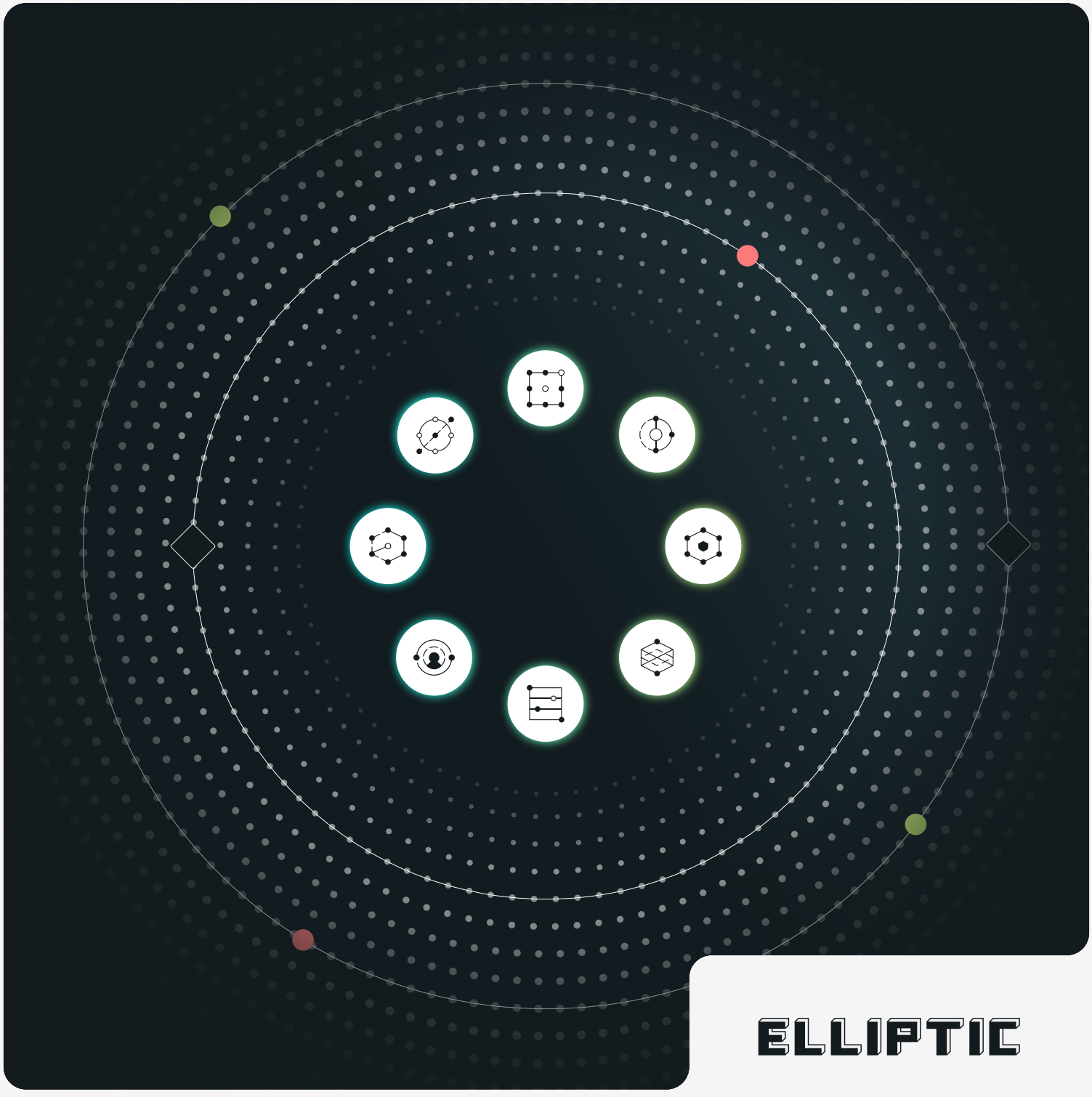
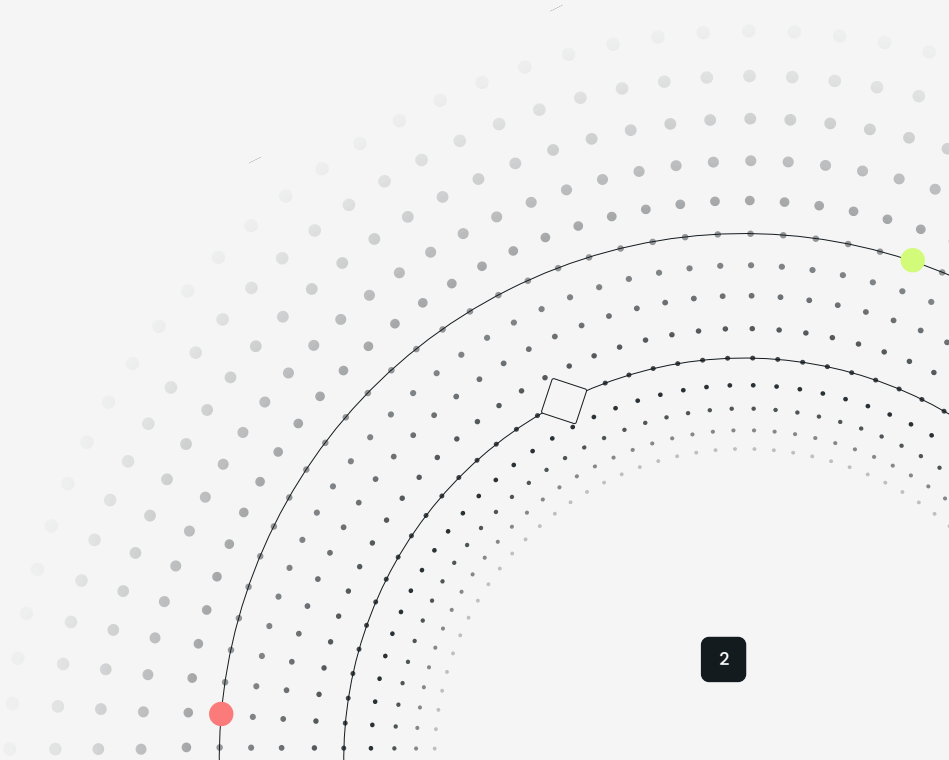
The Elliptic Standard

Principles for agentic on-chain risk



ELLIPTIC

Contents

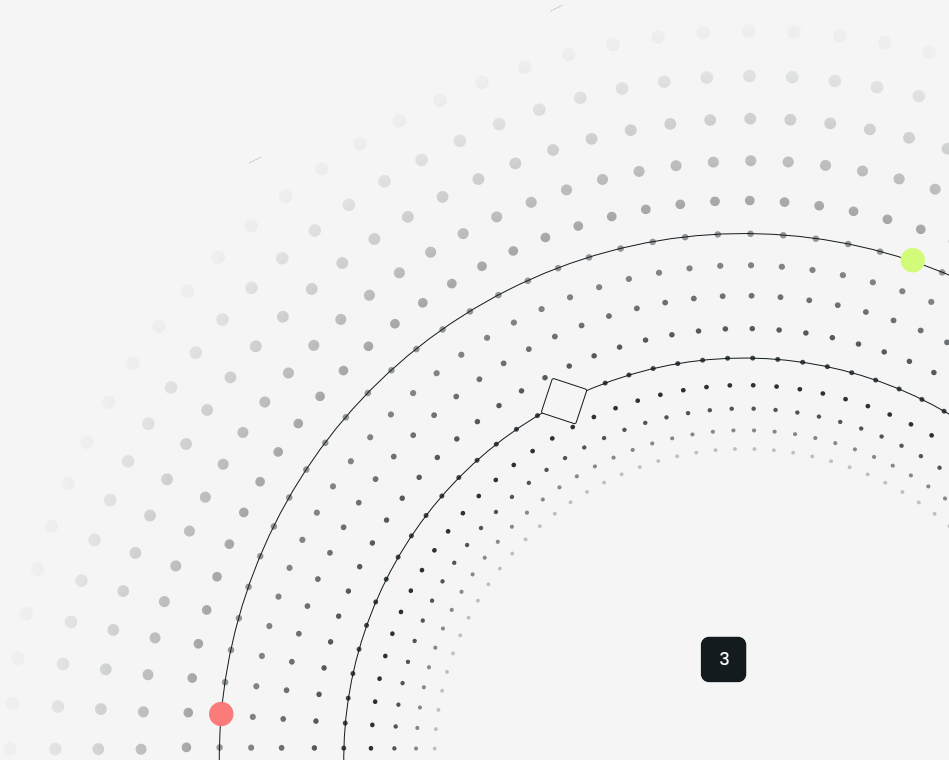


Introduction

The Elliptic Standard sets out eight principles for the responsible design and use of agentic artificial intelligence (AI) in on-chain risk solutions. It is written for the compliance, risk and policy leaders who will be accountable for these systems, and for the teams who will run them.

A company evaluating agentic on-chain risk should use these eight principles to assess a vendor and expect a substantive answer for every principle. Similarly, a company upgrading its own capabilities can treat the Standard as a design brief. Elliptic has adopted these principles as it develops its own on-chain risk solutions and expects to be held to them.

We developed the Standard with input from key partners, including regulated companies we are collaborating with to build AI-powered capabilities.



Why now

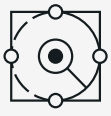
The Elliptic Standard addresses three significant developments that are reshaping how financial crime risk has to be managed:

On-chain finance is moving into the mainstream. Major financial institutions are developing tokenized financial services capable of reducing settlement times, providing 24/7 liquidity and supporting a programmable financial infrastructure suited to digitized markets. Concurrently, stablecoins are being integrated into merchant settlement and cross-border payment services at an unprecedented pace. Stablecoin transaction volumes reached $33 trillion in 2025, up 72% year on year.[1] Citi's base scenario has stablecoins supporting up to $100 trillion in annual transaction activity by 2030 if velocity approaches that of fiat currency.[2] Human review alone cannot keep up with these on-chain volumes and speeds, or the added complexity as funds switch between on-chain and off-chain rails.

Autonomous agents will originate, authorize and settle more and more transactions. Research by PHD and WARC puts agent-facilitated consumer spending at $944 billion in 2026, rising to $3.35 trillion by 2030, or 3.8% of global consumer spending.[3] Gartner's modeling is broader in scope, estimating that machine customers will directly influence or participate in purchases worth $30 trillion by 2030.[4] Risk management insights have to be machine-consumable to function in that environment, and controls have to respond to risk indicators that surface at agentic speed.

Illicit actors are already using AI. Fraudsters, cybercriminals, hostile nation states and other bad actors use AI across the full criminal lifecycle, from establishing access to moving proceeds. They use it to produce identity documents that pass Know Your Customer (KYC) checks, write and personalize the communications behind scams at scale, find and exploit technical vulnerabilities, automate the movement of funds through layers of intermediaries, and build the infrastructure that supports it all. The result is crime with lower barriers to entry, because criminals can develop faster and better tools.

Blockchain analytics as used today cannot meet these challenges. Detection arrangements built around post-hoc, investigation-heavy reviews already cannot keep pace. The Wolfsberg Group has observed that legacy tools "have challenges with adapting to the risk of today's financial services sector, which is rapidly expanding cross border transactions, increasing volumes, and executing faster payments."[5]

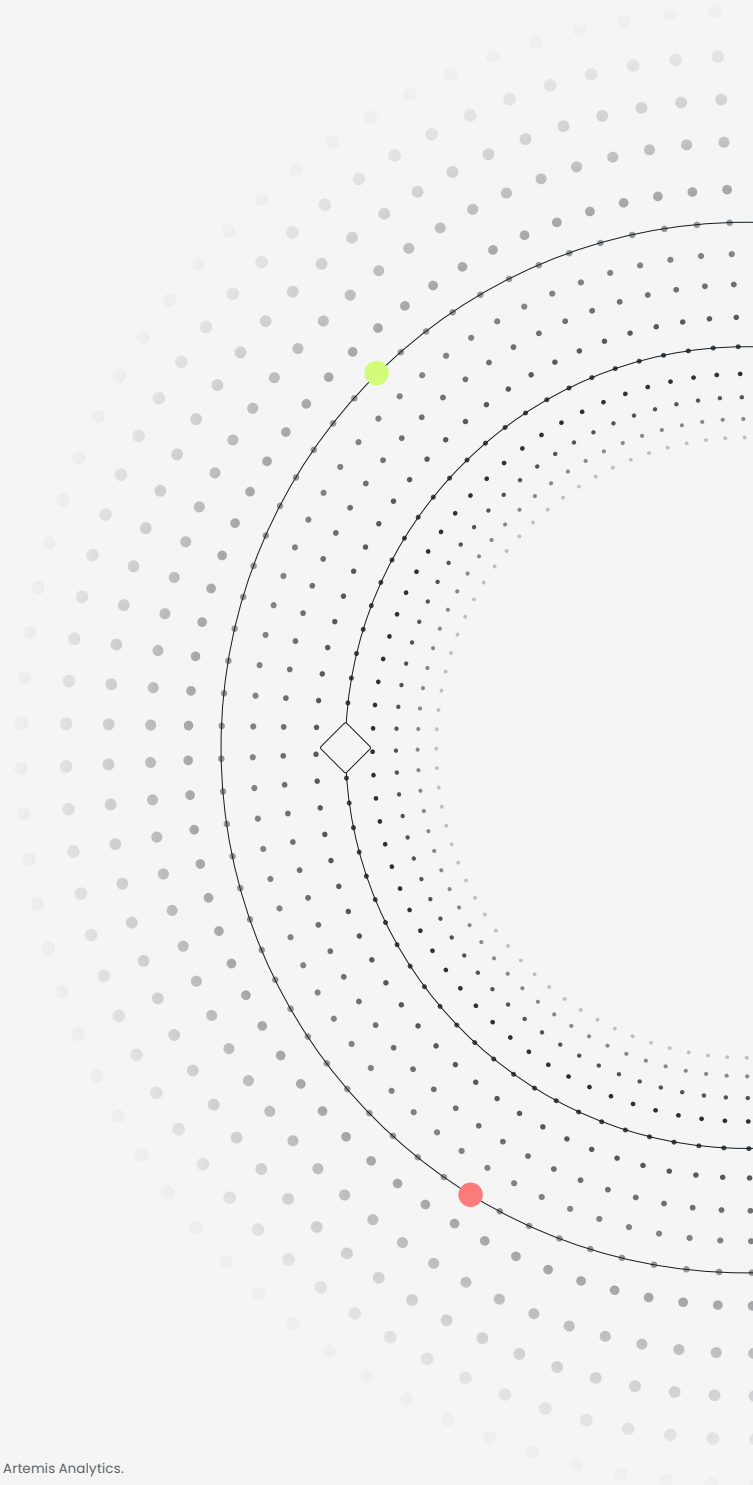


What has to change is the architecture built on top of it, which must move past a forensics-first model toward real-time detection that catches risk before it becomes systemic or entrenched.

When cross-border digital asset transactions settle between agents in real time, analyst teams cannot spend hours clearing false positives or reading unstructured data in the hope of spotting something. Nor can detection exist solely to satisfy a process, producing defensive suspicious activity report (SAR) filings that serve no investigative purpose.

The solution needs to be intelligence-led. Human attention should go to the signals that matter most, in real time, with the aim of disrupting activity rather than documenting it. This will improve the quality of the outputs of risk management processes and result in better filings. Wolfsberg makes a similar point about what a mature risk management approach looks like, accepting that it does "not detect everything historically considered suspicious, but importantly improves the identification of quality leads for law enforcement, uncovering new risks that the legacy platform did not."[6]

On-chain data remains as valuable as it has ever been. What has to change is the architecture built on top of it, which must move past a forensics-first model toward real-time detection that catches risk before it becomes systemic or entrenched.



[1] "Stablecoin transactions rose to record $33 trillion in 2025," Bloomberg, January 8, 2026, citing Artemis Analytics. https://www.bloomberg.com/news/articles/2026-01-08/stablecoin-transactions-rose-to-record-33-trillion-led-by-usdc

[2] Citi Institute, "Stablecoins 2030," September 2025. https://www.citigroup.com/global/insights/stablecoins-2030

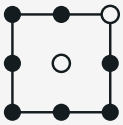
[3] PHD and WARC, "From abundance to agents," July 2026. https://www.phdmedia.com/new-phd-x-warc-research-from-abundance-to-agents-how-the-delegation-of-choice-is-transforming-marketing/

[4] Gartner, reported in Forbes, February 2025. https://www.forbes.com/sites/torconstantino/2025/02/18/machine-customers-ai-buyers-to-control-30-trillion-in-purchases-by-2030/

[5] The Wolfsberg Group, "Statement on Effective Monitoring for Suspicious Activity, Part II: Transitioning to Innovation," 2025, https://wolfsberg-group.org/resources/202/202

[6] Ibid

The eight principles



1. Exceptional data quality

On-chain data should not just reconstruct what happened after funds have reached their destination. A dataset cannot be judged solely on whether a particular attribution proves accurate or on the number of labeled addresses it contains. Agentic systems built upon a large dataset of poor quality will simply draw more incorrect conclusions, faster. For agentic risk intelligence capabilities, the standard for exceptional data quality must be higher, capable of guaranteeing the accuracy and defensibility of risk decisions at speed and scale, and must enable more than post-hoc analysis. The assessment of a dataset's quality must go beyond counting the number of on-chain attributions, taking account of a range of factors that evidence its breadth, depth, accuracy, timeliness and relevance. This requires:

Comprehensive, holistic data coverage. The dataset powering agentic models must be able to follow at least 99% of value-moving transactions across cryptoassets and blockchains, incorporate bridging events and off-chain intelligence. Cross-chain and cross-asset risk analysis must not be subject to an arbitrary hop limit and must enable agents to identify risk across the various pathways through which funds flow.

Attribution quality. The proportion of high-confidence attributions in a dataset determines what can be built on it. Attributions differ in how well they are supported, and a dataset can be extensive while much of what it contains is weakly evidenced. An agent operating on a dataset weighted toward low-confidence attributions will simply produce inaccuracies faster, which is why attribution quality has to be high throughout the entire dataset.

Contextualization. Each data point should arrive with enough supplementary information, such as hops, paths traversed or time-bound analysis, to offer meaningful insights and support a decision with clarity and confidence.

Real-time and reliable. Insights must be surfaced at agentic speed and at scale. Analysts and agents need contextualized intelligence in real time so they can act against the highest risks without losing accuracy. Insights about emerging risks should also feed back into risk controls, so the system keeps up with new threats over time.

Compatibility with off-chain data sources. On-chain data is one input among several, including off-chain transaction data, KYC information, sanctions data and other risk signals. On-chain data must be accessible in a format that integrates readily into multi-source, off-chain workflows.

Consistency of outcome. The same data state must produce the same conclusion. Reasoning paths may vary depending on the underlying model, but a system should never return different determinations for identical inputs.

Evidenced data. Any data an agent surfaces should be supported by an evidence trail that verifies claims about its accuracy, together with the reasoning and policy citations that show how a decision was reached.

Data quality must be measured by how much it supports better and more confident decisions in real time and at scale. It must also improve the effectiveness of risk management systems and controls over time through a feedback loop where it incorporates insights about emerging risks. Where coverage is incomplete, the gap must be surfaced, not absorbed into a confident output.

2. Model transparency and validation

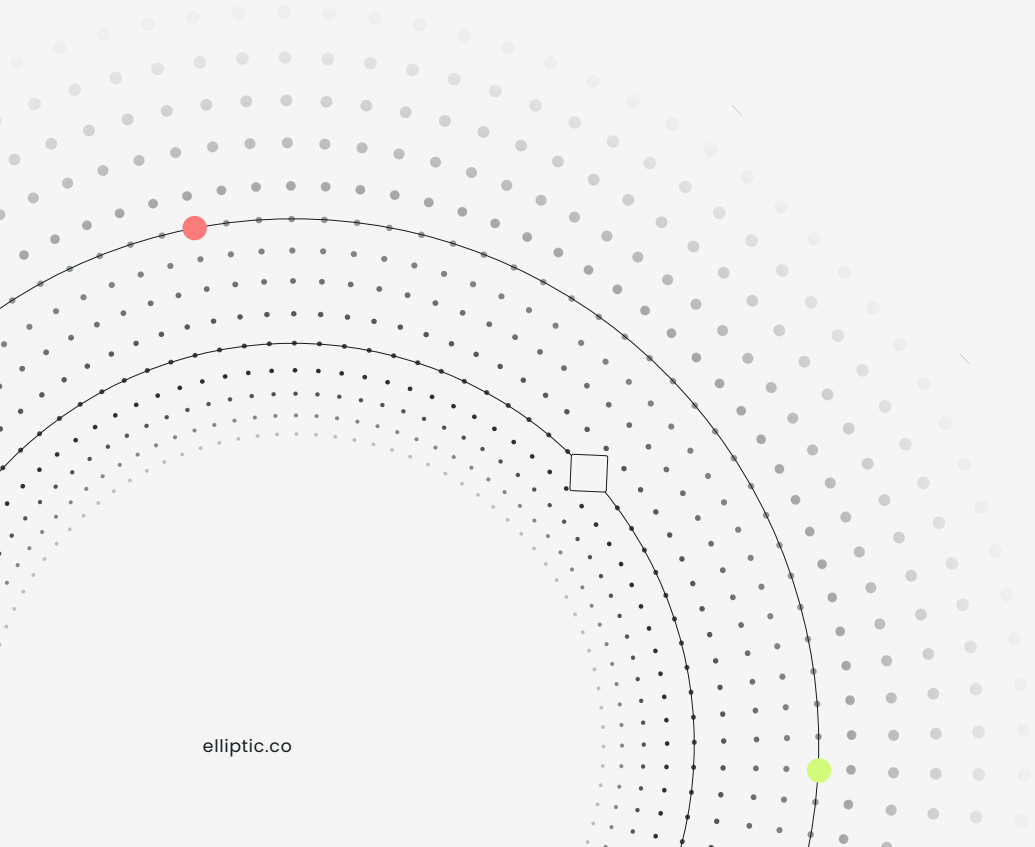
Agentic systems used across the financial crime risk management and intelligence lifecycle must not be black boxes. Their underlying logic cannot be opaque or incomprehensible to human reviewers, including those in charge of assessing the completeness of controls (e.g. regulators). How they arrive at an action or conclusion must also be explainable in relation to its intended use.[7]

Every model deployed to support on-chain risk management should be sufficiently auditable to verify that it operates in line with its intended goals, including through a coherent evidence trail of the steps taken to reach a given outcome. Providers must commit to robust testing of the models they use in training and ongoing deployment, with evidence of credible, objective third-party validation.

That evidence must be comprehensible and available to the people using the system and to relevant parties such as regulators. It should be available in three forms: a machine-readable verdict a downstream system can act on, a plain-language summary with accurate dates, values, hashes and (where relevant) designation dates, and the underlying structured evidence.

Documentation should be clear and detailed enough that a reviewer can reproduce the implementation and the conclusions. Evaluation methodologies and the thresholds used to validate performance should be documented on the same basis.

Equally, validation cannot stop at deployment. It must continue afterward to identify performance drift or degradation and cover models as they undergo recurrent training.[8]



3. AI safety

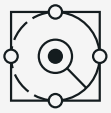
Agentic on-chain risk solutions must be designed and maintained to guard against common AI-related risks. The Wolfsberg Group has set out that the use of AI for financial crime detection should support "fair, effective, and explainable outcomes."[9]

Model validation and other testing and auditing must account for principal AI safety risks, and oversight arrangements before and after deployment must include a process for responding to identified safety issues. Testing and audit records should document how those issues were found and what was done about them.

Risks that AI safety arrangements should account for include, but are not limited to:

- Model drift
- Hallucinations
- Bias and potentially discriminatory outcomes
- Unintended behaviors
- Judgment errors and false negatives
- Misrepresentation and deception
- Prompt injection and data poisoning
- Unintended or unexpected black box effects

Data security and privacy belong in the development of agentic systems from the start. Data should be used in the agentic lifecycle only where it is fit for the purpose and appropriate to the use case, under robust data governance and access controls, and handled in line with applicable regulatory and legal requirements.



Data security and privacy belong in the development of agentic systems from the start.

[7] Monetary Authority of Singapore, "Artificial intelligence model risk management," December 2024. https://www.mas.gov.sg/publications/monographs-or-information-paper/2024/artificial-intelligence-model-risk-management

[8] Financial Conduct Authority, "AI and the future of retail financial services (The Mills Review)," July 2026. https://www.fca.org.uk/publications/corporate-documents/mills-review

[9] The Wolfsberg Group, "Principles for using artificial intelligence and machine learning in financial crime compliance," December 2022. https://wolfsberg-group.org/resources/202/93

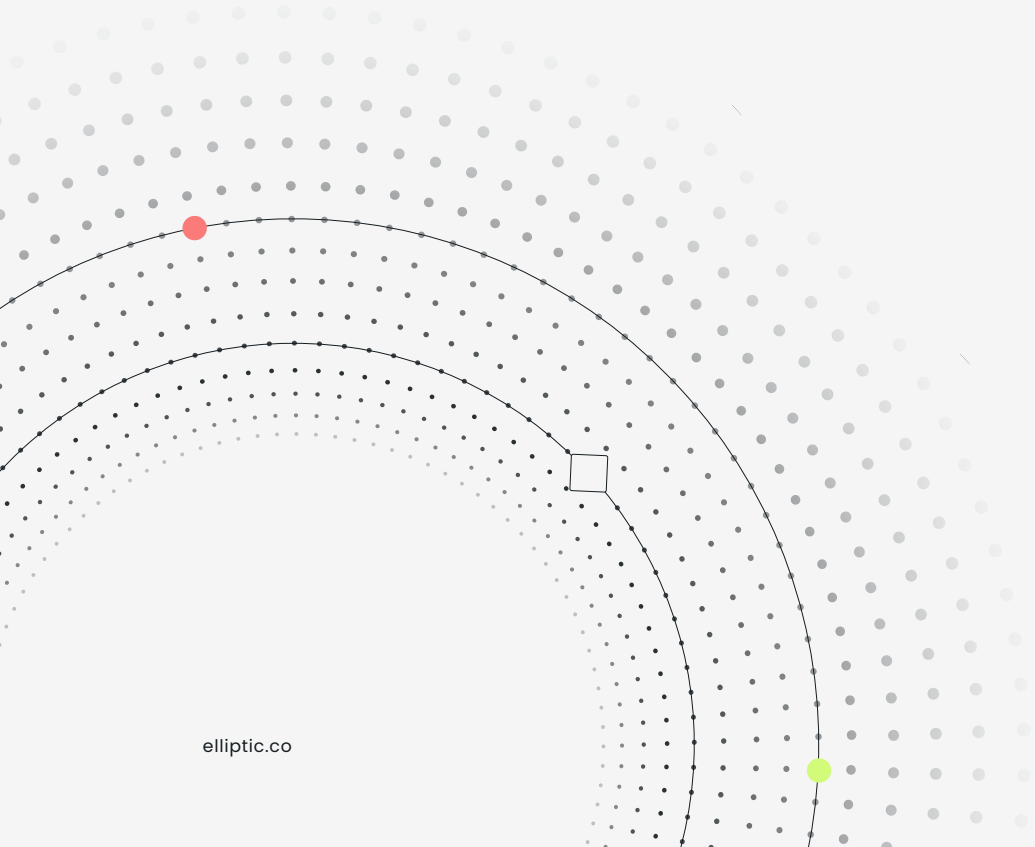
4. Foundation model agnosticism

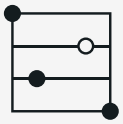
Agentic on-chain risk solutions must remain agnostic to the foundation models they use for reasoning. The US Treasury has identified reliance on third-party providers as one of the risks amplified by AI adoption in financial services, and noted the dependency this creates for smaller firms in particular.[10] Agentic capabilities compatible with only one model, or a very limited set, can hinder the effectiveness of agentic systems, increase risk for the user and run afoul of local regulations.

Model-agnostic agentic risk solutions reduce exposure to foundation model failure and other concentration risks.

They also help users address multi-jurisdictional regulatory and legal obligations that can arise from reliance on a particular model and give them the flexibility to work with blockchain data through agents regardless of their preferred underlying model.

Agentic on-chain risk solutions should also allow customized foundation models that are tuned to the risk requirements of financial institutions and allow for private models that avoid exposing sensitive data to third parties.





5. Configurability

Every regulated entity faces its own risk considerations, operational requirements, governance arrangements, regulatory obligations and jurisdictional realities, all of which evolve over time. Consequently, it has been a longstanding principle of financial crime risk management that regulated firms must not blindly rely upon "out-of-the-box" solutions from vendors and must retain the ability to configure and tune risk detection systems aligned to their firm-specific risks.

This principle remains imperative and must not be sacrificed as agentic systems are developed and deployed. Agentic systems must be adjustable to the circumstances of the company deploying them, so their use supports a genuinely risk-based approach and proportionate compliance arrangements. The fact that an agent can operate with a level of autonomy does not obviate the need to ensure that the underlying risk detection parameters used to guide their actions are configured to firm-specific considerations of risk. Indeed, ensuring the appropriate configuration of underlying systems and controls becomes even more important in the presence of agents, which can only be as effective as the risk-based approach that guides their actions.

Specifically, the user must retain ownership of the factors that determine what the system does. Configurability belongs in agentic on-chain risk capabilities from day one.

At a minimum, the user should retain the ability to:

- ✓ Determine the risk appetite thresholds the system must align to and reflect, based on institutional policy frameworks and regulatory obligations
- ✓ Define the escalation logic and human oversight thresholds across the agentic risk management lifecycle, in line with their own policies
- ✓ Align the execution of actions and delivery of outputs to their own standard operating procedures (SOPs)
- ✓ Train, judge and measure the reasoning of agents to improve their accuracy and recall
- ✓ Override agent actions or determinations
- ✓ Use their own narrative templates, including SAR templates, that meet specific evidentiary standards and requirements
- ✓ Define what counts as a match, so the system reports against the company's own definitions, not the provider's

A company using an agentic system must always have control over the decisions that matter. It must not be beholden to or limited by a vendor's default configurations or definitions and it must not assume that an agent is operating in alignment with the company's risk based approach simply because an agent can execute certain actions efficiently.

[10] US Department of the Treasury, "Artificial intelligence in financial services: report on the uses, opportunities, and risks of artificial intelligence in financial services," December 2024. https://home.treasury.gov/news/press-releases/jy2760

6. Human oversight by design

Human oversight is a non-negotiable principle that has to be present in the design of any agentic system or process. While agentic solutions will reduce the process burden on analysts, they do not absolve individuals from legal accountability. In fact, human oversight, accountability and governance become more critical the more agents are incorporated into a company's risk processes.

That said, a person who is answerable for an agent's actions needs the surrounding controls to show them what the agent is permitted to do, what it is doing, what it has done and on what basis, and to give them the means to intervene.

Human oversight can take various forms, and model design should be flexible enough for the user to implement the governance arrangements best suited to a company's use case.[11]

Specifically, features of human oversight arrangements should include:[12]

- **Scope of authority.** The actions each agent is permitted to take, the data and systems it can reach, and the thresholds beyond which it cannot act without human approval. Authority should be explicit, not implied by configuration, and an approval at one step should not carry authority into subsequent steps.
- **Identity and attribution.** Which agent took a given action, under whose authority it was operating and which person or team owns it.
- **Activity in progress.** Visibility into what agents are doing as they do it, so intervention is possible while it can still be effective.
- **Decision trail.** For each action or conclusion, the evidence considered, the reasoning applied and the policy or threshold that determined the outcome.
- **Points of human engagement.** Which categories of action execute automatically, which require approval before proceeding, which are escalated for review and which are recorded for observation only.
- **Means of intervention.** Appropriate authorizations and the ability to intervene and potentially override an agent's actions or conclusions, including the ability to halt a class of actions rather than a single instance.
- **Evidence of oversight.** Documented evidence that can be reviewed to measure and test how well oversight arrangements and the resulting decisions are working, and show that oversight was exercised, not merely available.

7. Business resilience

Agentic capabilities change the shape of the control environment they enter, and that environment has to be designed around them. Adding agents alongside controls built for human throughput leaves the surrounding process unable to keep pace with them. Retiring long-standing procedures on the assumption that agents now cover the work risks removing steps whose function was never fully documented and leaves nothing in place when the agentic process is unavailable.

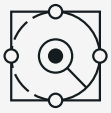
Continuity and robustness of operations must remain a priority. Agentic models and systems must be incorporated into existing workflows deliberately, with the surrounding steps adjusted to account for what the agent now does, and it must work with the other data sources and systems the company relies on. Both the agentic system and the process around it should be monitored before and after deployment for risks to resilience.

Where a process has been handed to agents, the company needs a defined fallback. What happens to transactions in flight? Which controls revert to manual review? Who is notified and how is the backlog cleared once service resumes?

Where a company depends on an agentic system for controls that operate continuously, availability itself becomes a control requirement. Providers should be able to evidence sustained uptime better than 99.9%. This should be measured and reported rather than asserted.

Systems should also reinforce institutional memory. Evidence and audit trails covering both agentic and human actions, documentation of the decisions that followed and records of changes made should all be built in, reducing the risk of knowledge loss when staff leave or are unavailable.

> Continuity and robustness of operations must remain a priority.



[11] Financial Stability Board (FSB), "Sounds Principles for Responsible Adoption of Artificial Intelligence (AI) - Consultation Report," June 2026. https://www.fsb.org/uploads/P100626.pdf

[12] Monetary Authority of Singapore, "Consultation paper on guidelines on artificial intelligence risk management," P017-2025, November 2025. https://www.mas.gov.sg/publications/consultations/2025/consultation-paper-on-guidelines-on-artificial-intelligence-risk-management

8. Empowering the risk professional of the future

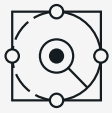
Human judgment and scrutiny carry more weight in a world of risk and intelligence agents. Systems must be designed with the human decision-maker in mind, and made intelligible, usable and intuitive to them.

Firms developing and deploying agentic solutions must prioritize and invest in efforts to equip risk management professionals with the capacity, skills and confidence needed to thrive in a new landscape. Risk management professionals must be empowered to acquire new skills within evolved roles that will feature less intensive process execution but will demand greater accountability for, and comprehension of, agentic systems and their outputs.

New systems should arrive with the training and educational resources that give employees the competence their roles require. Training and education must support institutional learning that covers proactive risk identification, critical thinking, typology comprehension, AI safety, human accountability and system configurability and testing. It should also give the workforce a clear understanding of how agentic workflows map to the company's SOPs and equip it to build the escalation pathways and controls that gate agent outcomes.

Equally, employees should learn to interrogate an agent's reasoning and evidence trail rather than accept or override a score. They also need the skills to ensure explainability and defensibility when working with agentic systems, including how to document and defend an override or escalation.

Human judgment and scrutiny carry more weight in a world of risk and intelligence agents.



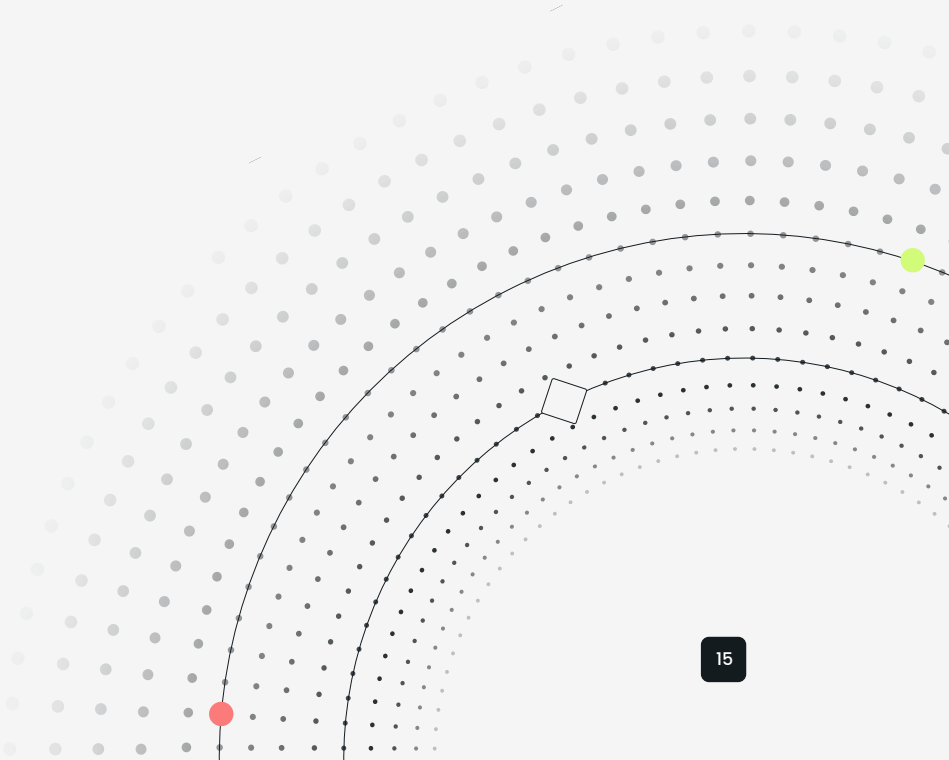
Conclusion

Financial crime risk management is being pulled in two directions at once. The activity it monitors is moving on-chain as stablecoin activity and tokenization scale, settling faster and increasingly being initiated by machines rather than people. At the same time, the illicit actors it exists to detect have the same access to AI as the institutions defending against them, but do not face the same constraints as regulated institutions when it comes to the speed of adoption.

Agentic capabilities are therefore necessary, not optional. But necessity does not lower the bar those capabilities must meet.

A system operating at agentic speed produces far more decisions, and the company deploying it remains answerable for every one of them. As a result, the system's design and how it is used must be defensible. This is only possible when the company responsible for a risk decision can demonstrate what the decision was, what evidence supported it, how the system reached it and who was accountable for it. That test does not change because an agent performed the work. The eight principles set out what an agentic system has to do, and what a company has to require of it, for that test to be met.

Elliptic is already following and implementing these principles as it develops agentic on-chain risk solutions and expects to be held to them. No company should accept less from any provider.



ELLIPTIC

Elliptic is recognized as a WEF Technology Pioneer and backed by investors including J.P. Morgan, Wells Fargo Strategic Capital, SBI Group, Santander Innoventures and HSBC

Founded in 2013, Elliptic is headquartered in London with offices in New York, Washington D.C., Dubai, Singapore and Tokyo.

For more information or to follow us, visit

www.elliptic.co

LinkedIn

X