LOCI

WHITE PAPER · V1 · JANUARY 2026

The Fraud Language Model (FLM)

Explainable, real-time fraud defense for the AI era. How specifications that function as risk models deliver transparency and rapid adaptation, without choosing between them.

EXECUTIVE SUMMARY

Fraud defense has to show its work

The democratization of large language models has fundamentally altered the fraud landscape. Attackers leverage AI to generate synthetic identities, automate social engineering, and rapidly evolve evasion strategies. Traditional defenses face a structural mismatch: rigid rules can't adapt quickly enough, while black-box ML systems struggle to meet transparency requirements for regulatory compliance and operational trust.

This white paper introduces the **Fraud Language Model (FLM)**: an integrated approach that combines domain-constrained AI, structured policy representation, and deterministic execution. FLM enables fraud teams to express detection intent in natural language while producing fully auditable, high-performance risk specifications.

> THE CORE CLAIM
>
> FLM is not a compromise between rules and ML. It represents a distinct approach where specifications function as risk models, weighted, composable, and calibratable, while remaining transparent enough for regulatory scrutiny and operationally trustworthy for front-line analysts.

CONTENTS

What's inside

CHAPTER 01

The Inflection Point

The problem

Financial fraud is evolving faster than defenses can adapt. The attack surface expands with every new payment rail, digital channel, and customer touchpoint.

$485B
global fraud losses in 2023, per ACAMS estimates

<1s
decision windows on instant payment rails, where funds move irrevocably

95%+
false positive rates common in AML transaction-monitoring alert queues

The pain

Financial institutions face compounding pressures:

Alert fatigue. False positive rates in AML transaction-monitoring alert queues commonly exceed 95%, burying genuine threats in noise.

Compliance burden. Regulators demand decision explanations that many systems cannot provide in operationally useful form.

Talent scarcity. Data scientists command premium salaries; specialized rule engineers are increasingly rare.

Speed mismatch. Rule deployment cycles are measured in weeks; attackers iterate in hours.

What changed

Three shifts have created an inflection point:

AI-equipped adversaries. The same large language models powering legitimate innovation now generate convincing phishing campaigns, synthetic identity documents, and automated reconnaissance. Fraud-as-a-service operations leverage AI to scale attacks that once required human expertise.

Real-time payment rails. Instant payment systems, from Nigeria's NIP to the US FedNow and Europe's SEPA Instant, demand sub-second decisions. There is no "review queue" when funds move irrevocably in milliseconds.

Regulatory convergence. Despite jurisdictional differences, global regulators are converging on a common principle: automated decisions affecting individuals must be explainable in operationally meaningful ways. The EU AI Act, US model risk guidance, and African central bank frameworks all point toward transparency as a baseline expectation.

> THE GAP
>
> The defenders' toolkit has not kept pace.

CHAPTER 02

Why Current Approaches Fall Short

Traditional rules: transparent but brittle

Rule-based systems offer clarity. An analyst can read the logic, understand the trigger, and explain the decision. This transparency comes at a cost:

Binary thinking. Rules fire or they don't. A transaction is either blocked or approved. There is no nuance, no graduated risk assessment.

Threshold fragility. A velocity rule checking "more than 5 transactions in 10 minutes" catches the attacker who sends 6, and misses the one who sends 5.

Slow adaptation. Each new fraud pattern requires manual rule creation, testing, and deployment. By the time a rule is live, attackers have moved on.

Combinatorial explosion. Addressing pattern variations means more rules, more conflicts, more maintenance burden.

Rules work for known, stable fraud patterns. They struggle against adaptive adversaries.

Machine learning: adaptive but operationally opaque

ML models learn patterns from data, adapting to signals humans might miss. This power introduces different challenges:

Operational opacity. When a model flags a transaction, the analyst triaging the alert needs to understand *why* in terms they can act on. Technical explainability tools like SHAP values and LIME have advanced significantly; they can identify which features contributed to a prediction. But feature importance rankings ("velocity_score contributed 0.3") differ from operational explanations ("flagged because this customer made 7 transfers to new beneficiaries in 20 minutes"). The gap between technical explainability and operational explainability remains.

Trust deficit. When analysts don't understand why a model flags something in actionable terms, alert fatigue compounds. The issue isn't just volume. It's confidence.

Data requirements. ML requires substantial labeled datasets. Fraud is rare by definition; confirmed fraud rarer still. Models trained on thin data overfit or underperform.

Cold start. New customers have no behavioral history. Models optimized for established patterns struggle on day-one accounts, precisely where fraud risk often concentrates. FLM faces similar constraints when customer history is sparse. The difference is that specifications can define explicit fallback logic for thin-file scenarios, such as elevated scrutiny on device signals, network indicators, or identity verification, rather than relying on implicit model handling of missing features.

Drift. Fraud patterns evolve. Models trained on last year's attacks degrade against this year's tactics. Retraining cycles lag the adversary's innovation.

Adversarial gaming. Sophisticated attackers probe model boundaries, engineering transactions that score just below detection thresholds.

The false dichotomy

> REFRAME
>
> The industry has often framed the choice as rules versus ML, transparency versus adaptability. This framing obscures a different question: whether an architecture can deliver operational transparency *and* rapid adaptation.

CHAPTER 03

Introducing the Fraud Language Model

FLM is an integrated approach purpose-built for fraud defense. It combines three capabilities:

Domain-constrained AI. Fraud analysts express detection intent in natural language. But unlike generic text-to-code tools, FLM's AI layer operates within rigorous domain constraints: a formal vocabulary of detection primitives, compositional rules for building valid specifications, and semantic understanding of fraud patterns. The system reasons about intent and constructs structured specifications, not merely translates text.

Structured policy representation. Natural language compiles into a formal specification, human-readable and machine-executable. This specification is versioned, auditable, and portable across environments.

Deterministic execution. The same input always produces the same output. Decisions are reproducible, explainable, and suitable for regulatory review. The AI assists specification authoring; it does not make runtime decisions.

The name explained

"Fraud Language Model" reflects three aspects:

- **Language Model:** AI that interprets natural language fraud scenarios within domain constraints
- **Fraud Language:** a domain-specific language for expressing fraud detection logic
- **Model:** specifications that model fraud behavior through weighted signal composition

> IN ONE LINE
>
> The AI interprets intent. The language structures it. The specification models risk. Each layer is distinct; together they form a coherent system.

What FLM is

FLM is an autonomous risk engineering system. When an analyst describes a fraud scenario, the system:

- Interprets the semantic intent (not just keywords)
- Reasons about which signals matter and their relative importance
- Constructs a valid specification within formal constraints
- Validates the result against schema requirements
- Produces an auditable artifact ready for testing and deployment

This goes beyond "assisted authoring." The system performs the engineering work, within defined guardrails that ensure specifications are structurally sound and semantically coherent.

Addressing the rules critique

FLM builds on rule-based foundations, which invites skepticism. Traditional rules carry known limitations:

"Rules are static and reactive."

Traditional rules are static because updating them requires development cycles. FLM specifications deploy rapidly. When a new pattern emerges, an analyst describes it, the system generates a validated specification, and shadow mode testing begins within minutes. The bottleneck shifts from engineering capacity to analyst insight.

"Rules don't scale. Managing thousands becomes unmanageable."

Rule sprawl occurs when each variation requires a separate rule. FLM's weighted approach consolidates variations into single specifications with configurable sensitivity. Fewer specifications, broader coverage.

"Rules generate too many false positives."

Binary rules force a choice: trigger or don't. FLM's weighted approach means partial signal matches register as elevated risk without necessarily triggering immediate action. Tuning happens at the threshold level, not by rewriting logic.

"Rules can't detect unknown fraud."

No system detects the truly unknown. But FLM reduces the unknown faster. Near-miss analysis surfaces emerging patterns. Integration with pattern discovery systems feeds novel signals into specifications. The gap between "pattern observed" and "detection deployed" compresses significantly.

FLM doesn't pretend rule-based approaches have no limitations. It architects around them.

CHAPTER 04

Architecture

The pipeline

FLM's architecture flows through distinct layers:

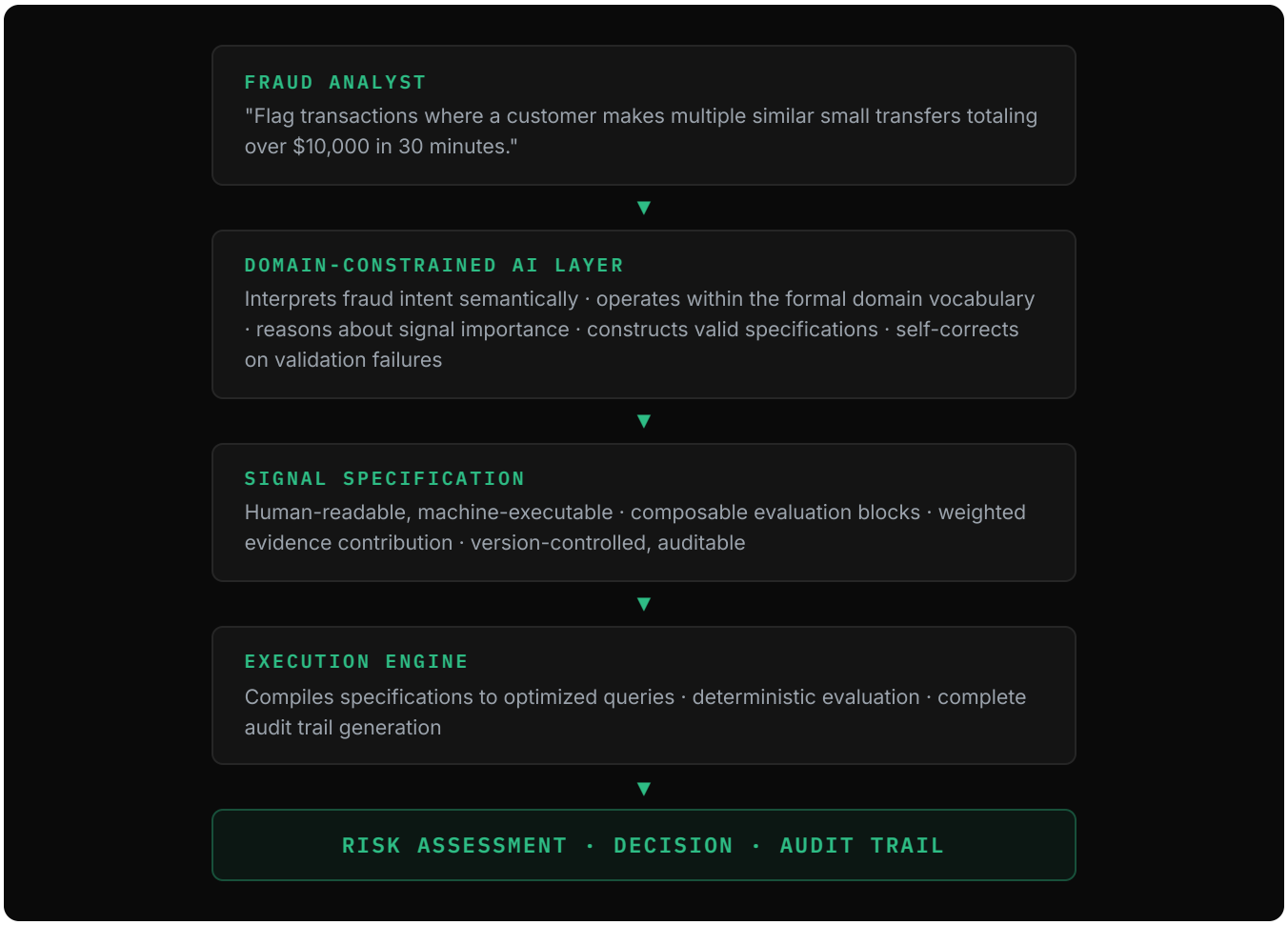

Fig. 1 · The FLM pipeline: AI authors, specifications define, execution decides

Separation of concerns

The architecture enforces strict boundaries:

LAYER	ROLE	CHARACTERISTICS
AI Layer	Interpret intent, construct specifications	Domain-constrained, semantic reasoning
Specification	Define detection logic	Deterministic, versioned, auditable
Execution	Evaluate transactions	High-performance, reproducible

> **WHY THIS MATTERS**
>
> The AI layer's role is confined to specification authoring. By the time a specification reaches execution, behavior is fully deterministic. This separation is what enables auditability: the specification, not the AI, defines runtime behavior.

Input flexibility

FLM accepts detection intent from multiple sources:

Analyst scenarios. Natural language descriptions of fraud patterns.

Pattern discovery systems. Structured outputs from statistical mining tools that identify anomalies in transaction data.

Visual editors. Direct manipulation through graphical interfaces.

ML model outputs. Scores and classifications from existing machine learning systems.

This last point deserves emphasis. FLM does not require abandoning existing ML investments. Model scores can flow into FLM as data fields, another signal to evaluate alongside velocity checks and amount thresholds. A specification might weight "ML risk score above threshold" as one evaluation among several.

This hybrid approach captures ML's pattern recognition while wrapping decisions in operationally explainable logic. When a transaction triggers, the audit trail shows which signals contributed, including any ML scores, rather than presenting a single opaque prediction.

The signal specification serves as the common interchange format, regardless of origin.

CHAPTER 05

The Signal Specification Language

Design principles

The signal specification language is designed for:

- **Human readability.** Analysts and auditors can understand detection logic without specialized training.
- **Machine precision.** Unambiguous semantics enable correct compilation and execution.
- **Composability.** Complex patterns built from simple, reusable primitives.
- **Auditability.** Every element has explicit naming and documentation.

Core structure

A signal specification contains:

Identity. Name, description, and version information.

Evaluations. Independent detection blocks, each with a weight and a sequence of operations. Each evaluation resolves to a boolean result and contributes its weight to the overall risk assessment.

Threshold. The risk level required to trigger action.

Actions. What happens when the threshold is met.

Evaluation independence

Each evaluation block operates independently. Evaluations cannot reference each other; this constraint ensures they can be reasoned about in isolation and executed efficiently. The weighted aggregation happens at the specification level, not through cross-evaluation dependencies.

Within an evaluation, operations can reference prior operation results, enabling compositional logic: compute an aggregate, compare the current value against it, combine results logically. This composition happens within a single evaluation's scope.

Schema validation

Every specification is validated against a formal schema before deployment. This catches structural errors at authoring time: required fields present and correctly typed, references resolving to valid prior operations, no circular dependencies within evaluations, and semantic constraints satisfied.

> HONEST BOUNDARY
>
> The validated deployment path rejects invalid specifications. But schema validation ensures structural correctness, not semantic correctness: whether the specification actually detects what the analyst intended. Shadow mode testing against real traffic validates semantic correctness.

CHAPTER 06

Rules as Models: Weighted Risk Assessment

Beyond binary triggers

Traditional rules produce binary outputs: triggered or not. A transaction either crosses a threshold or it doesn't. This forces fraud teams into uncomfortable tradeoffs: set thresholds too tight and miss subtle patterns; set them too loose and drown in false positives.

FLM takes a different approach. Evaluations contribute weighted evidence to an overall risk assessment. Strong indicators carry more weight than weak ones. The final determination considers the totality of evidence, not just whether any single condition fired.

Why "Rules as Models"

The term "model" carries specific meaning. In the ML context, it typically refers to learned parameters from data via optimization. But "model" has a broader engineering meaning: a structured representation of a system or phenomenon.

FLM specifications are models in this engineering sense. They *model fraud behavior* through weighted signal composition (compositional evaluation blocks, not flat feature weights), continuous risk scores rather than binary triggers, calibratable weights and thresholds, and structured representations that can be analyzed, compared, and versioned.

Traditional rule: `IF velocity > 5 THEN block`. This is logic, not a model.

FLM specification: multiple weighted evaluations, an aggregated risk assessment, a threshold determination. This models risk through structured signal composition.

The distinction matters. FLM specifications don't learn from data the way neural networks do. But they function as risk models: structured, calibratable representations that produce graduated risk assessments.

Practical example: structuring detection

Consider detecting "structuring" (smurfing): splitting large sums into smaller transfers to evade reporting thresholds. A traditional rule might check: "Block if more than 5 transfers in 30 minutes totaling over $10,000."

This binary approach has obvious problems. Four transfers totaling $9,800? Not flagged. Six transfers totaling $10,100 to a legitimate payroll? Blocked.

FLM enables a more nuanced approach:

EVALUATION	WEIGHT	WHAT IT CHECKS
Velocity Check	High	Multiple transfers in short window
Amount Similarity	High	Transfer amounts suspiciously uniform
Aggregate Threshold	Medium	Total approaches reporting limits

When a transaction arrives, each evaluation contributes its weighted assessment. Activity exhibiting high velocity and similar amounts, but falling short of the aggregate threshold, registers as elevated risk. It may not trigger immediate action, but it's logged, tracked, and available for investigation.

Conversely, activity that trips the aggregate threshold but shows normal velocity and varied amounts receives a lower risk assessment. Context matters.

SCENARIO	BINARY RULE	FLM WEIGHTED ASSESSMENT
4 transfers, $9,800, uniform amounts	PASS · invisible	RISK 0.64 · logged as near-miss
6 transfers, $10,100, legitimate payroll	BLOCK · false positive	RISK 0.64 · below threshold, no customer impact
6 transfers, $10,100, uniform amounts	BLOCK	RISK 1.00 · flagged with full evidence

Fig. 2 · The same activity, two verdicts. Try it live: runloci.com/resource

Near-miss intelligence

One of FLM's operational advantages is visibility into near-misses: transactions that exhibit risk signals but don't cross the action threshold. Traditional binary rules offer no insight here; a transaction either triggered or it didn't.

With weighted assessment, fraud teams can identify patterns clustering just below thresholds, discover emerging attack variations before they mature, generate richer intelligence for investigation teams, and inform ongoing calibration decisions.

Threshold flexibility

Different fraud types warrant different sensitivity levels. A specification detecting potential sanctions violations might require strong evidence across multiple evaluations. One for low-value suspicious patterns might act on weaker signals.

FLM supports configurable thresholds: the same underlying specification can serve different operational contexts by adjusting sensitivity.

CHAPTER 07

Regulatory Alignment

The transparency imperative

Global regulators are converging on explainability as a baseline expectation. While specific mandates vary, the direction is consistent: automated decisions affecting individuals must be understandable in operationally meaningful ways.

EUROPEAN UNION

AI Act Framework

Regulation 2024/1689 establishes a risk-based framework. The Act signals regulatory philosophy: transparency, human oversight, and documentation across AI in financial services.

UNITED STATES

Model Risk Management

Federal Reserve SR 11-7 and OCC Bulletin 2011-12 expect documented validation, conceptual soundness, ongoing monitoring, and independent review.

AFRICA

Emerging Frameworks

Central bank guidelines mandate real-time monitoring, clear escalation procedures, and consumer transparency, with structured reimbursement mechanisms emerging.

European Union: AI Act framework

The EU AI Act (Regulation 2024/1689) establishes a risk-based framework for AI systems. Annex III classifies creditworthiness systems as high-risk except where they are used to detect financial fraud; adjacent use cases such as credit scoring and insurance risk assessment face stricter requirements. More broadly, the Act signals regulatory philosophy: transparency, human oversight, and documentation are expected across AI applications in financial services.

FLM's deterministic execution and complete audit trails align with this direction, providing compliance margin regardless of specific classification.

KEY REQUIREMENTS ADDRESSED

- Human oversight capability (specifications are human-readable)
- Decision documentation and traceability (complete audit trails)
- Clear system logic for regulatory review (the specification *is* the documentation)

United States: model risk management

Federal Reserve SR 11-7 and OCC Bulletin 2011-12 establish model risk management expectations for US financial institutions:

- **Documented validation:** models must be tested and validated with clear documentation
- **Conceptual soundness:** the theoretical basis must be defensible
- **Ongoing monitoring:** performance tracking and periodic review required
- **Independent review:** validation by parties not responsible for development

FLM specifications are inherently documented; the specification itself serves as documentation. Validation is straightforward because behavior is deterministic and reproducible.

Africa: emerging frameworks

African regulators are developing frameworks suited to rapid digital payment adoption. The Central Bank of Nigeria's guidelines mandate real-time monitoring capabilities, clear escalation procedures, and consumer transparency. The emerging regulatory direction points toward structured reimbursement mechanisms and investigative timelines that require institutions to demonstrate clear detection rationale.

FLM's explicit decision logic provides ready-made compliance artifacts for these emerging requirements.

Operational vs. technical explainability

A distinction worth emphasizing: technical explainability tools (SHAP, LIME, attention mechanisms) have advanced significantly. They can identify which features contributed to a prediction and by how much.

But feature importance rankings differ from operational explanations. "velocity_score contributed 0.3 to the prediction" tells a data scientist something useful. "Flagged because this customer made 7 transfers to new beneficiaries in 20 minutes, exceeding the velocity threshold while transfer amounts were suspiciously uniform" tells an analyst what to investigate.

> **FLM'S TARGET**
>
> FLM targets operational explainability. The specification *is* the explanation: readable by analysts, auditors, and regulators without translation.

CHAPTER 08

Operational Benefits

Speed to deployment

Traditional rule development involves specification, coding, testing, and staged rollout: a process measured in weeks. FLM compresses this significantly:

PHASE	TRADITIONAL	FLM
Specification	Days	Minutes
Implementation	Days	Automatic
Validation	Days	Shadow mode
Deployment	Days	Configuration change

When a new fraud pattern emerges, the response gap shrinks from weeks to hours.

Analyst empowerment

FLM shifts ownership from technical specialists to domain experts. Fraud analysts, the people who understand attack patterns, can express detection logic directly. This reduces translation loss between "what we need" and "what got built."

The system performs the engineering work within defined constraints. Analysts focus on fraud patterns, not implementation details.

Shadow mode testing

New specifications can run in shadow mode, evaluating transactions and logging assessments without taking action. This enables:

- Performance measurement before production deployment
- Comparison against existing detection methods
- Threshold calibration with real traffic patterns
- Risk-free experimentation with new detection approaches

Operational trust

> THE TRIAGE DIVIDEND
>
> Analysts trust what they understand. When a specification is readable, its behavior is predictable. Alert triage becomes investigation, not guesswork. The "why did this flag?" question has an immediate, verifiable answer in the specification itself.

Complementing existing investments

FLM integrates with existing fraud infrastructure rather than replacing it:

- ML model scores flow in as input signals
- Existing data pipelines feed transaction context
- Alert management systems receive FLM outputs
- Investigation workflows remain unchanged

Organizations can adopt FLM incrementally, starting with specific fraud types or transaction segments.

CHAPTER 09

Looking Ahead

The adversarial trajectory

The next decade will see continued AI-enabled fraud evolution:

Synthetic identity at scale. LLM-generated identities with coherent multi-year histories and realistic document artifacts.

Adaptive evasion. Attacks that probe detection boundaries and evolve tactics in response to defenses.

Cross-channel coordination. Unified fraud campaigns spanning payments, cards, accounts, and emerging rails.

Deepfake-enabled social engineering. Convincing audio and video for authorization fraud and account takeover.

Defense systems must match this adaptability while maintaining operational trust and regulatory compliance.

FLM evolution

HORIZON	DIRECTION
Near-term	Graph-aware operations for network analysis. Streaming evaluation for continuous monitoring. Enhanced pattern discovery integration.
Medium-term	Multi-modal input support. Automated calibration assistance. Regulatory reporting integrations.
Long-term	Cross-institutional pattern sharing frameworks. Adversarial robustness testing. Industry-standard specification libraries.

The explainable future

The trajectory is clear: operationally opaque automated decisions face increasing scrutiny, operational friction, and regulatory pressure. Systems designed for transparency from the ground up, rather than retrofitting explainability onto opaque architectures, will have structural advantages.

FLM is built for this future.

CHAPTER 10

Conclusion

The fraud defense landscape has reached an inflection point. AI-equipped adversaries, real-time payment rails, and regulatory convergence on transparency create demands that neither traditional rules nor black-box approaches fully address.

The Fraud Language Model offers a different path:

- **Domain-constrained AI** interprets fraud intent and constructs valid specifications: autonomous risk engineering within defined guardrails.
- **Structured specifications** provide human and machine readability: the logic itself serves as documentation and explanation.
- **Deterministic execution** preserves auditability: AI assists authoring while runtime behavior remains reproducible and verifiable.
- **Weighted risk assessment** models fraud behavior through signal composition: graduated responses rather than binary triggers.

This is not rules versus ML. FLM represents an integrated approach where specifications function as risk models: transparent, composable, and calibratable.

The question facing fraud defense teams is not whether to pursue operationally explainable approaches, but when. Institutions that move early will build structural advantages in operational efficiency, regulatory relationships, and customer trust.

The future of fraud defense is explainable.
FLM is how to get there.

ABOUT LOCI

Fraud control that shows its work

Loci Fraud AI builds intelligent fraud defense infrastructure for financial institutions across Africa, Europe, and North America. Our platform, including transaction monitoring, AccessGate continuous authentication and biometrics, and Autographer AI fraud pattern discovery, helps banks and fintechs detect fraud in real time while maintaining full regulatory compliance.

Founded by practitioners with deep experience in scalable technology and AI, Loci combines rigorous engineering with operational pragmatism. We believe fraud defense should be powerful *and* understandable.

WHY NOW

Real-time fraud is rising, but legacy tools are too slow and opaque. Emerging markets face unique fraud tactics and regulatory pressure. Loci empowers teams to build, test, and deploy defenses fast, without needing ML experience.

SEE IT LIVE

Read the interactive edition of this paper and run the structuring example yourself at runloci.com/resource, or book a walkthrough with our team.