is most prevalent in cyber-related crimes and fraud (for example, malware development, phishing attacks and hacking). In general, criminal networks have found AI to be especially useful in automating discrete tasks associated with perpetrating fraud, such as generating phishing emails and replies. In early 2024, for example, a UK engineering firm in Hong Kong fell victim to a phishing scheme whereby criminals used an AI-generated video call to dupe an employee into sending

14. Europol, *European Union Serious and Organised Crime Threat Assessment: The Changing DNA of Serious and Organised Crime* (Luxembourg: Publications Office of the EU, 2025), p. 42, <https://doi.org/10.2813/0758057>, accessed 3 November 2025.
15. Satish Lalchand et al., 'Generative AI Is Expected to Magnify the Risk of Deepfakes and Other Fraud in Banking', Deloitte Insights, May 2024, <https://www.deloitte.com/us/en/insights/industry/financial-services/deepfake-banking-fraud-risk-on-the-rise.html>, accessed 28 October 2025.
16. Department of Homeland Security, 'Impact of Artificial Intelligence on Criminal and Illicit Activities', 2024, p. 16, <https://www.dhs.gov/sites/default/files/2024-10/24_0927_ia_aep-impact-ai-on-criminal-and-illicit-activities.pdf>, accessed 28 October 2025.

nearly $27 million to the criminals. The Arup deepfake video purportedly realistically portrayed the UK company's senior officers.[17]

Not only have cyber criminals deployed AI in novel ways to propagate malware attacks, but they have increasingly resorted to crime-as-a-service models, many of which are based around the use and integration of AI tools. XanthoroxAI, for example, is a large language model (LLM) available through common online platforms such as GitHub and Discord, which allows its users to generate deepfakes, phishing emails and ransomware – all available for $200.[18]

Other sectors are also seeing significant growth in AI-enabled crime. Specifically, finance, real estate and healthcare are all seeing large increases in the use of identity and document fraud driven by GenAI.[19] In early 2024, cybersecurity researchers identified a highly sophisticated suite of malware (specifically GoldPickaxe) targeting banking users in Southeast Asia. This was a direct evolution of traditional phishing, using AI to bridge the gap between stealing credentials and accessing the accounts.[20]

Broadly speaking, the use of AI in PF can be defined by two key categories: AI-assisted and AI-enabled. Whereas AI-assisted sanctions evasion merely augments traditional circumvention methods, AI-enabled sanctions evasion involves the strategic deployment of AI technologies to drive evasion through automation and scalability. This is not merely an incremental evolution, but a fundamental change which threatens to overwhelm current sanctions implementation practices.

The following sections outline a broad taxonomy of how AI is currently shaping PF and sanctions evasion, focusing specifically on the consequences of scaling deception through GenAI, automating obfuscation and enhancing anonymity through cryptocurrencies.

GenAI: The AI Forgery Foundry

Advances in GenAI – the ability to create content, whether text, audio or images – fundamentally challenges the ability to detect sanctions evasion and PF activities. Already, GenAI is playing a significant role in global fraud. A recent alert by the US Financial Crimes Enforcement Network (FinCEN) notes that starting in 2023 and continuing in 2024, 'FinCEN has observed an increase in suspicious activity reporting by financial institutions describing the suspected use of deepfake media in fraud schemes

17. Dan Milmo, 'UK Engineering Firm Arup Falls Victim to £20m Deepfake Scam', *The Guardian*, 17 May 2024.
18. Cristos Velasco, 'XanthoroxAI and the Crime-as a Service Model', LinkedIn, 27 May 2025, <https://www.linkedin.com/pulse/xanthoroxai-crime-as-service-model-cristos-velasco-vxkbc/>, accessed 28 October 2025.
19. Department of Homeland Security, 'Impact of Artificial Intelligence on Criminal and Illicit Activities', p. 22.
20. Andrey Polovinkin and Sharmine Low, 'Face Off: Group-IB Identifies First iOS Trojan Stealing Facial Recognition Data', Group-IB blog, 15 February 2024, <https://www.group-ib.com/blog/goldfactory-ios-trojan/>, accessed 28 October 2025.

… These schemes often involve criminals altering or creating fraudulent identity documents to circumvent identity verification and authentication methods'.[21]

The use of fraudulent documentation is common in sanctions evasion schemes, regardless of actor. Such methods are not novel; states have used fraudulent documentation to facilitate sanctions evasion and export control violations for decades. During the 1970s, for example, the then Soviet Union often falsified shipping documents and used third-party intermediaries to skirt nascent export controls systems, acquiring Western dual-use and military goods and technologies.[22]

North Korea has long used fraudulent documents to perpetrate its sanctions evasion schemes. The UN Panel of Experts reports routinely highlighted the country's illicit shipping activities (for example, the importation of oil in excess of UN caps and illicit exportation of coal) that relied, in part, on fraudulent documentation. This documentation has included, for example, fake vessel registration documents, which helped the country conceal the true owner of the vessels involved in smuggling.[23] North Korean overseas workers also use fraudulent documents to hide their nationality and obtain access to banking overseas. In one case, North Korean national Kim Chol Sok was found operating several hotels, restaurants and casinos in Cambodia, using fake passports and national identity documents.[24] The revenue generated by these establishments directly supported North Korea's nuclear weapons and ballistic missile programmes.

The widespread availability and increasing sophistication of GenAI and deepfake technology creates new avenues for sanctions evasion and PF networks. First, GenAI excels at creating a wide array of convincing fraudulent documents. This includes fake identification documents like driver's licenses and professional credentials, as well as financial documents such as bank statements, pay stubs and tax returns.[25] These AI-generated documents can then be used to open bank accounts, apply for credit, register front companies, or act as identity documents for directors and shareholders in complex ownership structures, designed to obscure the involvement of sanctioned entities.

21. Financial Crimes Enforcement Network, 'FinCEN Alert on Fraud Schemes Involving Deepfake Media Targeting Financial Institutions', FIN-2024-Alert004, 13 November 2024, <https://www.fincen.gov/system/files/shared/FinCEN-Alert-DeepFakes-Alert508FINAL.pdf>, accessed 28 October 2025.
22. Aaron Arnold and Daniel Salisbury, 'The Sanctions-Busting Architects: Moscow's Preparations for the West's Sanctions', Lawfare, 4 March 2024, <https://www.lawfaremedia.org/article/the-sanctions-busting-architects-moscow-s-preparations-for-the-west-s-sanctions>, accessed 28 October 2025.
23. UN, 'Final Report of the Panel of Experts Submitted Pursuant to Resolution 2515 (2020)', S/2021/211, 4 March 2021, p. 18, <https://docs.un.org/en/S/2021/211>, accessed 28 October 2025.
24. UN, 'Final Report of the Panel of Experts Submitted Pursuant to Resolution 2569 (2021)', S/2022/132, 1 March 2022, p. 76, <https://docs.un.org/en/S/2022/132>, accessed 28 October 2025.
25. Financial Industry Regulatory Authority, 'Protecting Your Investment Accounts From GenAI Fraud', 15 January 2025, <https://www.finra.org/investors/insights/gen-ai-fraud-new-accounts-and-takeovers>, accessed 28 October 2025.

GenAI is also useful for creating synthetic identities, or meticulously crafted online personas with plausible backstories, social media presences and even fabricated financial histories. Examples of synthetic identities created with GenAI include records of synthetic parents to bolster the legitimacy of manufactured online personas.[26] According to a report by the US Federal Reserve Bank, a potential benefit of using GenAI is that the systems used to generate synthetic identities 'can learn from its mistakes and churn out more of what works'.[27] Importantly, this can be done at scale, in a way which is contextually accurate, grammatically correct and consistent with genuine documentation.[28]

GenAI is particularly problematic for sanctions regimes related to specific goods, where falsified documents can effectively obscure the true origin, destination or nature of the items being traded, directly contravening sanctions. Falsifying the nature, origin or destination of goods is a common tactic in sanctions evasion and PF. Much of Iran's illicit procurement operations, for example, employed the use of third-party jurisdictions to transship export-controlled or dual-use goods and technologies. During the transshipment, goods were often intentionally mislabelled or misidentified to avoid scrutiny by export authorities.[29] This might entail, for example, the use of falsified end-user certificates, as well as other fraudulent trade documents, including invoices, bills of lading (shipping documents), certificates of origin and customs declarations.

The ability to mass-produce high-quality fraudulent documents fundamentally challenges the reliance on document-based verification processes, particularly in international trade and finance, where physical inspection of goods or direct verification of document issuers is often impractical or impossible. Many compliance frameworks, especially within trade finance, depend heavily on the presumed authenticity of supporting documentation. Furthermore, enforcement investigations related to sanctions evasion and PF often centre around document verification and identifying inconsistencies. The use of GenAI could easily overwhelm traditional investigative methods.

26. Mike Timoney, 'Gen AI Is Ramping up the Threat of Synthetic Identity Fraud', Federal Reserve Bank of Boston, 17 April 2025, <https://www.bostonfed.org/news-and-events/news/2025/04/synthetic-identity-fraud-financial-fraud-expanding-because-of-generative-artificial-intelligence.aspx>, accessed 28 October 2025.
27. *Ibid*.
28. Sachin Dixit, 'Generative AI-Powered Document Processing at Scale with Fraud Detection for Large Financial Organizations', *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* (Vol. 10, October 2024), pp. 1038–65.
29. See, for example, the case studies of Iran in Jonathan Brewer, 'Study of Typologies of Financing of WMD Proliferation', King's College London, Project Alpha, 12 October 2017, pp. 79–153, <https://www.kcl.ac.uk/csss/assets/study-of-typologies-of-financing-of-wmd-proliferation-2017.pdf>, accessed 28 October 2025.

The Evolution of North Korea's IT Labourers

In 2017, the UN Security Council banned North Korea's overseas labourers, requiring all member states to repatriate North Korean workers by December 2019.[30] Historically, North Korea has sent its nationals overseas to generate significant sources of revenue for the regime. Such activities have included construction, restaurants and Eastern medical clinics, among others. After the country effectively shuttered its borders due to the pandemic, its traditional revenue streams began to run dry. To compensate, North Korea moved to a 'work from home' model, deploying thousands of IT labourers to work remotely on IT projects for companies worldwide.[31] According to 2022 guidance from the US government, these individuals – mostly located throughout China, Russia and parts of Southeast Asia – could generate up to $300,000 per year in revenue.[32]

The first wave of North Korea's IT labourers used relatively common tactics to hide their identity and location. These include, for example, the use of network services like VPNs (virtual private networks) to hide their location, use of fraudulent documentation, and the recruitment of both witting and unwitting actors to facilitate their activities.[33] In a recent case, a US citizen was sentenced to prison for helping North Korea establish an online presence within the US, which helped the country's IT labourers obscure their identity and location – ultimately resulting in the IT workers securing jobs at 309 companies across the US.[34]

30. UN Security Council Resolution 2397, S/RES/2397, 22 December 2017, <https://main.un.org/securitycouncil/en/s/res/2397-%282017%29>, accessed 28 October 2025.
31. Aaron Arnold and Daniel Salisbury, 'Remote Sanctions-Busting: A Post-COVID New Normal?', *Washington Quarterly* (Vol. 47, No. 4, October 2024), pp. 63–77.
32. US Department of State, 'Guidance on the Democratic People's Republic of Korea Information Technology Workers', 16 May 2022), <https://ofac.treasury.gov/media/923126/download?inline>, accessed 28 October 2025; Multilateral Sanctions Monitoring Team, 'The DPRK's Violation and Evasion of UN Sanctions through Cyber and Information Technology Worker Activities'.
33. For an overview of North Korea's IT labour network methods, see Chandana Seshadri, 'How DPRK IT Workers Exploit Identity Management Vulnerabilities', *RUSI Journal* (Vol. 170, No. 4, June 2025), pp. 74–84.
34. US Department of Justice, 'Arizona Woman Sentenced for $17M Information Technology Worker Fraud Scheme That Generated Revenue for North Korea', press release, 24 July 2025, <https://www.justice.gov/opa/pr/arizona-woman-sentenced-17m-information-technology-worker-fraud-scheme-generated-revenue>, accessed 1 November 2025.

As governments began to raise awareness of the risks of hiring North Korean IT workers, however, North Korea began adopting new, AI-enabled methods to evade detection. This has included the use of GenAI to create online personas, cover letters and resumes for job applications, as well as live deepfakes to obscure their true identity during online job interviews.[35]

Shell Companies and Beneficial Ownership

One of the more significant limitations of current sanctions evasion practices are human resources; in other words, the individuals behind evasion networks have a limited capacity to manage the administrative minutiae of obfuscation. Generating shell companies, for example, still requires paperwork to be filled out, filed and maintained correctly, so errors are a common way of getting caught.[36]

GenAI has the potential to automate most aspects of creating and managing extensive networks of shell companies. AI algorithms can now effectively generate diverse and seemingly unrelated registration details (such as names, addresses and contact information) for numerous shell entities, manage their minimal digital footprints (through basic websites or social media snippets), and even automate rudimentary administrative tasks or small, seemingly legitimate transactions to lend an air of authenticity. Such networks could lie dormant, with AI programmes monitoring for specific triggers or optimal conditions – such as a new sanctions designation or a specific transactional need – before being activated for an evasion scheme.

Identifying beneficial ownership information is a cornerstone of anti-money laundering (AML) and sanctions compliance. Threat actors can use GenAI to architect multi-layered, cross-jurisdictional ownership chains involving an intricate web of trusts, nominee shareholders and AI-generated synthetic identities serving as directors or shareholders.[37] Furthermore, the potential for AI to design ownership structures that are mathematically optimised for obfuscation is a significant escalation of such a threat. As previously mentioned, current approaches to identifying true beneficial information often rely on finding human error in setting up structures, finding

35. Google Threat Intelligence Group, 'Adversarial Misuse of Generative AI', Google Cloud blog, 29 January 2025, <https://cloud.google.com/blog/topics/threat-intelligence/adversarial-misuse-generative-ai>, accessed 28 October 2025.
36. In one of many such instances, UN Panel of Experts investigators found several registration errors related to the vessel, MT Koti. This led the investigators to connect the errors to other sanctions evasion networks that were previously unknown. See UN, 'Final Report of the Panel of Experts Submitted Pursuant to Resolution 2345 (2017)', S/2018/171, 5 March 2018, p. 33, <https://docs.un.org/en/S/2018/171>, accessed 28 October 2025.
37. Andres Knobel, 'When AI Runs a Company, Who Is the Beneficial Owner?', Tax Justice Network blog, 19 May 2025, <https://taxjustice.net/2025/05/19/when-ai-runs-a-company-who-is-the-beneficial-owner/>, accessed 28 October 2025.

common links (such as shared addresses, directors or service providers), or recognising established patterns in how complex ownership arrangements are typically created.

Enhanced Anonymity and Laundering

Advances in AI will ultimately supercharge actors' ability to exploit virtual assets and virtual asset service providers to facilitate sanctions evasion schemes. Over the last six years, virtual assets have played a significant role in PF and sanctions evasion. As previously mentioned, North Korea was one of the first to capitalise on cryptocurrencies' relative anonymity and ease of use in its 2017 WannaCry ransomware attack. In March 2025, hackers linked to North Korea's famed Lazarus Group pulled off the largest-ever heist of cryptocurrency, valued at over $1.5 billion.[38] Iran, too, has become a prolific user of cryptocurrency to evade sanctions. In September 2025, the US Department of the Treasury announced sanctions against a network of Iranian nationals, stating that 'Iranian "shadow banking" networks like these – run by trusted illicit financial facilitators – abuse the international financial system, and evade sanctions by laundering money through overseas front companies and cryptocurrency'.[39]

Cryptocurrency mixers, or tumblers, are services that attempt to break the traceability of blockchain transactions by pooling funds from multiple users and then distributing them in a manner that obscures the link between senders and receivers.[40] AI-powered systems can analyse blockchain transaction patterns in real time and dynamically adjust their mixing strategies – such as varying transaction sizes, timing, pathways and the number of intermediary wallets – to more effectively evade detection by blockchain analysis tools. This creates an adaptive adversary that continuously improves its obfuscation capabilities over time, making the tracing of illicit cryptocurrency flows increasingly challenging for law enforcement and compliance teams.[41]

AI can manage a vast number of cryptocurrency wallets, execute intricate transaction chains across multiple cryptocurrencies and decentralised finance (DeFi) protocols, and vary transaction patterns (including amounts, timing and recipient addresses) to

38. Taylar Rajic and Julia Brock, 'The ByBit Heist and the Future of U.S. Crypto Regulation', Center for Strategic and International Studies, 18 March 2025, <https://www.csis.org/analysis/bybit-heist-and-future-us-crypto-regulation>, accessed 23 April 2026.
39. US Department of the Treasury, 'Treasury Targets Financial Network Supporting Iran's Military', press release, 16 September 2025, <https://home.treasury.gov/news/press-releases/sb0248>, accessed 28 October 2025.
40. US Secret Service, 'Public Advisory Cryptocurrency Mixers', 2025, <https://www.secretservice.gov/sites/default/files/reports/2025-06/Public-Alert-Cryptocurrency-Mixing.pdf>, accessed 23 April 2026.
41. For a description of how AI is scaling the illicit use of crypto currency, see TRM Labs, 'How AI Is Changing the Scale and Speed of Crypto Fraud', TRM Labs blog, 23 February 2026, <https://www.trmlabs.com/resources/blog/how-ai-is-changing-the-scale-and-speed-of-crypto-fraud>, accessed 5 March 2026.

make them appear less suspicious to automated transaction monitoring systems. In such scenarios, vast sums of illicitly obtained cryptocurrency could be rapidly obfuscated through thousands, or even tens of thousands, of micro-transactions executed across a sprawling network of wallets in a very short period. This speed and scale could overwhelm the real-time monitoring and response capabilities of financial institutions and law enforcement agencies, allowing illicit actors to move and obscure funds before effective action can be taken to freeze assets or comprehensively trace the transaction path.

AI could also be used to create and manage 'autonomous liquidity pools' on DeFi platforms, specifically designed for sanctions evasion or money laundering.[42] This means that AI could automate the creation of numerous accounts, funded with illicitly sourced cryptocurrency, using AI-generated synthetic identities to make these accounts appear as legitimate liquidity providers. This would create a decentralised laundering infrastructure that is significantly harder to identify, monitor and shut down than centralised illicit financial services.

42. Fintech Staff Writer, 'When DeFi Protocols Become Self-Evolving Organisms', GlobalFinTechSeries blog, 12 January 2026, <https://globalfintechseries.com/featured/when-defi-protocols-become-self-evolving-organisms/>, accessed 3 March 2026.

Next-Generation Proliferation Financing AI Threats

To date, AI has largely augmented PF by enhancing scalability and existing tactics, for example in fraud-related schemes. As the technology becomes more ubiquitous and advanced, there will probably be more novel applications in PF and sanctions evasion that rely on more autonomous features. The following sections explore these novel applications and horizon threats, such as the use of agentic (or autonomous) AI networks and its implications for CPF.

Adversarial AI: Probing and Exploiting System Vulnerabilities

While financial institutions and other regulated entities increasingly incorporate AI into their 'know-your-customer' and 'customer due diligence' systems, these systems are still very much focused on detecting suspicious human activities. Such processes may not yet be ready for emerging adversarial AI systems, which use AI to intentionally mislead or manipulate other AI systems. One potential use case is the deployment of adversarial AI to more effectively circumvent export controls. Both North Korea and Iran have found varying degrees of success manipulating shipping records and using third-country transshipment hubs to acquire proliferation-sensitive goods and technologies.

In a recent report on North Korea's illicit procurement activities, Daniel Salisbury notes that:

> The use of procurement agents outside North Korea, intermediaries in third countries (those not hosting the supplier or customer), and the use of a range of deceptive practices have historically allowed North Korea's procurement operatives – and indeed those of other proliferating countries – to dupe industry that is trying to comply with export controls.[43]

Adversarial AI could feasibly make many of these operational steps more efficient, such as the analysis of customs codes, tariff schedules and regulatory frameworks across multiple jurisdictions, to identify the most effective ways to misclassify sanctioned goods or dual-use items to avoid detection. This capability could evolve into 'dynamic misclassification', where the declared nature of goods is algorithmically altered based on perceived risk levels at different ports of entry or in response to changing customs enforcement priorities. Customs scrutiny and priorities can vary significantly by jurisdiction and can change over time due to new regulations or intelligence.

An AI system could analyse data on customs seizures, inspection rates for specific Harmonised System codes (the standardised numerical system for classifying traded products) and recent regulatory updates to build a risk profile for different classifications at various global chokepoints.[44] Based on this continuous analysis, AI could then select the misclassification for a particular shipment that offers the statistically lowest probability of detection for that specific route and time and automatically generate the necessary fraudulent shipping documents to support this declaration.

Defensive AI tools are already being developed to monitor trade flows and identify misclassifications and semantic anomalies in trade documents. The tools are easily adopted by sanctions evaders.

43. Daniel Salisbury, 'Shopping for Mass Destruction: North Korea's Illicit Procurement Networks', *RUSI Occasional Papers* (August 2024), p. 43, <https://www.rusi.org/explore-our-research/publications/occasional-papers/shopping-mass-destruction-north-koreas-illicit-procurement-networks>, accessed 28 October 2025.
44. Transaction-level data, like bills of lading, are now easily sourced through commercial data providers like S&P Global. Trade enforcement data and statistics are routinely reported in court and legal filings and are often open to the public. For an example of how this data is obtained and used, see the recent report by the Open Source Centre – a UK nonprofit that uses open source intelligence to shine a spotlight on international security issues – that explores illicit oil trade between North Korea and Russia. Open Source Centre, 'Follow the Money: Exposing Russian Financing of North Korean Oil Transfers', 20 November 2025, <https://opensourcecentre.org/research/follow-the-money>, accessed 28 October 2025.

Autonomous AI Agents in Evasion Networks

One of the biggest emerging challenges is contending with autonomous AI systems (also known as agentic AI networks). Unlike current use cases, which typically augment a singular aspect of a PF or sanctions evasion scheme – like generating false documents – agentic AI networks operate autonomously, with many agents each executing discrete tasks. The concept of AI agents operating with a significant degree of autonomy represents a frontier in AI development, with profound implications for sanctions evasion.

Such autonomous agents, given broad strategic objectives by human operators (such as 'move X amount of funds from sanctioned entity A to offshore account B while minimising detection risk'), could potentially manage complex segments of sanctions evasion schemes. This could include autonomously managing the finances of a network of shell companies, executing sequences of cryptocurrency transactions through mixers and DeFi protocols, or even generating and deploying deepfake content for social engineering purposes as needed – all acting in concert to achieve a specific objective and with little or no human interaction.

A key driver of agentic networks is the rise in application programming interfaces (APIs) and model context protocols (MCPs) within banking and other financial services.[45] An API is a programmatic interface that allows two different software programmes to communicate and exchange data, acting as an intermediary by taking requests from one application (the client) and delivering a response from another (the server). In a banking context, APIs promote the programmatic automation of a range of tasks from managing accounts to processing payments. Similarly, MCP is a newer open-source standard that is built specifically for agentic AI to communicate with data sources. Whereas a developer may need to write explicit functions to interface with an API, an MCP allows an AI agent to 'plug and play' with different data sources without needing custom code for each one.

According to one recent market trends report, API banking is becoming a priority for most banks to further streamline growing business demands.[46] Furthermore, it is not just banks deploying APIs; companies providing online business services are hosting APIs that provide access to business formation and registration, back-office systems, patent and IP management, transactions and asset management, and tax optimisation services. It is important to note that APIs themselves are neither malicious nor a threat.

45. For a discussion of open banking and the growing use of APIs in finance, see Hakan Eroglu et al., 'Opening Doors to Open Finance: Evidence from the International Experience', Bank for International Settlements, 2026, <https://www.bis.org/publ/bppdf/bispap168.pdf>, accessed 23 April 2026.
46. Lukas Everding et al., 'APIs in Banking: From Tech Essential to Business Priority', McKinsey and Company, January 2023, <https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/tech-forward/apis-in-banking-from-tech-essential-to-business-priority>, accessed 28 October 2025.

However, the rapid adoption of APIs and MCPs to specifically handle AI within finance sectors will attract both licit and illicit actors.

Another potential challenge to sanctions enforcement is agentic networks' persistence and non-reliance on social structures. Traditional criminal and evasion networks often degrade or collapse when key human operatives are arrested, interrogated or otherwise neutralised. However, if AI agents are performing critical operational functions – such as fund movement, maintaining covert communication channels, managing synthetic identities, or adapting tactics in response to detected surveillance – the network could potentially continue functioning, reconfigure itself or even initiate pre-programmed contingency plans, even if its human overseers are compromised or removed.

This potential threat ultimately undermines over two decades of analytic knowledge, which has promoted the need to identify and stop key nodes within a given network. In an agentic AI network, nodes are software instances that are redundant, instantly replaceable and infinitely copyable. The core premise is that criminal organisations (sanctions evasion networks, narco-trafficking groups and terrorist cells) are fundamentally social structures constrained by human limitations like trust, cognitive capacity and physical geography paradigms. Agentic networks threaten to render these traditional paradigms obsolete by removing the human constraints; for example, removing a key node in a narco-trafficking network creates a vacuum that might take months to fill.

AI-Enhanced Social Engineering and Disinformation

Perpetrating a sanctions evasion scheme often requires participation from unwitting actors. In the case of Iran, for example, this meant using third-party intermediaries to assist with transshipment operations. A common scheme entailed Iranian procurement agents sending dozens of enquiries to Western-based manufacturers, purporting to be from locations not subject to export controls and sanctions.[47] In other instances, Iran would enlist the support of Iranian nationals located in Western countries to procure and then illegally export controlled goods.

AI, especially agentic networks, will probably enhance the capacity and capability of proliferating states to recruit unwitting actors into sanctions evasion and PF schemes, by dramatically enhancing the effectiveness of phishing and spear-phishing attacks. By

47. For a recent analysis of Iranian procurement operations, see Simon Mairson and Valerie Lincy, 'U.S. Targets Procurement Network Supplying Machine Tools to Iran', Wisconsin Project on Nuclear Arms Control, 31 October 2019, <https://www.wisconsinproject org/u-s-targets-procurement-network-supplying-machine-tools-to-iran/>, accessed 28 October 2025.

scraping and analysing publicly available data from social media, professional networking sites, company websites and data breach repositories, AI can help craft highly personalised and contextually relevant fraudulent communications. An employee tricked by such sophisticated, AI-driven deception might inadvertently facilitate an evasion scheme, becoming an unwitting insider. This blurs the traditional lines of culpability for insiders and makes it much harder for organisations to rely solely on standard employee awareness training as a primary defence against such targeted and manipulative attacks.

Current Challenges in AI Proliferation Financing and Sanctions Evasion

AI-enabled PF and sanctions evasion pose not merely an incremental evolution, but a fundamental disruption. Preparing for this next wave of threats requires both technological solutions and robust governance frameworks. While governments and financial institutions are taking steps to mitigate AI-enabled threats, more needs to be done. The preceding sections have detailed both current and over-the-horizon technological challenges that both governments and private sector institutions will face. What follows is an examination of these challenges in the context of existing global governance frameworks and initiatives, as well as recommendations for governments, private sector institutions and international organisations.

Current Legal and Regulatory Challenges

The EU Agency for Law Enforcement Cooperation (Europol), in its EU Serious and Organised Crime Threat Assessment (SOCTA) 2025, notes that AI is accelerating crime, lowering the barriers to entry for digital crimes, and enabling the automation of activities like phishing campaigns and the creation of deepfakes for fraud.[48] Similarly, the International Criminal Police Organization (Interpol) has highlighted that financial fraud is being boosted by technologies like AI, LLMs and cryptocurrencies.[49] These organisations, however, have stopped short of making concrete recommendations about how frontline organisations like financial institutions should deploy countermeasures.

48. Europol, *European Union Serious and Organised Crime Threat Assessment*, pp. 10–12.
49. Interpol, 'Financial Fraud Assessment: A Global Threat Boosted by Technology', 11 March 2024, <https://www.interpol.int/News-and-Events/News/2024/INTERPOL-Financial-Fraud-assessment-A-global-threat-boosted-by-technology>, accessed 28 October 2025.

From a technological perspective, financial institutions are increasingly replacing outdated models of fraud detection with machine learning systems. According to a recent survey conducted by the Bank of England, 75% of UK financial firms reported using AI, citing AML and compliance as key usage areas.[50] Similarly, the European Central Bank found that 90% of large European businesses use AI in some capacity.[51]

Traditional AML models largely rely on rule-based detection, which filters for specific criteria, such as whether a transaction meets a reporting threshold or matches an entity on a sanctions list. These methods, however, are increasingly inadequate and unable to flag growing threats and new tactics posed by AI. Instead, machine learning systems attempt to dynamically identify patterns consistent with fraud. These models have proven quite effective but are not without complication. The most compelling evidence for AI's usefulness comes from the Bank for International Settlement's Innovation Hub, which runs controlled experiments comparing traditional rules-based systems against machine learning systems. The programme, called Project Hertha, analysed real-time retail payment systems to identify financial crime patterns. The results showed a significant leap over previous methods: AI models, for example, achieved a 26% improvement in identifying previously unseen criminal behaviours that rules-based systems missed entirely. These results demonstrate that advanced models can effectively draw on network patterns rather than personal data.[52]

Furthermore, banks and sanctions evasion networks are not on a level playing field. AI models deployed by proliferator states can be trained on large repositories of open-source data, leaked financial records and shared criminal knowledge bases (such as dark web forums). Conversely, defensive AI models within financial institutions are often constrained by data privacy laws, such as GDPR.[53] A bank in Singapore, for example, cannot train its detection model on the transaction data of a bank in London. Consequently, this can leave blind spots that global proliferation networks exploit; in other words, the AI models become the equivalent to global governance gaps that proliferation networks exploit to evade sanctions or procure proliferation-sensitive goods and technologies. As previously mentioned, adversarial AI does not necessarily need massive training sets. It merely needs to probe for weaknesses by conducting incremental transactions to identify flagging thresholds, for example.

50. House of Commons Treasury Committee, 'Artificial Intelligence in Financial Services', Fifteenth Report of Session 2024–26, 20 January 2026, <https://publications.parliament.uk/pa/cm5901/cmselect/cmtreasy/684/report.html>, accessed 30 March 2026.
51. Laura Lebastard and David Sondermann, 'Artificial Intelligence: Friend or Foe for Hiring in Europe Today?', European Central Bank, 4 March 2026, <https://www.ecb.europa.eu/press/blog/date/2026/html/ecb.blog20260304~d9e34fc95f.en.html>, accessed 23 April 2026.
52. Bank for International Settlements, 'Project Hertha: Identifying Financial Crime Patterns in Real-Time Retail Payment Systems', 5 June 2025, <https://www.bis.org/publ/othp96.htm>, accessed 28 October 2025.
53. See Article 22(1), Article 5(1)(c), and Articles 13–15 in General Data Protection Regulation (GDPR), 'General Data Protection Regulation (GDPR)', <https://gdpr-info.eu/>, accessed 23 April 2026.

There are, however, new initiatives that seek to address these data problems. Project Aurora, for example, is an initiative hosted by the Bank for International Settlements' Innovation Hub, which focuses on multi-jurisdictional data collaboration to use AI detection tools against money laundering. While the project has demonstrated that collaborative analytics can improve detection rates,[54] the legislative frameworks to support such data pooling at a global scale remain nascent. This data deficit means that while offensive AI is learning from the entire ecosystem, defensive AI is learning only from fragmented slices. Another project by the US Defense Advanced Projects Research Agency is attempting to not only derive algorithms to detect money laundering and sanctions evasion, but also to 'learn a precise representation of how bad actors move money around the world without sharing sensitive data'.[55]

Moving these models from theory and proof-of-concept to implementation is not always straightforward. The need for data privacy, for example, is a limiting factor for the use of AI to identify and fight fraud. The effectiveness of an AI model's ability to identify fraud is directly related to the volume and quality of training data. This can lead to a fundamental tension over data privacy. For example, biometric data, including behavioural information, is increasingly useful for identity management. If compromised, however, a breach of a biometric database could expose individuals to even higher risks of identity theft and fraud.

Furthermore, the rise of adversarial AI specifically targeting compliance systems could trigger an AI security dilemma within the financial industry. While sharing information about AI models, detection typologies and system vulnerabilities can help the entire industry strengthen its collective defences against sanctions evasion, the fear of such information falling into the wrong hands could lead to increased secrecy and compartmentalisation.[56]

The current global governance landscape regarding AI and financial crime is dangerously fragmented. While PF networks operate globally and without borders, the regulatory frameworks designed to contain them are strictly jurisdictional. This disconnect creates a patchwork quilt of rules that generates exploitable points of divergence – gaps that sophisticated actors like North Korea and Iran can leverage for regulatory arbitrage. Furthermore, while many of the legal and regulatory frameworks that do exist seek to balance AI safety with economic competitiveness, they can have the unintended effect of curtailing financial institutions' ability to employ AI-based countermeasures.

54. Bank for International Settlements, 'Project Aurora: The Power of Data, Technology and Collaboration to Combat Money Laundering across Institutions and Borders', updated 7 July 2025, <https://www.bis.org/about/bisih/topics/fmis/aurora.htm>, accessed 28 October 2025.
55. David Dewhurst, 'A3ML: Anticipatory and Adaptive Anti-Money Laundering', DARPA, <https://www.darpa.mil/research/programs/a3ml-anticipatory-adaptive>, accessed 30 March 2026.
56. Kyle A Kilian, 'Beyond Accidents and Misuse: Decoding the Structural Risk Dynamics of Artificial Intelligence', AI and Society (Vol. 41, 2026), pp. 23–42.

The most visible example of this tension between privacy and security exists between the hard law (or legally binding) approach of the EU and the data-exploitation tactics of adversarial states. The EU's 2024 legal framework classifies AI systems by risk, prohibiting those that use deceptive techniques or social scoring. The broad-based framework classifies AI systems and imposes requirements according to four different levels of risk. The highest, 'unacceptable risk', for example, prohibits AI systems that could be used for 'social scoring; and/or systems that use deceptive or exploitative techniques to materially distort a person's behavior in a manner that can cause harm'.[57]

While the EU's approach is by far the most robust, it is nonetheless geared towards mitigating potential risks and harms of legitimate AI use and is lacking when it comes to malicious use. The unintended consequence, of course, is that the regulations may inadvertently deter EU institutions from developing AI-enabled countermeasures. Defensive AI models require vast datasets to learn new evasion patterns; however, strict data privacy laws (like GDPR) prevent the cross-border pooling of transaction data necessary to train these models effectively.

In contrast, state-sponsored evasion networks face no such constraints. They train offensive AI on unrestricted repositories of open-source data, leaked financial records and dark web criminal knowledge bases. This creates a stark asymmetry: offensive AI learns from the entire global ecosystem, while defensive AI is restricted to learning from fragmented, national slices of data.

A second disconnect lies between the EU's restrictive approach and the 'pro-innovation' stance of the US and the UK.[58] The US lacks comprehensive legislation, relying instead on voluntary technical standards from the US National Institute of Standards and Technology (NIST) and an overall stance towards deregulation to maintain global AI leadership.[59] This divergence creates a compliance gap. In the US and the UK, the lack of statutory mandates for AI security in the financial sector means that adherence to defence standards (like the NIST Cyber AI Profile) remains voluntary. Consequently, adoption is inconsistent, meaning that adversarial AI agents can test the defences of multiple institutions, identifying those that have opted for lower standards, and funnel illicit transactions through the path of least resistance.

Finally, a critical gap exists regarding the explainability of AI decisions. Current global guidance on AI remains rudimentary and focuses largely on fraud rather than PF. In

57. Council of the European Union and European Parliament, 'Regulation (EU) 2024/1689 of the European Parliament and of the Council', 13 June 2024, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>, accessed 28 October 2025.
58. HM Government, *A Pro-Innovation Approach to AI Regulation*, Cm 815 (London: The Stationery Office, 2026).
59. The goal of the framework is to help companies address unique challenges posed by AI, such as data poisoning, model evasion (manipulating input data to deceive an already trained machine learning model, causing it to produce incorrect outputs, misclassify data or fail to detect threats), and extraction attacks (stealing or replicating a proprietary machine learning model). As a guideline, however, the framework is voluntary. See Katerina Megas et al., *Cybersecurity Framework Profile for Artificial Intelligence (Cyber AI Profile)* (Gaithersburg, MD: National Institute of Standards and Technology, 2025).

its most recent report on PF threats, for example, FATF acknowledges growing risks from AI but states that 'evidence and trends in this area are too nascent to draw conclusions and few case studies are present'.[60] Without clear international standards, financial institutions are hesitant to deploy black-box deep learning models, which are superior at detecting complex evasion patterns, because they cannot easily explain the machine's decision-making process to regulators. This regulatory hesitation forces institutions to rely on outdated rules-based systems, while adversaries freely deploy unexplainable, high-complexity AI to defeat them.

60. FATF, 'Complex Proliferation Financing and Sanctions Evasion Schemes', p. 38.

Key Recommendations

Recommendations for National Authorities

- **Explore safe harbour provisions for AI-driven CPF**. Banks and other financial institutions remain hesitant to implement black-box models for AML/CPF detection due to their lack of explainability. In some cases, like in the EU, they may be prohibited by law. In other words, while such models may be superior at detecting AML/CPF, it is nearly impossible to explain how the models did so. National regulators, like financial intelligence units (FIUs), should explore safe harbour provisions for financial institutions that deploy AI-driven CPF solutions, which are consistent with national rules and regulations and mitigate any potential risks to privacy.
- **Explore public–private partnerships to implement collaborative technologies**. The conflict between privacy laws (like GDPR) and defensive AI training creates a blind spot. Collaborative analytics technology (like Project Aurora) exists, but the legal framework to use it does not. Governments should explore public–private collaborations, such as a National Security Data Enclave exemption. Here, a legally defined sandbox (namely, an isolated virtual environment where computer code and data can be stored and analysed safely) could be set up in which financial institutions could pool pseudonymised cross-border transaction data exclusively for training counter-proliferation AI models, immune from standard consumer data privacy liabilities. Such frameworks would help with faster adoption of proof-of-concept models.
- **Establish 'compute-KYC' and cloud liability standards**. Most sanctions regimes focus on financial transactions. However, AI-enabled evasion requires significant computational power ('compute') and cloud hosting to run autonomous agents. Governments should expand export controls and sanctions enforcement to the infrastructure-as-a-service (IaaS) layer. Cloud providers should be required to implement compute-KYC to verify the identity of clients renting high-performance capacity capable of running agentic AI networks.

- **Pilot mandated adversarial red-teaming certifications**. Relying on passive compliance checklists is insufficient against active AI agents. Regulators should move from checklist compliance to adversarial certification. Financial institutions above a certain size must subject their screening systems to mandatory, periodic red teaming by certified third-party ethical AI hackers. These red teams would use adversarial AI to probe for vulnerabilities (such as model evasion or data poisoning) as described in the NIST framework.
- **Implement circuit breakers for API-initiated finance**. Agentic AI networks rely on APIs to execute transactions at a speed and scale that is impossible for humans to match. Central banks and payment regulators should pilot circuit breakers for API-initiated transactions. For example, if an API triggers a series of cross-border transfers that exceeds a specific complexity-to-time ratio (for example, moving funds through 10 jurisdictions in 10 seconds), the system must automatically freeze the chain for human review.
- **Increase global AI sanctions evasion awareness**. Because the nature of AI threats evolves rapidly, awareness-raising efforts must also be continuous and adaptive. Donor countries and international organisations (like the US, the EU and the UK) should fund AI-specific capacity-building programmes for FIUs in high-risk regions.

Recommendations for International Organisations and Institutions

- **Update FATF PF national risk assessment criteria**. Current PF national risk assessments focus on human actors and do not account for the speed and scale of autonomous AI agents. While the FATF October 2025 Horizon Scan is a start, it needs to be operationalised.[61] The FATF should update Recommendation 1 (Risk Assessment) and Recommendation 15 (New Technologies) to explicitly recommend that jurisdictions assess the vulnerability of their financial sectors to autonomous AI agents.
- **Develop model legislative frameworks**. There is an urgent need for the international community to develop guidance and best practices regarding responsible use of AI, and development of AI-specific legislative and regulatory requirements to govern the use and deployment of AI systems and to criminalise AI-enabled offences. International organisations like the UN should develop model legislative frameworks that consider the malicious use of AI.

61. According to FATF, the 2025 Horizon Scan 'provides a forward-looking perspective of current and potential Artificial Intelligence (AI) related risks and trends. It forms part of the FATF's staged approach to emerging technologies, with this study aiming to identify and explain developing risks and vulnerabilities associated with AI through the lens of Anti-Money Laundering, Countering the Financing of Terrorism, and Countering the Financing of Proliferation (AML/CFT/CPF)'. See FATF, 'Artificial Intelligence and Deepfakes: Impacts on Money Laundering, Terrorist Financing, and Proliferation Financing', December 2025, <https://www.fatf-gafi.org/en/publications/Methodsandtrends/horizon-scan-ai-deepfake.html>, accessed 28 October 2025.

- **Develop new KYC protocols for AI**. The technology sector, specifically organisations like Coalition for Secure AI, should establish 'know-your-API' (KYA) standards for AI agents. For example, a KYA could require AI agents accessing financial APIs to maintain a verifiable identity token linked to a verified individual.
- **Increase investigations into proliferator use of AI**. In the absence of the UN North Korea Panel of Experts, the Multilateral Sanctions Monitoring Team should form a working group dedicated to identifying North Korea's use and evolution of AI in its cyber and sanctions evasion activities.

Recommendations for Private Sector Institutions

- **Financial institutions should update CPF KYC procedures**. The Arup deepfake fraud demonstrates that static biometric checks (for example, a simple selfie or voice print) are no longer sufficient proof of identity against AI-enabled adversaries. Financial institutions should upgrade CPF KYC procedures and standards.
- **Deploy defensive AI against trade documentation**. Financial institutions should deploy defensive AI to audit trade documentation, which analyses the content of trade documents for semantic inconsistencies (for example, a shipment of textiles that weighs as much as heavy machinery) that GenAI might overlook.
- **Deploy dynamic compliance modelling**. Financial institutions should adopt compliance frameworks and investigative approaches that are more dynamic, incorporating behaviour-based analytics and predictive modelling – often leveraging defensive AI capabilities – to effectively counter these evolving threats.

Ultimately, the shift towards AI-enabled sanctions evasion requires a corresponding shift within public and private sectors. For this to happen, governing rules and regulations must empower these institutions to prepare for the use of adversarial AI.

About the Author

Aaron Arnold is a Senior Associate Fellow with the Centre for Finance and Security at RUSI, where his work focuses on sanctions and proliferation financing.

Prior to joining RUSI, Aaron served as the finance and economics expert on the UN Panel of Experts for North Korea sanctions, where he monitored global sanctions implementation and investigated instances of sanctions violations. Before joining the Panel of Experts, Aaron was a fellow with the Project on Managing the Atom at the Harvard Kennedy School's Belfer Center, where he published work on the extraterritorial use of sanctions and the efficacy of WMD trade controls.

He also previously worked as a counterproliferation subject matter expert in the U.S. Department of Defense and the U.S. Justice Department, where he specialised in WMD counterproliferation investigations and operations, with an emphasis on threat finance and sanctions evasion.

Aaron has a PhD and a Master's in Public Policy and National Security from George Mason University and a BA in Political Science from Virginia Tech.

195 years of independent thinking on defence and security

The Royal United Services Institute (RUSI) is the world's oldest and the UK's leading defence and security think tank. Its mission is to inform, influence and enhance public debate on a safer and more stable world. RUSI is a research-led institute, producing independent, practical and innovative analysis to address today's complex challenges.

Since its foundation in 1831, RUSI has relied on its members to support its activities. Together with revenue from research, publications and conferences, RUSI has sustained its political independence for 195 years.

RUSI is registered as a charity in England and Wales

Charity number: 210639

VAT number: GB752275038