KORASTRATUM

Agentic AI in Financial Services

22 Autonomous Agents. 6 Products. Zero LLM Cost.

A Korastratum White Paper

April 2026

korastratum.com

Table of Contents

1. Executive Summary

African and emerging-market financial institutions face a core challenge: manual processes that handle identity verification, compliance checks, banking operations, and customer protection require substantial human effort, cause delays that can take hours or days, and increase risk with each transaction.

This white paper introduces Korastratum's Agentic AI layer: 22 autonomous agents deployed across six products — KoraIDV, CEngine (Kora Compliance), Kora CBA, Kora Digital Banking, Kora Sentinel (Elder Protection), and Stratum Remit. These agents observe data, make decisions, and take action through existing APIs — without modifying any existing code or database schema.

The key insight is that 70% of the value comes from rule-based and ML-powered agents (Tier 1 and Tier 2) at zero LLM cost. These agents use deterministic logic, deploy scikit-learn and ONNX models, and rely on existing API infrastructure. LLM-powered agents (Tier 3) are ready architecturally but are deferred until transaction volume justifies the cost.

Key Results

- 60-80% reduction in manual work across compliance, reconciliation, and identity verification
- $0 LLM cost for Tier 1 and Tier 2 agents (22 of 22 initial deployments)
- Zero regression guarantee -- agents are API consumers, not code modifiers
- Shadow mode validation with 90% agreement threshold before any agent goes live
- Multi-tenant isolation -- each bank runs independent agent instances

2. The Problem: Manual Processes in Financial Services

Financial institutions in Africa and emerging markets rely heavily on manual processes. While transaction volumes are increasing rapidly—driven by mobile money, digital banking, and cross-border remittance—the ability to process, verify, and monitor these transactions remains largely dependent on human effort.

Identity Verification Queues

Even with automated document OCR and selfie matching, 15-25% of identity verifications need manual review. These cases involve borderline document quality, uncertain face match confidence, or liveness detection indicating possible spoofing. Each manual review takes 4-24 hours, creating a bottleneck that directly affects customer onboarding time and conversion rates.

For a microfinance bank processing 500 verifications daily, this means 75-125 cases are in the review queue at any moment. The cost isn't just operational; it's the customers who abandon onboarding because they are waiting for a human reviewer to examine their documents.

Compliance Case Backlogs

Sanctions screening and AML monitoring produce large volumes of potential matches. The issue is that 70% of these matches are ultimately approved after hours of manual investigation — they are false positives that a human analyst must review, document, and clear. Each case investigation takes between 30 minutes and 2 hours, depending on complexity.

For institutions screening against over 1 million watchlist entries from 13 sources, the daily number of potential matches can overwhelm a compliance team. The result: increased backlogs, missed SLAs, and regulatory risks from delayed case resolution.

Reconciliation Breaks

Core banking operations involve daily reconciliation of internal ledgers, payment switch records (NIBSS, Interswitch), card network settlements, and nostro accounts. Breaks—mismatches between expected and actual settlement amounts—necessitate manual investigation. A mid-sized bank typically faces 50-200 reconciliation breaks each day, with each taking 15-45 minutes to resolve.

Settlement Processing

Settlement cycles for NIP (instant payments), NEFT (next-day), and card transactions require manual triggers, monitoring, and exception handling. Settlement officers must verify positions, initiate settlement runs, monitor completion, and investigate failures. This is a repetitive, time-sensitive process that occurs multiple times a day.

Fraud Response Times

Fraud detection in digital banking channels depends on transaction scoring and rule-based alerts. However, the response to a flagged transaction is manual: an analyst reviews the alert, investigates the account history, and decides whether to block, hold, or release. Response times are measured in hours, not seconds — and each hour of delay increases the risk of additional fraud.

Elder Exploitation Detection

Financial exploitation of elderly customers—such as account draining, POA misuse, and scam-induced transfers—is typically detected reactively rather than proactively. By the time a complaint is filed or a pattern emerges, significant financial damage has often already happened. Current systems lack the ongoing behavioral monitoring necessary to identify exploitation in its early stages.

3. Why Traditional AI Falls Short

Most financial institutions have invested in AI and machine learning capabilities. Fraud scoring models, credit risk algorithms, and document classification systems are common. However, these implementations share a key limitation: they are passive.

The Passive AI Problem

- Passive AI assigns scores when prompted, but a human still makes decisions and takes action. A fraud model generates a score; a human analyst reviews the alert and responds accordingly.
- No autonomous action capability. Traditional ML models cannot independently trigger API calls, update records, or start workflows.
- No feedback loop from outcomes. Models are trained periodically on historical data but do not learn from the immediate results of their recommendations.
- High cost of LLMs for simple tasks. Some institutions have implemented LLM-based solutions for jobs that could be managed with deterministic rules or light ML models — paying per-token costs for work that doesn't require language understanding.

The result is a "human-in-the-middle" architecture where AI offers recommendations, but humans remain the bottleneck for every decision and action. This architecture does not scale— as transaction volumes increase, the human layer becomes the limiting factor.

Agentic AI addresses this by shifting from "score and recommend" to "observe, decide, and act." Agents do not replace humans -- they manage the 60-80% of cases that are routine and predictable, allowing human experts to focus on the 20-40% that require true judgment.

4. The Three-Tier Agent Architecture

Korastratum's agent architecture is organized into three tiers based on computational approach and cost profile. This tiered design ensures that the simplest, most affordable solution is always implemented first — with more expensive capabilities only activated when they provide measurable ROI.

Tier 1: Rule-Based Agents (12 Agents)

Tier 1 agents utilize deterministic logic: Python functions, SQL queries, and REST API calls. They encode the decision rules that experienced human operators follow — but execute them in milliseconds instead of minutes.

Agent	Product	What It Does	Cost
Reconciliation	CBA	Auto-match settlement records, flag breaks, attempt resolution	$0
Settlement	CBA	Trigger settlement cycles, monitor completion, handle exceptions	$0
Compliance Monitoring	CEngine	Continuous re-screening based on risk events and watchlist changes	$0
Exploitation Detection	Sentinel	Pattern detection for account draining, POA misuse, scam signatures	$0
Caregiver Monitor	Sentinel	Track POA activity, detect anomalous behavior patterns	$0
Bill Pay Recovery	Digital	Retry failed bill payments, notify customers of resolution	$0
Branch Ops	CBA	Monitor cash positions, trigger replenishment alerts	$0
Escalation Routing	CEngine	Route cases by complexity, analyst expertise, and SLA pressure	$0
Continuous Monitoring	CEngine	Trigger re-screening on risk events and watchlist updates	$0
Match Verification	CEngine	Disambiguate true vs. false positives via entity resolution rules	$0
Adaptive Liveness	KoraIDV	Adjust liveness thresholds by region and fraud pattern	$0
Quote Optimization	Remit	Select optimal liquidity provider based on rate, speed, reliability	$0

Key characteristics: deterministic results, complete audit trail, no cost per decision, and immediate rollback through configuration toggle.

Tier 2: ML-Powered Agents (10 Agents)

Tier 2 agents utilize machine learning models that are already in production. These include ONNX-exported scikit-learn models, FaceNet embeddings, and statistical models running on existing infrastructure. There is no additional cost because these models are already providing predictions.

Agent	Product	ML Approach	Cost
Decision Intelligence	KoraIDV	Risk scoring ensemble (face match + doc quality + device signals)	$0
Fraud Ring Detection	KoraIDV	Graph analysis on face embeddings and device fingerprints	$0
Name Matching	KoraIDV	Multi-algorithm fuzzy matching (Jaro-Winkler, Soundex, N-gram)	$0
Case Investigation	CEngine	Automated evidence gathering and disposition recommendation	$0
Predictive Risk	CEngine	Customer risk trajectory forecasting from transaction patterns	$0
Compliance Disambiguation	CEngine	Entity resolution across conflicting screening results	$0
Lending Decision	CBA	Credit scoring with financial and behavioral features	$0
Treasury Liquidity	CBA	Cash position forecasting and nostro optimization	$0
Fraud Prevention	Digital	Real-time transaction scoring with device and behavioral signals	$0
Rewards Optimization	Digital	Personalized offer selection based on transaction patterns	$0

Key insight: these agents provide advanced ML-driven decisions at no additional cost because the underlying models are already deployed and generating predictions for other platform features.

Tier 3: LLM-Ready Agents (10 Agents)

Tier 3 agents are structurally designed and integration-ready but not yet active. They need LLM inference (Claude API, GPT-4, or self-hosted models) and incur per-token costs. Activation is postponed until transaction volume yields enough ROI.

- Conversational Banking: natural language banking across SMS, WhatsApp, USSD, and in-app chat
- Document Understanding: automated extraction from KYB documents, financial statements, trade documents
- Financial Advisory: personalized savings, investment, and insurance recommendations
- Regulatory Report Generation: automated drafting of CBN and other regulatory submissions
- Customer Communication: context-aware notification drafting and customer service responses

The key design principle: 70% of agent value comes from Tier 1 and Tier 2 at no LLM cost. Tier 3 is additional — it broadens capability but isn't essential for the main value offer.

5. Product-by-Product Impact

KoraIDV (4 Agents)

Metric	Current (Manual)	With Agents
Manual review rate	15-25% of verifications	3-5% (agent handles routine cases)
Review turnaround	4-24 hours	<5 minutes (agent) / 4 hrs (escalated)
Fraud ring detection	Reactive (post-incident)	Proactive (real-time graph analysis)
Name matching accuracy	85-90% (single algorithm)	95%+ (multi-algorithm ensemble)

Decision Intelligence Agent: Monitors verification results, including document OCR confidence, face match score, liveness check, and device signals. Uses a risk scoring ensemble to automatically approve low-risk verifications and reject high-risk cases. Cases that are borderline are forwarded to human review with agent-prepared evidence summaries.

Fraud Ring Detection Agent: Continuously analyzes face embedding similarity across accounts and clusters device fingerprints. Identifies potential fraud rings—groups of applications sharing facial features or devices—and flags them for investigation before individual verifications are approved.

CEngine / Kora Compliance (6 Agents)

Metric	Current (Manual)	With Agents
Case investigation time	30 min - 2 hours per case	<2 minutes (agent-investigated)
False positive resolution	70% approved after manual review	Auto-resolved by agent with audit trail
Case backlog	Growing with volume	Near-zero backlog (real-time processing)
Continuous monitoring	Periodic batch re-screening	Event-driven, real-time re-screening
Escalation accuracy	Round-robin assignment	Skill-based routing by case type

Case Investigation Agent: When a screening match is flagged, this agent automatically collects evidence such as entity details from the watchlist source, customer transaction history, previous screening results, and related entity information. It generates a structured investigation report with a recommended disposition (true positive, false positive, or escalate). For clear false positives, it can automatically close cases with full documentation.

Kora CBA (5 Agents)

Metric	Current (Manual)	With Agents
Reconciliation breaks resolved	50-200/day, 15-45 min each	80% auto-resolved, remainder flagged
Settlement processing	Manual trigger and monitoring	Automated with exception handling
Loan application processing	2-5 business days	Pre-scored in real-time, decision in

		minutes
Treasury forecasting	Spreadsheet-based	ML-driven with intraday updates
Branch cash management	End-of-day reporting	Real-time monitoring with proactive alerts

Reconciliation Agent: Processes settlement files from NIBSS NIP, NEFT, Interswitch, and card networks. Matches transactions to internal GL entries using amount, reference, timestamp, and counterparty. Automatically resolves common break types such as timing differences, duplicate postings, and rounding errors. Flags complex discrepancies with initial analysis for operations staff.

Kora Digital Banking (3 Agents)

Metric	Current (Manual)	With Agents
Fraud detection response	Hours (analyst review)	Real-time scoring, <1 second decision
Bill payment failure rate	3-5% with no auto-retry	<1% with intelligent retry and routing
Rewards engagement	Generic offers, 5-10% redemption	Personalized offers, 15-25% projected

Fraud Prevention Agent: Rates every digital banking transaction in real-time using device fingerprinting, behavioral biometrics (typing speed, navigation habits), transaction velocity, and recipient risk signals. High-risk transactions are blocked immediately; medium-risk transactions prompt step-up authentication.

Kora Sentinel / Elder Protection (2 Agents)

Metric	Current (Manual)	With Agents
Detection approach	Reactive (complaint-driven)	Proactive (behavioral monitoring)
Time to detection	Weeks to months	Hours to days
POA monitoring	Manual periodic review	Continuous automated tracking

Exploitation Detection Agent: Monitors account activity for patterns indicative of financial exploitation—such as sudden changes in withdrawal frequency, pressure-driven addition of new beneficiaries, account balance draining, and transactions that diverge from historical behavior. It generates alerts prioritized by severity, with supporting evidence for the compliance team.

Stratum Remit (2 Agents)

Metric	Current (Manual)	With Agents
Quote selection	Fixed provider per corridor	Dynamic selection across providers
Corridor monitoring	Manual dashboard review	Automated health monitoring and rerouting
Settlement optimization	Conservative liquidity buffers	ML-optimized position management

Quote Optimization Agent: For each remittance request, assesses available liquidity providers across the corridor based on real-time exchange rates, settlement speed, reliability score, and compliance status. Selects the best provider and route, possibly splitting large transfers across multiple providers for optimal execution.

6. Safety by Design

Autonomous agents in financial services need strong safety measures. Korastratum's safety design guarantees that no agent can cause harm, and every decision made by an agent is explainable, auditable, and reversible.

Shadow Mode

Every agent starts in shadow mode. During shadow mode, the agent monitors production data, makes decisions, and logs those decisions — but does not take action. Human operators continue their usual workflows. The shadow mode analyzer compares the agent's decisions with human decisions over a minimum observation period of two weeks.

- Minimum 2-week observation period before any agent can be activated
- 90% agreement threshold: the agent must agree with human decisions on at least 90% of cases
- Disagreement analysis: every case where the agent and human disagree is logged and reviewed
- Activation requires explicit approval from a designated administrator

Human-in-the-Loop

Even after activation, agents function within specified limits.

- High-risk actions always need human approval (e.g., blocking an account, rejecting a loan above the threshold, closing a compliance case as a true positive).
- Configurable risk thresholds per agent per tenant -- each bank defines what "high risk" means
- Escalation paths are always available -- if an agent is uncertain, it escalates rather than guessing
- Human override: any agent decision can be overridden by an authorized operator

Audit Trail

Every agent decision is logged with:

- Structured reasoning: a machine-readable explanation of why the agent made this decision
- Input data hash: cryptographic hash of the data the agent used, ensuring reproducibility
- Confidence score: the agent's self-assessed confidence in its decision
- Outcome tracking: what happened after the decision (was it correct? was it overridden?)
- Timestamp and agent version: for regulatory audit and debugging

Instant Rollback

Each agent can be disabled with a simple configuration toggle. Rollback happens within a minute — no deployment or code changes needed. When an agent is disabled, the platform functions exactly as it did before the agent was turned on. This is possible because agents are API consumers — they do not alter any existing service code or database schema.

Zero Regression Guarantee

- Agents are API consumers only: they call existing REST APIs but never modify existing service code
- New database tables only: agent state is stored in agent_config, agent_decisions, and agent_metrics — no changes to the existing schema.
- Graceful degradation: if an agent process fails, the platform continues operating normally
- No dependency injection: existing services have no awareness of agents and no code paths that depend on agent availability

Multi-Tenant Isolation

Each bank tenant operates its own independent agent instances. Decisions made by one tenant's agent do not access data from other tenants. Tenant-specific configuration, metrics, and audit trails are maintained separately. This separation also applies to the agent orchestrator, scheduler, and shadow mode analyzer.

7. Implementation Architecture

Agent Orchestrator: kora-ai-service

The agent layer is implemented as a new microservice called kora-ai-service, built with Python FastAPI. This service is the 33rd microservice in the Korastratum platform, operating alongside the 32 existing Go microservices that comprise the CBA, IDV, Compliance, Digital Banking, Sentinel, and Remittance products.

Key components:

- Agent Registry: configuration store for all 22 agents, including enable/disable toggles, risk thresholds, and tenant-specific overrides.
- Agent Scheduler: cron-based and event-driven execution of agent observation-decision-action cycles
- Shadow Mode Analyzer: compares agent decisions against human decisions during shadow mode, computes agreement metrics.
- Decision Logger: writes every agent decision to the audit trail with structured reasoning.
- Metrics Collector: tracks agent performance (accuracy, latency, volume) for monitoring dashboards.

BaseAgent Pattern

Every agent inherits from a BaseAgent class that enforces the observe-decide-act pattern:

- observe(): Gather data from platform APIs. Read-only. No side effects.
- decide(): Apply rules, ML model, or LLM to the observed data. Produce a structured decision with reasoning.
- act(): Execute the decision via platform APIs. Only runs if the agent is active (not in shadow mode) and the decision passes risk checks.

This pattern guarantees consistency across all 22 agents, regardless of their tier. A Tier 1 rule-based agent and a Tier 3 LLM agent follow the same lifecycle, logging, and safety procedures.

Database Schema

Table	Purpose	Key Columns
agent_config	Agent registry and configuration	agent_id, tenant_id, enabled, mode (shadow/active), risk_thresholds, schedule
agent_decisions	Immutable audit trail	decision_id, agent_id, tenant_id, input_hash, decision, reasoning, confidence, outcome, created_at
agent_metrics	Performance monitoring	agent_id, tenant_id, period, decisions_count, accuracy, avg_latency, agreement_rate
agent_shadow_comparisons	Shadow mode analysis	agent_id, tenant_id, agent_decision, human_decision,

		agreed, case_ref

These tables are built in the kora-ai-service database. No existing service databases are altered.

Integration with Existing Services

Agents integrate with existing microservices exclusively through REST APIs -- the same APIs that the bankapp frontend, admin dashboard, and third-party integrations use. This means:

- No new internal APIs are needed for agent integration
- Agents are subject to the same authentication, rate limiting, and audit logging as any other API consumer
- If an agent is disabled, no API consumer is affected -- the APIs continue serving all other clients normally

8. Multi-Tenant Banking-as-a-Service

Korastratum functions as a Banking-as-a-Service platform where multiple bank tenants share the platform infrastructure while maintaining strict separation. The agentic AI layer expands this multi-tenant architecture:

Tenant-Scoped Agent Configuration

- Each tenant has independent agent configurations -- one bank may enable all 22 agents while another enables only 5
- Risk thresholds are configurable per tenant -- a conservative bank may set higher human-approval thresholds
- Shadow mode runs independently per tenant -- agents can be in shadow mode for one bank and active for another

Credential Isolation

Each tenant's agents authenticate to platform APIs using tenant-specific credentials encrypted with AES-256-GCM. Agent service accounts are provisioned per tenant with the minimum permissions needed for each agent's function.

Network Isolation

For integrations with external payment gateways like NIBSS NIP, Interswitch, and card networks, each tenant uses a dedicated VPN tunnel. Agent-initiated settlement and payment operations go through the tenant's isolated network path.

Application Isolation

The CBA core banking platform is shared infrastructure, but each tenant's bankapp (mobile app, web portal, admin dashboard) is a separate deployment with tenant-specific branding, configuration, and feature flags. Agent decisions that impact customer-facing experiences (e.g., rewards offers, fraud blocks) are limited to the tenant's bankapp instance.

9. Results: First Production Deployment

FMFB (Firstmidas Microfinance Bank)

Firstmidas Microfinance Bank (FMFB) is the first production tenant on the Korastratum platform, providing a real-world validation environment for the full agent suite.

Deployment Scope

- 22 agents deployed across KoraIDV, CEngine, CBA, Digital Banking, Sentinel, and Remit
- Integration with NIBSS NIP for instant payments and Interswitch for card processing
- Full KYC flow with KoraIDV (BVN/NIN verification, document OCR, selfie matching, liveness detection)
- AML/CFT screening via CEngine against OFAC, UN, EU, UK, and Nigerian NIGSAC sanctions lists
- Core banking operations on Kora CBA with double-entry GL, deposits, lending, and treasury

Shadow Mode Status

As of April 2026, agents are operating in shadow mode -- monitoring production data, making decisions, and recording outcomes to compare with human operators. Initial shadow mode metrics are being gathered across all agent categories to establish baseline agreement rates.

Expected Outcomes

Based on shadow mode observations and comparable deployments:

Metric	Projected Impact
Manual KYC review reduction	60-70% of reviews handled by Decision Intelligence agent
Compliance case resolution	70%+ false positives auto-resolved by Case Investigation agent
Reconciliation automation	80% of daily breaks auto-resolved by Reconciliation agent
Fraud detection speed	Real-time scoring replacing hour-delayed manual review
Elder protection	Proactive detection reducing time-to-alert from weeks to hours

10. Regulatory Alignment

Autonomous agents in financial services must function within regulatory frameworks. Korastratum's agent architecture is built to ensure compliance from the beginning.

CBN Compliance Requirements

The Central Bank of Nigeria (CBN) mandates financial institutions to uphold strong KYC/AML procedures, transaction monitoring, and regulatory reporting. Korastratum's agents assist in fulfilling these requirements by:

- Automating KYC verification while maintaining human oversight for high-risk cases
- Providing continuous transaction monitoring that exceeds the capacity of manual review
- Generating structured audit trails that satisfy regulatory examination requirements
- Supporting regulatory reporting with automated data collection and formatting

Explainability

Every agent decision includes structured reasoning—a machine-readable and human-readable explanation of the factors that influenced the decision, the weights applied, and the threshold used. This explainability is essential for regulatory review: a compliance examiner can trace any agent decision back to its inputs, logic, and outcome.

AI Governance Framework

- Model Registry: each ML model used by Tier 2 agents is documented with its version, training data description, performance metrics, and approval status.
- Performance Monitoring: agent accuracy, false positive/negative rates, and agreement rates are tracked continuously and reported on dashboards
- Bias Detection: agent decisions are analyzed for demographic and geographic bias on a regular cadence
- Change Management: agent configuration changes (threshold adjustments, activation/deactivation) require approval workflows and are logged in the audit trail

11. Future: Tier 3 LLM Agents

Tier 3 agents represent the next stage of the platform -- natural language interfaces and document understanding capabilities powered by large language models. These agents are architecturally ready, but activation is delayed based on ROI analysis.

Conversational Banking

Natural language banking across SMS, WhatsApp, USSD, and in-app chat. Customers can check balances, initiate transfers, apply for loans, and resolve issues through conversational interfaces in English, Pidgin, Yoruba, Hausa, and Igbo. The agent understands context, maintains conversation state, and executes banking operations via the CBA API layer.

Document Understanding

Automated extraction and analysis of KYB documents (such as certificates of incorporation, shareholder registers, and financial statements), trade finance documents (including letters of credit and bills of lading), and regulatory filings. The agent extracts structured data, cross-checks it against external sources, and generates compliance-ready reports.

Financial Advisory

Personalized financial recommendations based on customer transaction history, savings patterns, and life events. The agent can suggest savings products, loan options, insurance products, and investment opportunities — all within regulatory guidelines and suitability standards.

When to Activate

Tier 3 activation is driven by volume thresholds and ROI analysis:

- Conversational Banking: when customer service ticket volume exceeds 1,000/month per tenant
- Document Understanding: when KYB document processing exceeds 100/month per tenant
- Financial Advisory: when the bank introduces savings and investment products with cross-sell potential.

The architecture supports both cloud LLM APIs (Claude, GPT-4) and self-hosted models for organizations that need data residency compliance.

12. About Korastratum

Korastratum is the technology partner for African financial institutions building the next generation of banking infrastructure. The platform consolidates six products into a single integration.

- KoraIDV: Identity verification with document OCR, selfie matching, and liveness detection
- CEngine (Kora Compliance): AML screening, sanctions checking, case management, and transaction monitoring
- Kora CBA: Core banking with 25 microservices, 350+ API endpoints, and full GL
- Kora Digital Banking: White-label mobile app, web portal, and admin dashboard
- Kora Sentinel: Elder financial exploitation detection and prevention
- Stratum Remit: AI-powered cross-border remittance across 40+ corridors

Contact

Website: korastratum.com

Email: hello@korastratum.com

Demo: korastratum.com/demo

Korastratum | April 2026 | korastratum.com