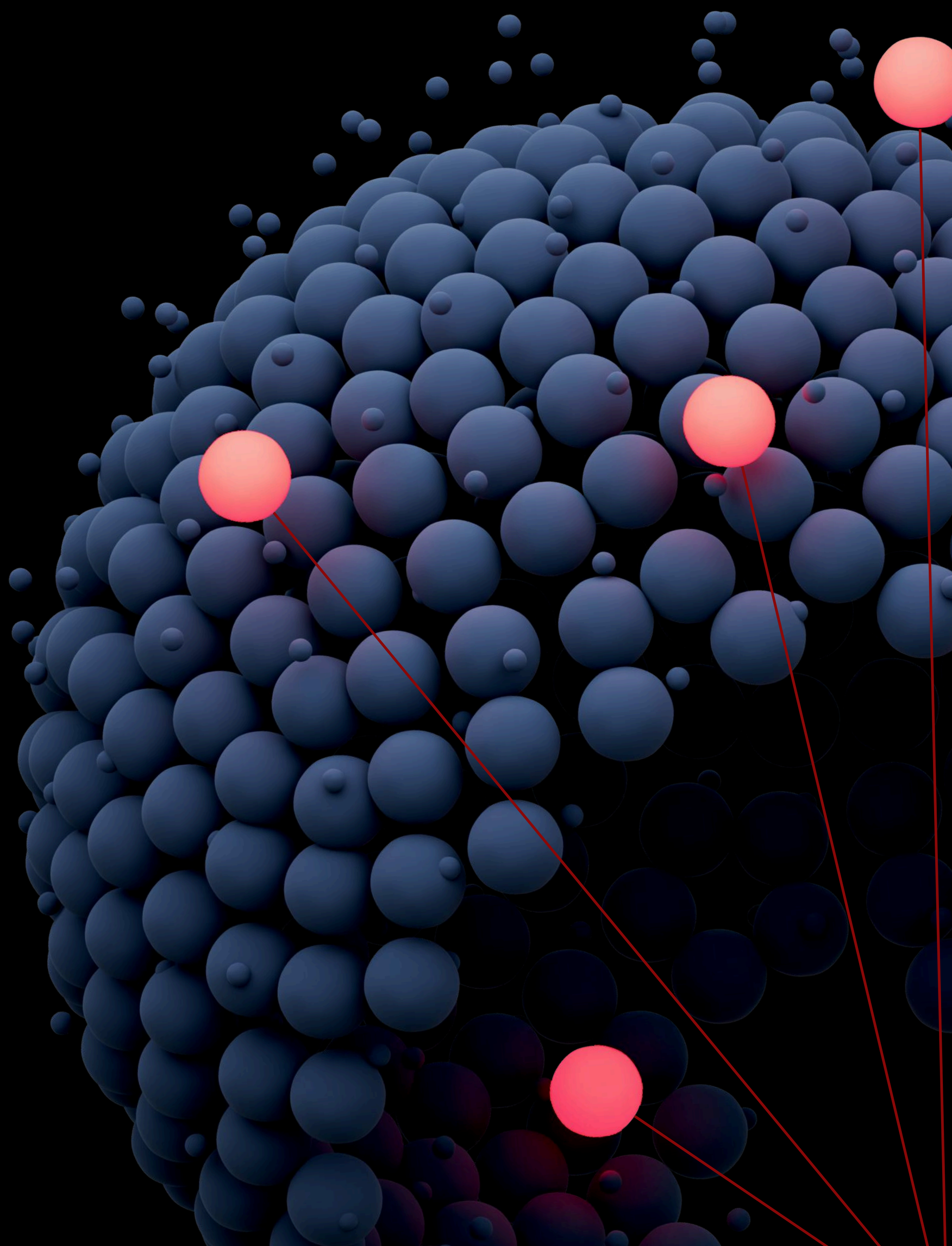
CORE·3

CORE3 Probability of Loss
The agent-readable risk benchmark for digital asset issuers and exchanges.

Authors: **CORE3 Research team**
Date: **June 8, 2026**



Abstract

Risk decisions are only as sound as the data they read. When that data is wrong or poisoned, the actor consuming it carries the full blast radius of the error, and that radius is widening: AI agents are becoming the first reader of crypto data ahead of an allocation or trade, so a single bad signal propagates to every decision that agent makes.

Public trust signals such as audits, market cap, TVL, and named partners describe only a fraction of the surface through which capital is lost.

This paper defines the Probability of Loss, an index ranging from 0 to 100 that is computed from public data and calibrated against the largest documented Web3 failures. In this document, we demonstrate why a standardized risk benchmark is currently required by the blockchain market, detail the methodology behind the Probability of Loss index, and show how gaps in a public risk posture, consumed identically by a human reviewer and by the automated systems acting on a humans's behalf, can be used to predict both the probability of an incident and the potential exploit surface.

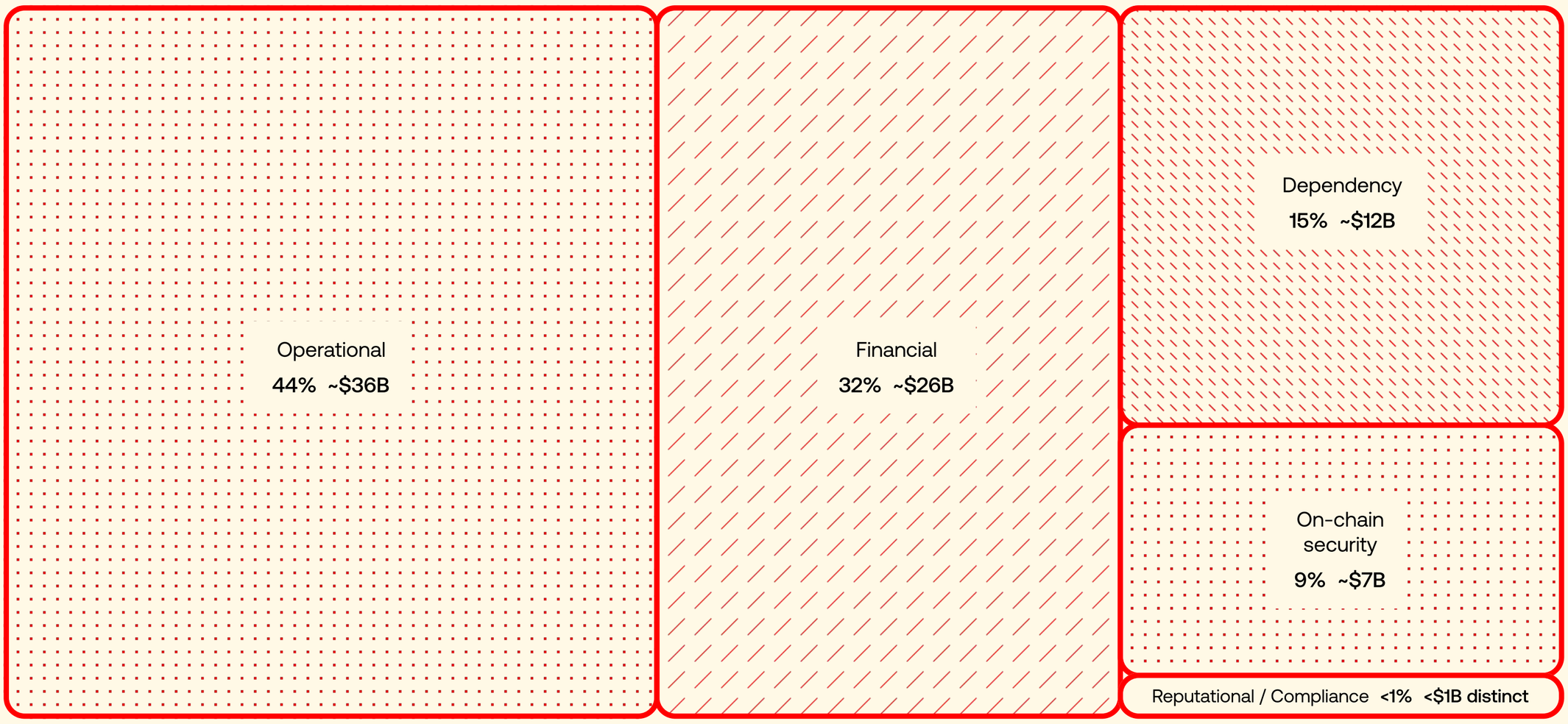
01

Web3 losses are cataloged, yet risks are not priced.

Since 2011, documented Web3 failures have destroyed approximately $81B in user capital across more than 3,000 cataloged events. Sorted by what was missing in the public record the day before each event, the losses cluster into six categories.

$81B was lost in crypto category

Order of magnitude: **~$81B**
Events catalogued: **3,000+**
Period: **2011 to May 2026**



Sources:

DefiLlama Hacks ↗
Crystal Intelligence Top-10 CEX Hacks ↗
Root Cause of $1B Smart Contract Losses, arXiv ↗
Chainalysis 2026 Crime Report ↗
Phemex DeFi Hacks 2026 ↗

For context, the global average corporate data breach in 2025 cost $4.44M per incident. April 2026 set the monthly record for crypto exploits: 28 to 30 separate incidents, $629M stolen, and three days in the month without an event above $10K. May 2026 continued at a similar cadence of 36 incidents in 30 days.

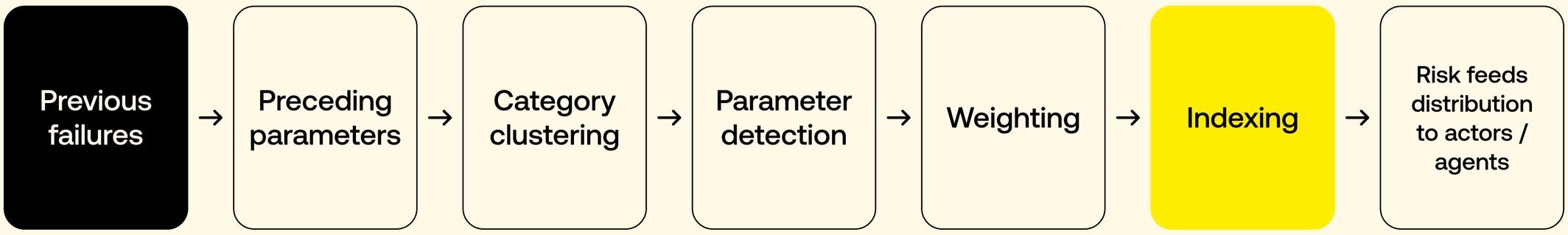
Historically, several collapses have shaped the current state of risk infrastructure in the Web3 market: FTX collapse—publishing proof of reserves. Celsius—sustainable yields. Terra—economic design of self-dependent ecosystems. Because the cryptocurrency market is decentralized by design, enforcing risk standardization is impossible. Consequently, risk management practices are primarily formed to mitigate the likelihood of exploiting the most common vulnerabilities. This approach introduces two drawbacks: first, it is highly reactive; second, it only tracks the most prevalent risk vectors at any given moment. As a result, certain historical patterns continue to be re-exploited, regardless of how simple or extensively documented they are at the time of the exploit.

To monitor both historical attack patterns prone to re-exploitation and to evaluate a digital asset issuer's alignment with modern risk management practices, a public risk-reading layer is required. This layer analyzes disclosed controls, weights their absence based on historical loss correlation, and publishes the findings in a format accessible to all market participants.

Accessibility extends to both modes in which the market now reads risk. A risk officer requires the composite index, as well as the parameters that drove it, while automated screening agent needs same data, but in callable format requested via API. CORE3 is one implementation of that risk data layer, being unbiased toward the project being evaluated.

02

Measuring risk exposure of digital asset issuers based on historical failure patterns.



A familiar reader can name numerous metrics that suggest a Web3 project is healthy: TVL, high-profile partners, a doxxed team, a smart contract audit, high trading volume, rising market cap, or frequency of GitHub commits. While some of these metrics are commonly treated as proxies for stability or early indicators of future failure, in practice, they do not accurately reflect actual risk exposure.

The CORE3 methodology names a framework of over 85 parameters that serve as direct indicators of risk exposure. The vast majority of these metrics evaluate transparency, security practices, incident monitoring and prevention, operational maturity, economic health, and regulatory and reputational posture.

These evaluation parameters were derived directly from post-mortems of the largest documented Web3 failures. For each adverse event, the analysis identified specific practices or controls that would have contained, minimized, or prevented the failure of the respective protocol or asset the issuer. Following this analysis, the risk parameters were clustered into six core categories. Both individual parameters and their overarching categories were then assigned weights based on their severity and historical correlation with resulting losses.

Each parameter is structured as a discrete binary or threshold-based condition and contributes to the Probability of Loss index only when verifiable. Crucially, an absent control is scored as a risk not because its absence definitively proves the existence of a vulnerability, but because the public record cannot rule one out. Therefore, the methodology is deliberately conservative, treating undisclosed information as unmitigated risk.

Project Probability of Loss weights

Security
35%

- Audit coverage
- Bug bounty
- 3rd-party monitoring
- Post-incident posture

Operational
20%

- Team identity
- Key custody (CCSS)
- Incident response
- ISO 27001 / SOC 2

Financial
15%

- Treasury quality
- Revenue source
- Yield sustainability
- Token supply mechanics

Dependency
15%

- Bridge / cross-chain
- Oracle architecture
- RPC / infra providers
- Key rotation cadence

Reputational
10%

- Past incidents reaction
- Founder track record
- Social authenticity
- Partner credibility

Compliance
5%

- Jurisdiction
- License posture
- Disclosure practices
- AML / KYC framework

03

Probability of Loss scoring and weights.

Probability of Loss (PoL) is the likelihood that an adverse event will occur at a digital-asset issuer or exchange and cause users, counterparties, or the project itself to incur financial losses outside of organic price movement. The estimate is derived from the issuer's publicly disclosed control posture, calibrated against the historical record of documented failures, and expressed on a 0-to-100 scale.

Project risk exposure calculation score

=

$$(0.35S + 0.20O + 0.15F + 0.15D + 0.10R + 0.05C) * \text{the Scale multiplier}$$

S - Security O - Operational F - Financial D - Dependency R - Reputational C - Compliance

Probability of Loss = Inverse [Calculation score]

Weights justification.

Security parameters are proven to be effective at detecting and mitigating the vulnerabilities that are most commonly exploited.

Operational domain exposes structured controls over cryptographic asset custody, key management, and broader information security under an auditable management framework.

Finance assesses economic models driving both gradual value erosion and event-driven liquidity failures, severe but slower to trigger than a direct exploit.

Dependency captures systemic and cascading failure scenarios, applying conditionally to a project's architecture and third-party integration solutions rather than uniformly across all issuers.

Reputational domain evaluates how the project has responded to past incidents, the credibility of its founders, and whether its track record and community presence reflect a project built to last.

Compliance carries the least weight, as non-compliance rarely triggers an immediate loss, functioning instead as a structural signal that a project may be designed to evade accountability.

Score bands.

0 to 20	20 to 40	40 to 60	60 to 99
Exceptional	High	Moderate	Low
AAA AA A	BBB BB B	CCC CC C	DDD DD D
Demonstrate strong fundamentals across assessed dimensions, with minimal Probability of Loss.	Projects show solid performance with manageable risk exposure. While some weaknesses may exist in certain areas, the overall profile remains sound.	Projects exhibit a mixed risk profile with notable gaps or uncertainties in key assessed areas.	Projects present significant red flags or critical lack of disclosed information, with a high probability of loss.

The composite index delivers risk exposure as a single, interpretable figure. By design, this composite score is lossy; consequently, two issuers with identical Probabilities of Loss scores may exhibit materially different risk postures. To address this, every composite risk index is published alongside the six domain scores and individual parameter outputs. This multi-layered approach mirrors the reporting standards of major credit rating agencies, providing a headline grade at the apex supported by a comprehensive analytical breakdown underneath.

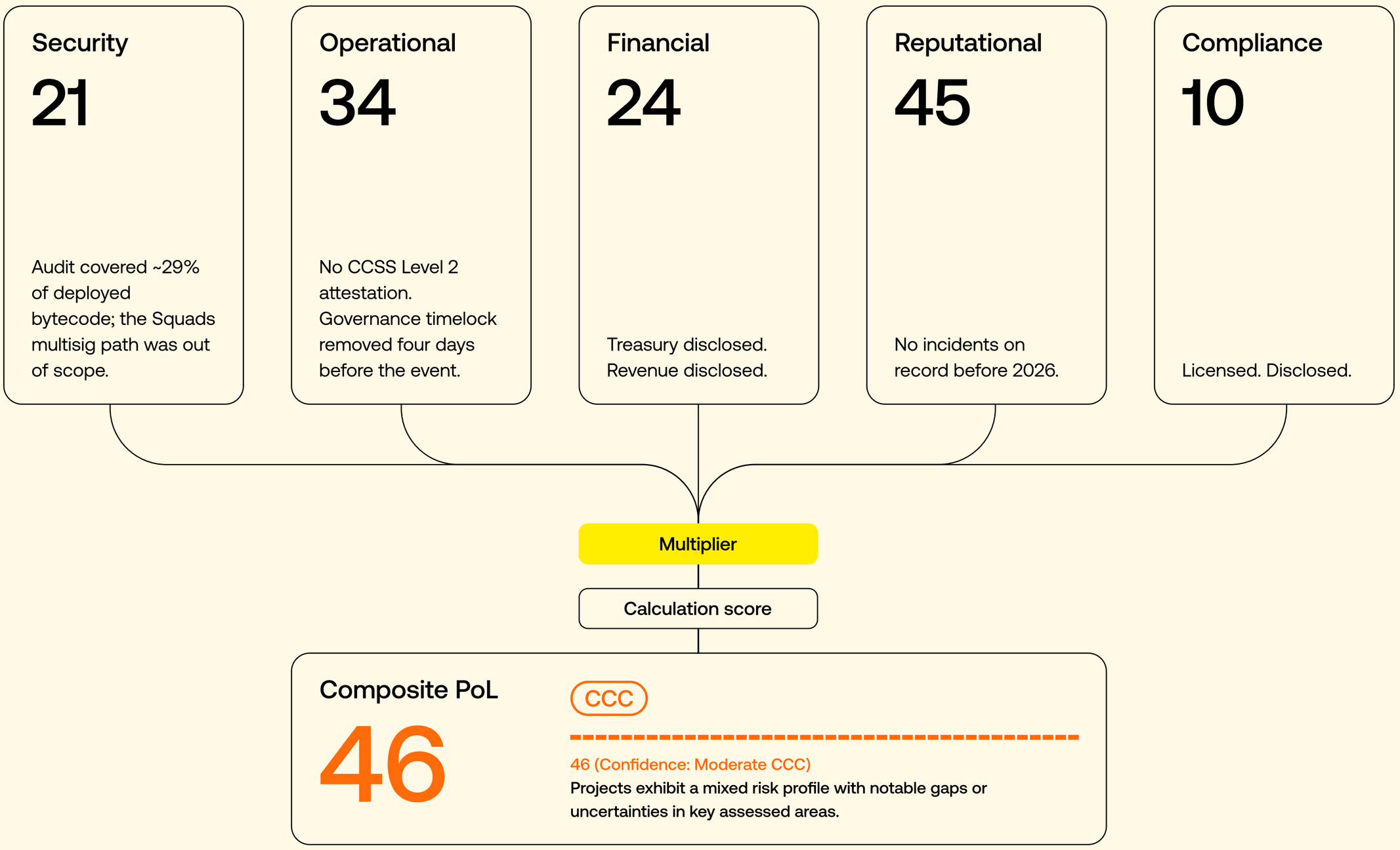
04

Probability of Loss showcases & validation

This section validates the methodology against adverse events occurring after the April 2, 2026, MVP release. Because the methodology assesses disclosed risk posture rather than predicting specific exploit vectors or timing, the validation test measures whether the model accurately identified the exposed attack surface, not the exact moment of failure.

Case #1 Case #2

Drift Protocol, 1 April 2026. $285M drained.



The category-level PoL breakdown identified two specific operational vulnerabilities. First, the project held no standard security certifications, which would have enforced device verification and personnel verification for machines holding signing keys. Second, the governance architecture lacked time locks on administrative actions, allowing any two of the five multisig keys to instantly modify the protocol.

Both gaps were visible in the public record. Over a six-month period, adversaries compromised the team’s devices to secure two signing keys. On April 1, they executed an administrative takeover, manipulated the protocol's price feed, and drained $285 million within twelve minutes.

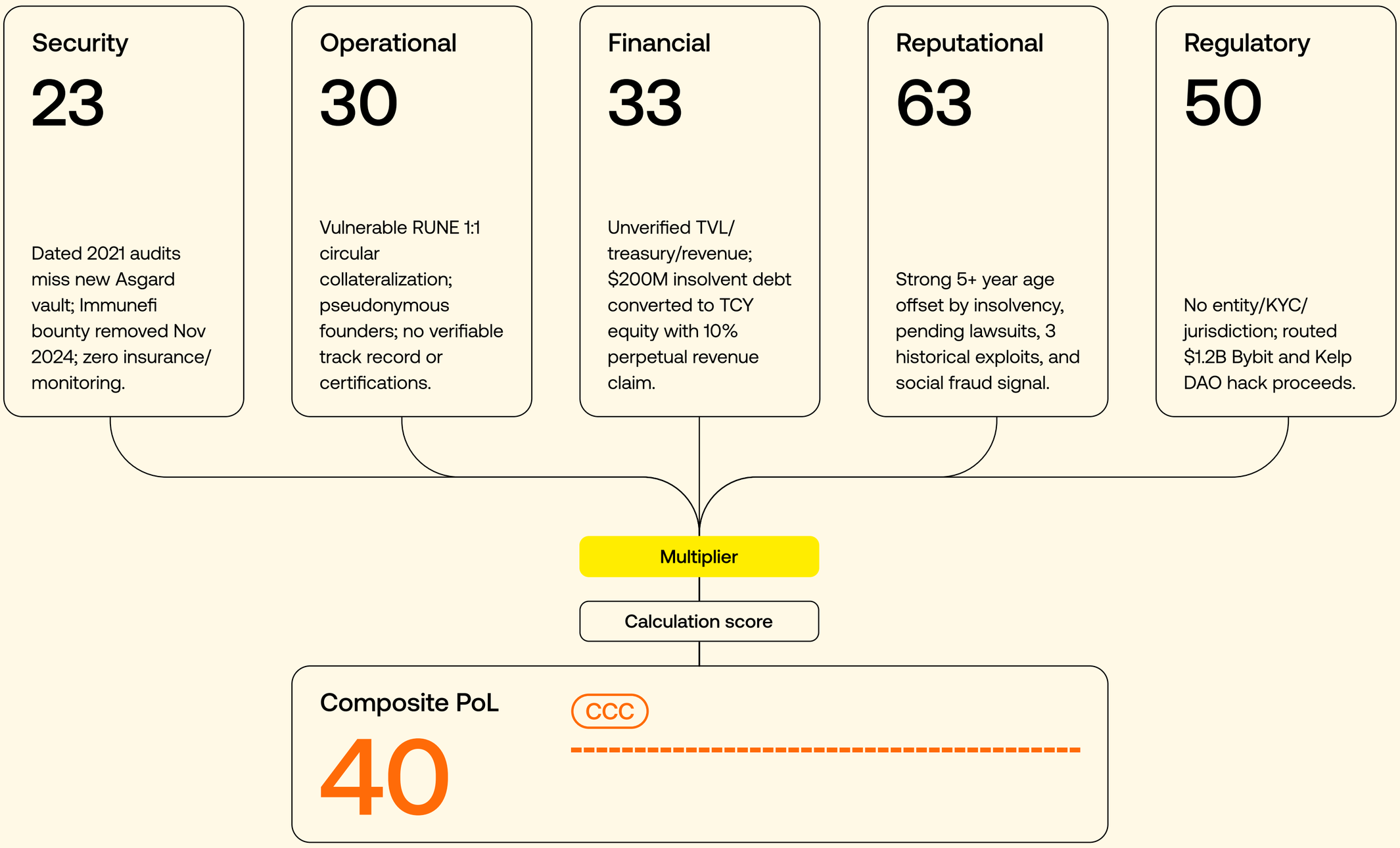
The exploit directly targeted the specific openings identified by the PoL. While the methodology did not predict the timing or specific social-engineering vector, it accurately recorded the absent defenses through which the loss occurred.

04

Methodology tested in the wild

Case #1 **Case #2**

THORChain, 15 May 2026. ~$10.8M drained from an Asgard vault.



The category-level PoL breakdown identified THORChain's vault infrastructure as highly exposed across several security controls. Recent audits failed to cover the vaults or the key-coordination layer securing them. Furthermore, the bug bounty program had been removed eighteen months prior, eliminating any structured channel or incentive for researchers to report vault-level flaws. Concurrently, the absence of vault outflow monitoring or rate-limiting caps meant an active drain would face no automated response.

On May 15, an adversary compromised one of the protocol's six vaults and extracted approximately $10.8 million across four blockchains. Due to the lack of automated outflow monitoring, the funds moved for thirteen hours before governance manually halted the protocol.

This marked the third on-chain exploit of THORChain's cross-chain vault infrastructure in five years, directly occurring on the surface that the model had flagged. Just like with Drift, methodology did not predict the timing or the method of compromise; instead, it identified the missing defenses that enabled the drain.

What the two cases together demonstrate

In both cases, the failure surface was visible in public data prior to the event, differing only in the specific domain where the risk signal originated. The analysis directly confirms that a lack of public evidence signaled a total absence of operational controls. Had these defenses existed privately, the exploits would have been obstructed.

Instead, both attacks progressed unimpeded through the exact gaps flagged by the model, proving that the structural weaknesses visible from the outside matched the actual operational deficiencies on the inside. Ultimately, the domains exhibiting the highest weighted gaps experienced exploits directly on the surfaces flagged by the methodology.

06

Limits

01

PoL does not predict timing. The methodology produces a probability surface, where a project at 80 PoL carries observable failure conditions that historical evidence correlates with eventual loss. It is not "about to fail" in the trader sense. Yet, time is included in PoL benchmark as a factor, that shows how project teams improve or deteriorate their risk posture over time, allowing any market participant to see the momentum the project carries in terms of building secure product.

02

PoL does not subsume Value-at-Risk. The two answer different questions. VaR estimates how much capital a portfolio could lose under a defined statistical distribution. PoL estimates the probability that a project carries the structural conditions associated with capital-loss events. Allocators using both will need to run them in parallel rather than substituting one for the other.

03

PoL reads public data by design. Where risk management evidence cannot be published, projects may submit it through the disclosure channel to improve their visible Probability of Loss. Where no evidence exists or is submitted, this methodology treats a risk parameter as absent and contributing to worsening the risk index. This way, CORE3 team prompts all digital asset issuers that stay in the grey, to get back into the crypto ethos, where trustless evidence is the only medium of truth.

05

PoL as the agentic risk layer for crypto

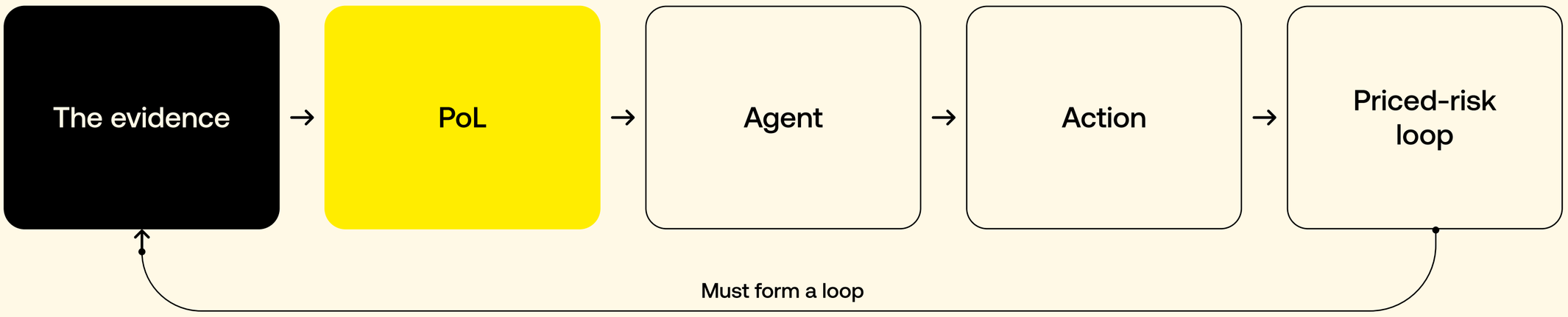
Forecasts hold that AI agents will soon move funds on-chain at scale, already outnumbering human workers in financial services by roughly 100 to 1.1 Each remains bound to an accountable principal: the agent executes, a human answers for it. On a risk, listings, or DD desk, that human is the reader of this document.

The Probability of Loss is constructed to be the risk layer those agents read. Its determinism is what qualifies it for the role: an agent that cites PoL produces reasoning a third party can reproduce, the precondition for any automated decision, a due diligence must later back-test.

Its evidence anchoring is what makes the output trustworthy at machine speed: every parameter resolves to a verifiable public artifact, which closes the hallucination surface that dominates model-generated crypto research.

And because the score is delivered in callable form, one signal serves both readers: the dashboard a person reads and the pre-trade check an autonomous agent runs are the same number, not two.

The result is a closed loop. Public evidence resolves into a PoL score; an agent reads the score and acts; the action prices the project's risk into the capital that reaches it. Risk stops being assessed after a failure and starts being priced before one, by the systems doing the routing.



This is the same wiring credit-rating data has in traditional finance, where a rating is not advisory but an input every underwriting system clears before capital moves. Agent-readable risk data takes that position in the systems that increasingly route crypto capital on a user's behalf. Once the risk layer is machine-readable, a project's disclosed posture becomes a direct constraint on what an agent will recommend or execute against it: the score is no longer a report a project can ignore, it is a parameter in someone else's automated decision.

07

Conclusion

CORE3 assesses the risk posture of Web3 issuers and exchanges from public evidence alone, on a standard applied the same way to every counterparty. The benchmark is built for the principal accountable for the decision and read in parallel by the systems that increasingly act on that principal's behalf, so one bar holds whether a person or a machine runs the first pass.

Since 2011 the industry has run a single cycle: each failure teaches a lesson the next failure forgets. That cycle ends when risk is priced before failure occurs rather than catalogued after, and a standard legible to both human reviewers and the systems routing capital is what lets the market hold that price.

08

Disclaimer

CORE3 is an independent analytics platform offering a data-driven Probability of Loss framework to quantify risk in Web3 projects. The platform is not a ratings agency, and its metrics do not constitute investment advice. Low probability of loss doesn’t mean a project is certified, approved, or risk-free.

The following project breakdown is based on processed information obtained through research relying primarily on publicly available data. As a result, some of the assessed projects may have mitigation measures or controls in place that are not visible to us.

core3.io → docs.core3.io → github → X →

CORE3 is an independent analytics platform. The Probability of Loss metric does not constitute financial advice.