Cleafy | NYX

Closing the Investigation Gap

Using AI Agents to Fight AI-Powered Fraud

POWERED BY

Cleafy

Autonomous Cyber Fraud Defence

AI-powered fraud is not only changing the external threat landscape. It is exposing an internal limitation in the operating model of many banks.

The bank sees more…

Detection has improved materially — more signals, more alerts, more visibility.

…but does not understand more.

Reconstruction, attribution, and meaning have not kept pace with detection.

The decisive weakness has moved.

From the detection layer to the gap between detection and understanding.

From the threat, to the bottleneck, to the response.

I

How AI Reshapes Fraud

Not new motives — new economics, tempo, and operating logic. Trust itself becomes an attack surface.

II

The Investigation Gap

Why detection can improve while operational control still erodes — the new bottleneck in fraud defence.

III

Using AI to Fight AI

A disciplined architectural shift — autonomous autonomous investigation, proportional prevention, human orchestration.

PART ONE

I

How AI Reshapes Fraud

Not by inventing new crimes, but by changing the economics, tempo, and operating and operating logic of every familiar one.

Economics, Not Invention

AI does not invent fraud. It removes the constraints that previously limited it.

WHAT STAYS THE SAME

A scam is still a scam.

An account takeover is still an account takeover.

Greed, coercion, deception, mule recruitment — the criminal motives are old.

WHAT CHANGES

Skill barriers **collapse**

Time between attack stages **compresses**

Scale and personalisation **now coexist**

Campaigns are **industrial in reach, intimate in tone**

From Episodes to Operations

Modern fraud is no longer an event. It is a workflow run by adaptive operators.

Each stage looks ambiguous in isolation. **Only when stages connect does the institution see the campaign.**

Opportunism with Structure

Attackers do not strike randomly. They follow asymmetry — toward institutions whose resilience is incomplete.

Attackers do not seek the most visible target. They seek the target whose controls reveal themselves under probing.

WHERE ATTACKERS GRAVITATE

Abundant trust, fast digital payments, deeply embedded habits, response slower than attack iteration.

WHY RECONNAISSANCE MATTERS

It is not prelude. It is where the economics of the attack are determined.

THE SHIFT-LEFT RATIONALE

Disrupt early to collapse the campaign's economic value before losses materialise.

Trust as an Attack Surface

AI erodes the cues people use to recognise legitimacy — voice, language, urgency, channel coherence.

OLD PROXIES FOR LEGITIMACY

Familiar voice

Tone, cadence, accent

Quality of language

Grammar, register, fluency

Coherence

Narrative consistency

Channel continuity

Same place, same person

WHAT AI NOW GENERATES

Persuasive language at **local fluency**

Cloned voice and tuned **emotional resonance**

Scripts that **adapt in real time**

Scenarios that feel **situationally credible**

The fraud has occurred earlier than the transaction.

It has occurred **in the shaping of intent.**

Control, Not Anomaly

When fraud hides inside valid sessions, anomaly is not enough. The right question is who is in control.

/ 01 · Session

Session control - ATO

Compromised when an attacker hijacks or assumes possession of an authenticated interaction.

/ 02 · Device

Device control - DTO

Compromised when malware, RATs, overlay abuse, accessibility, or relay techniques mediate the user.

/ 03 · Intent

Intent control - SCAM

Compromised when the user is present, authenticated, authenticated, and human — yet manipulated into into authorising what they would not have chosen. chosen.

Modern fraud lives **between events** — cross-channel, longitudinal, relational. Across the space between functions.

The Digitally Mature Threat Surface

In markets where digital trust runs deep, attackers no longer break security — they hijack normality.

Pattern 01

Fast-payment social engineering

Manipulation logic that attaches to instant-payment habits and peer-to-peer trust dynamics.

Pattern 02

Mule-account orchestration

Monetisation as a coordinated network — not an afterthought to the attack.

Pattern 03

Device-mediated fraud

NFC relay, contactless abuse, overlay attacks where mobile is deeply embedded.

Pattern 04

Scam × mocking behaviour

High digital confidence creates the very behavioural surface modern fraud imitates.

The real target of modern fraud is not only the customer. It is often the bank's operations.

THEY BENEFIT WHEN…

Alert volumes exceed review capacity

THEY BENEFIT WHEN…

Fraud, cyber, identity remain siloed

THEY BENEFIT WHEN…

Signals are visible but not connected

THEY BENEFIT WHEN…

Analysts spend time clearing queues, not finding campaigns

Detection without explanation is too easily absorbed into delay. Delay is the attacker's ally.

PART TWO

II

The Investigation Gap

Detection vs. Understanding

More signals do not produce proportionate control. The relationship between seeing and understanding is not linear.

Defenders do not keep up with attackers

A structural mismatch — not an analyst-competence problem.

ATTACKER PACE

- Automated message generation
- Continuous infrastructure testing
- Behavioural variation at scale
- Compressed compromise → monetisation
- Learning bounded only by compute

MACHINE_SPEED

VS.

DEFENDER PACE

- Alerts routed into human review queues
- Data stitched manually across tools
- Session, device, transaction inspected separately
- Cross-case linkage rarely made
- Peaks in alert volume become peaks in blindness

QUEUE_SPEED

Why Traditional Tools Fall Short

The insufficiency is structural, not parametric. Tuning thresholds will not resolve it.

/ Transaction monitoring

Sees the attacks too late

Necessary, but cannot be the main locus of control.

/ Behavioural analytics

Cannot decode mediated activity

Behaviour alone won't separate coerced or hijacked sessions.

/ Device intelligence

Sees signals, not stages

Won't say whether fraud is being prepared, attempted, or executed.

/ Cyber tooling

Compromise without context

Doesn't connect compromise to customer-level monetisation.

/ Case management

Organises human work — does not create machine-speed correlation

The architecture still waits for events, separates data across functions, and depends on humans for correlation.

The New Bottleneck Is Interpretation

THE OLD QUESTION

"How can we detect more?"

→

THE NEW QUESTION

"How can we understand sooner?"

The bank that interprets attack progression quickly enough **retains strategic initiative.**

The bank that sees only disconnected suspicious events becomes reactively informed, but not in control.

A Governance Problem, Too

Without understanding, response is either delayed, excessive, or incomplete — and proportionality matters in banking.

RISK: TOO AGGRESSIVE

Block legitimate customers

Friction, complaints, damaged trust.

RISK: TOO CAUTIOUS

Permit fraud, delay containment

Loss, reimbursement pressure, reputational damage.

A bank cannot operate on opaque suspicion alone.

Decisions affect real customers in sensitive moments.

The investigation gap is not closed by faster scoring.

It is closed by faster, better-grounded **understanding** able to support both **action and accountability.**

Analysts Shift From Case Reviewers to Control Strategists

The shift is performed by the humans themselves — analysts move upward in value, away from mechanical correlation toward judgment, governance, and orchestration.

TODAY

Case Reviewer

- Faster reviewer of repetitive cases
- Manual correlator under pressure
- Time consumed clearing queues
- Cross-case linkage rare

TOMORROW

Control Strategist

- Defines **response posture**
- Tunes **escalation logic**
- Evaluates **ambiguous edge cases**
- Shapes **customer-protection policy**
- Continuously improves **risk stance**

PART THREE

III

Using AI to Fight AI

The Wrong Answer: More AI, Same Architecture

If the architecture stays transaction-centric, reactive, siloed, and queue-based — another model will not close the gap.

The right question is not whether to use AI.
It is where and how AI must be applied to make the operating model more efficient.

Many teams think: “If we deploy AI, we are more advanced”. **That is false.**

AI defends correctly only when it is embedded in a system that sees control loss early, understands context, adapts continuously, and responds proportionally.

AI ON FRAGMENTED DATA

- Late signals — arrives after fraud succeeds
- Outcome-based — trained on what already failed
- Isolated events — no cross-channel context
- Static models — periodic retraining

VS.

AI ON CONNECTED INTELLIGENCE

- Real-time signals — before monetisation
- Control-based — who is in control of this interaction?
- Correlated context — relationships across signal planes
- Adaptive learning — updates continuously as attackers do

AI Must Be Explainable at Decision Time

Many organisations deploy powerful models that produce a risk score but cannot explain why.
In cyber fraud that is dangerous because users are often legitimate victims, decisions trigger friction or blocks, and regulators expect accountability.

SCORE-ONLY AI

A number without a narrative

- Risk score with no observable cause
- Cannot justify a block to a legitimate customer
- Cannot withstand regulatory or audit scrutiny
- Forces binary allow / block decisions

VS.

AI USED CORRECTLY

Evidence tied to observable causes

- Produces **explainable signals**
- Ties decisions to **device compromise, session hijack, coercion**
- Enables **proportional responses** — not binary allow/block
- Stands up to challenge by analysts and regulators

AI Must Be Adaptive, Not Periodically Retrained

AI-enabled attackers adapt continuously — they probe thresholds, learn what passes controls, and adjust behaviour in real time. Defensive AI that retrains monthly will always lag behind.

STATIC AI — THE PROBLEM

- Retrains monthly or quarterly
- Relies on static features and fixed thresholds
- Learns only from **labelled past cases**
- Always lags behind attacker iteration speed

VS.

ADAPTIVE AI — THE SOLUTION

- Learns from **streaming data**
- Updates risk **continuously during a session**
- Detects **emerging patterns before labels exist**
- Closes the gap between attacker and defender tempo

AI Must Be Applied to Control Signals, Not Just Outcomes

Fraud lives in the relationships, not in the event. Most AI today learns what fraud looked like after it succeeded . Defensive AI must learn what control loss looks like before it does.

MOST AI TODAY

Trained on outcomes. Teaches the model what fraud looked like **after it succeeded**.

- Past fraudulent transactions
- Chargebacks
- Known bad events

VS.

AI USED CORRECTLY

Trained on signals that precede fraud. Detects the **loss of control** in the moments before monetisation.

- Loss of **session continuity**
- **Device integrity** degradation
- **Behavioural distortion**
- **Coercion** and manipulation indicators

Six Architectural Shifts

The minimum viable operating model for serious fraud defence.

/ 01

Detection moves upstream

Reconnaissance, probing, scripted onboarding, infrastructure reuse — before monetisation.

/ 02

Detection becomes control-based

From "does this look bad?" to "what is happening to agency, authority, and execution?"

/ 03

Correlation across the whole narrative

Network, device, application, behaviour, transaction — reconstructed reconstructed at machine speed.

/ 04

Prevention becomes proportional

Step-up, cooling-off, containment, propagation — not allow vs. block. vs. block.

/ 05

Detect once, protect everywhere

One confirmed pattern updates posture across channels, accounts, accounts, journeys.

/ 06

Continuous review & proactive hunting

Surface campaigns, infrastructure, and risk drift before a discrete event discrete event triggers an alert.

Human-in-the-Loop Orchestration

Explainability is not optional. **Outputs must be evidence-ready, cross-correlated, and transparent** — fluency is not enough.

THE MACHINE ABSORBS

- Scale
- Continuity
- Speed
- Mechanical correlation

THE HUMAN RETAINS

- Intent & risk appetite
- Governance
- Edge-case judgment
- Accountability

PART III · CASE STUDY

Cleafy Nyx

What changes from the legacy workflows

LEGACY FRAUD WORKFLOW

- Alerts accumulate, analysts triage and sample
- Data stitched manually across systems
- ~30% of cases abandoned for lack of time
- Cross-case patterns rarely surfaced
- 4–5 hours per investigation
- False-positive rate often above 80%

VS.

NEW OPERATING MODEL

- Continuous correlation across device, behaviour, network, application, transaction
- Structured, evidence-ready case output
- 100% case coverage — not sampled review
- Attack-pattern recognition, not opaque scoring
- Investigations under 5 minutes — ~77× faster
- Decoupled growth: defence scales without linear headcount

What an Agentic AI solution can do

Each validated understanding becomes a protective strategy. The bank learns at the pace of the attacker, creating a defence that scales along with attack volume, without scaling headcount linearly.

/ 01

Reconstruct

A full digital narrative across users, devices, sessions, and channels — not isolated alerts.

/ 02

Correlate

Multiple signal planes resolved into one coherent, evidence-ready case at machine speed.

/ 03

Protect

One confirmed pattern propagated propagated automatically across every every related channel, account, and and journey.

/ 04

Govern

Humans remain on the loop — setting strategy, strategy, overseeing outcomes, challenging anomalies.

Cleafy

NYX Platform

- Overview

THREAT MANAGEMENT

- Cases
- Hunts

AUTONOMOUS AI

- Executions
- Automate
- Agents

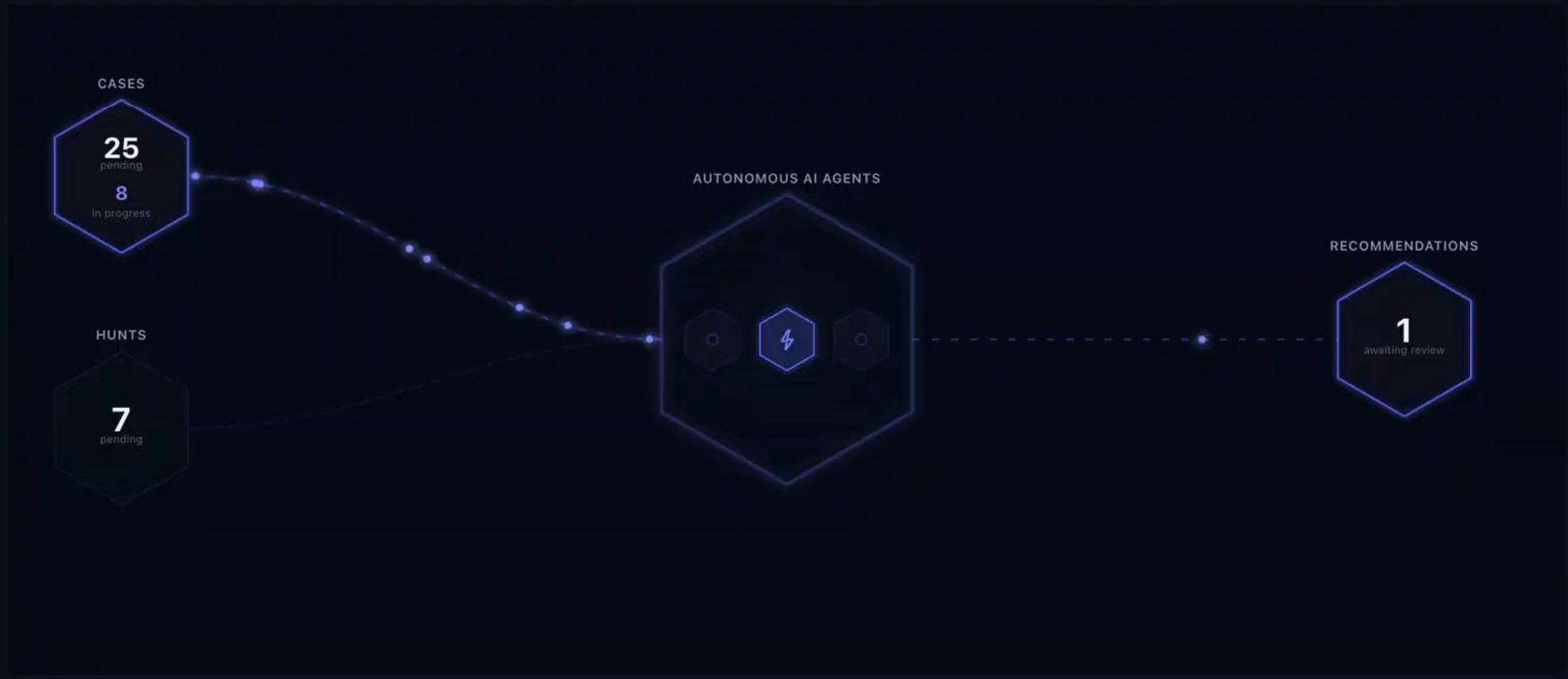
RESULTS

- Recommendations

AU Admin User

Overview

Real-time view of autonomous agent operations



SUCCESS RATE	AVG DURATION	QUEUE DEPTH	ACTIVE AGENTS
84%	2m 43s	4	3/3
0 completed today	per execution	pending executions	1 currently working

The Strategic Meaning of "AI vs. AI"

Using AI to fight AI means employing machine intelligence not just to score risk, but to restore the bank's capacity to interpret what is happening before delay becomes loss.

SYMMETRY

See how control is changing across the interaction.

SYMMETRY

Link the case to the campaign.

SYMMETRY

Propagate protection across the institution.

SYMMETRY

Preserve human expertise where judgment matters most.

IN CLOSING

The banks that succeed will not be those that merely detect more.

They will be those that learn how to understand sooner, intervene earlier, and place human intelligence where it matters most.

Thank you for your attention!

LETS CONNECT!

Cristina Bottoni

Cyber Fraud Strategist

Cleafy

Autonomous Cyber Fraud Defence