VINOD SHARMA · ENTERPRISE CTO & PRINCIPAL AI SYSTEMS ARCHITECT

The Architect Behind the Rails

Twenty five years of mission critical financial infrastructure, and what it taught me about where artificial intelligence is allowed to touch the money.

13	100M+	$1B+	4
Years multi-market Group CTO	Active financial lives	Daily transaction value	Production AI systems shipped

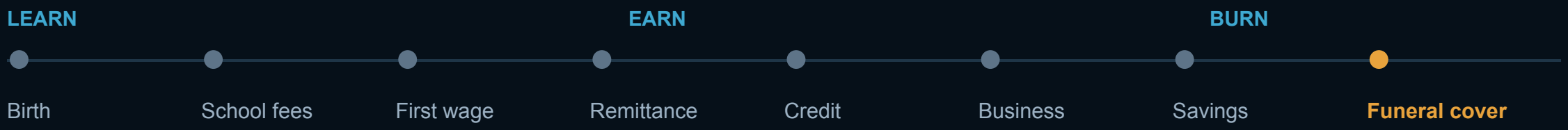
Perfection or Nothing

M-Pesa Africa · Ascend Money / TrueMoney · Econet & EcoCash · Saolix

THE HUMAN LIFECYCLE · NURSERY TO MORTUARY

Learn. Earn. Burn.

Financial technology is not transactional software. It is the invisible backbone of a human economic journey, and every outage lands somewhere on that line.



Blessings. Masvingo. She needed to bury someone.

She did not ask me for help. She did not know who I was. And nothing about her situation could wait for our next release.

2M → 9M active users

99.99% core availability

$70B annual value

Zimbabwe, 2014-2019. Built while physical cash was disappearing from people's hands.

THE GENESIS & EVOLUTION OF GLOBAL RAILS

When money left the branch

Mobile money was not innovation for its own sake. It was market friction finding a release, and everything that used to be physical control had to become architecture.

BRICK & MORTAR LEDGER BANKING	BOUNDARYLESS MOBILE ECOSYSTEMS
Branch bound, hours of travel	Agent networks and USSD reach
Cash settlement, manual reconciliation	Real time value transfer
Identity proven on paper	eKYC and digital identity
Trust carried by the building	Trust carried by the system

FOUR FRICTIONS FORCED THE SHIFT

Distance **Cost** **Trust** **Identity**

THE NON-NEGOTIABLE CORE

GMPI

Four things every platform is doing at once, whether the organisation admits it or not.

GMPI is not four projects. It is one budget conversation.

G Grow

New markets, cross border interoperability, unbanked and digital native on one rail.

25% YoY transaction value expansion

M Maintain

The unglamorous pillar. Throughput, availability, 24/7 operations that nobody applauds.

5,000+ TPS · 99.9%+ group uptime

P Protect

Zero trust, regulatory defensibility, AML and sanctions, central bank data sovereignty.

This pillar is the licence

I Improve & Transform

Replatforming, vendor displacement, applied AI. Funded first. Should be funded last.

45% faster delivery · 15% lower cloud spend

Most enterprises fund Improve and starve Maintain. Then Protect fails, and Grow stops.

THE SCALED ENTERPRISE REALITY

Fragmented heterogeneity

Tens of thousands of disparate legacy nodes, multi-currency ledgers and distributed databases.

You cannot rip this out. You connect it, and you pay for it either way.

THE GST OF ENGINEERING

Architectural debt is not a metaphor. It is a tax. Track it, tax it, amortize it, or it collects silently.

The stack a payments platform actually runs on

Six layers. Every one of them has to hold at 5,000 TPS while a regulator can still ask what happened to a single transaction three years ago.

CHANNELS	USSD · Mobile app · QR · NFC · Agent · Web
EXPERIENCE & APIS	API gateway · Partner onboarding · Developer portal · Rate limiting
INTEGRATION	Event streaming · ESB and micro-integration · CQRS · Transactional outbox
CORE FINANCIAL SERVICES	Wallet · Payment switch · Remittance · Lending · Merchant · Wealth
TRUST & CONTROL	eKYC · AML and sanctions · Fraud scoring · HSM · Zero trust · Audit
DATA & INTELLIGENCE	Ledgers · Streaming data · Vector stores · Models · Drift monitoring

MIDDLEWARE · MODERNISATION

Open source flexibility, commercial grade assurance

The vendor commoditisation debate misses the point. What you are choosing is whether you can read the code when it breaks, and whether someone answers the phone when it breaks at 3am.

WSO2 OPEN SOURCE

- Source available, no black box under the rails
- Extend and adapt to local market reality
- No licence wall in front of experimentation
- Engineering team owns the failure mode

WSO2 COMMERCIAL

- Supported releases and security patching
- Escalation path when production is on fire
- Regulatory and audit posture you can defend
- Roadmap alignment for multi-market scale

MODERNISE

You do not modernise by replacing everything. You modernise by putting an integration layer in front of what you cannot replace, and then replacing it quietly, behind the interface.

THE M-PESA JOURNEY

Middleware as the digital nervous system

Africa's largest financial rails, connected through an integration layer rather than a rewrite.

This is the part of the architecture that customers never see and never forgive.

01

Fragmented estate

Multiple operating companies, legacy cores, partner platforms, multi-currency ledgers.

→

02

Integration layer

APIs, events and micro-integration placed in front of what could not be replaced.

→

03

Core replatforming

Kenya and Tanzania to hybrid cloud, 5,000+ TPS capacity, behind stable interfaces.

→

04

Scale held

200M+ daily transactions, 100M+ lives, 99.9%+ group availability.

200M+
daily transactions

100M+
active lives

20+
payments programmes delivered

BEYOND THE GENERATIVE HYPE

Two professions, one word

A customer care assistant and a sanctions screening engine are both called AI. They are not the same discipline, and they should never share a risk appetite.

CONVERSATIONAL ASSISTANTS

A wrong answer is a bad review.

Tolerance for hallucination

Failure is recoverable

Optimised for fluency

Cloud API is acceptable

FINANCIAL CRIME SYSTEMS

A wrong answer is a licence.

Zero tolerance for hallucination

Failure is a regulatory event

Optimised for defensibility

Sovereign deployment required

The question was never whether AI belongs in finance. It is where exactly you let it touch the money.

Five different technologies, one overloaded word

Everything on this slide gets called AI. They have different failure modes, different audit properties, and they do not belong in the same place in a financial system.

Deterministic rules Policy, thresholds, ruleset versions	Fully explainable. Repeatable. Auditable.	**DECIDES**
Classical ML Fraud scoring, risk models, regime classification	Statistically validated. Drift monitored.	**SCORES**
LLM / Gen AI Language understanding, summarisation, drafting	Fluent. Probabilistic. Never authoritative.	**EXPLAINS**
RAG Grounded retrieval over a governed corpus	Cites evidence. Reduces, not removes, error.	**RETRIEVES**
Agentic AI Multi-step execution of intent under governance	Acts. Therefore needs identity and revocation.	**ACTS**

Only one of these five is allowed to decide. Everything else advises.

THE MODERN INTELLIGENCE STACK

The Determinism Boundary

Probabilistic intelligence explores. Deterministic policy decides. Nothing crosses the line, and the unfashionable half of AI turned out to be the only half you can put under money.

AI-ASSISTED INTELLIGENCE

Sovereign sub-3B SLMs · Agentic RAG · ONNX CPU runtime · Vector retrieval · Drift detection

THE DETERMINISM BOUNDARY

AI assists. Deterministic rules decide.

DETERMINISTIC AUTHORITY

Authorization · Ledger · Settlement · Pinned ruleset and thresholds · Sealed evidence

Match chain

Bloom → trigram → Jaro-Winkler → Daitch-Mokotoff → vector fusion

Audit ledger

SHA-256 tamper evident, byte identical verdicts

Regulator replay

The regulator reconstructs my decision themselves

WHERE THE LINE CAME FROM

The system did not hold.

Not in a test environment. In production, at volume, with licences attached. Everything I just showed you exists because of two months I do not enjoy talking about.

01

The failure

Screening degraded under real traffic. The load was not theoretical and neither were the consequences.

02

The two months

Not features. Precision, throughput, and what the chain actually does when volume is real.

03

What was left

The match chain, the audit ledger and the replay endpoint. The boundary became a product.

Anybody can code. Engineering is what happens when the system meets real traffic and real business.

Outcome · 18% lower infrastructure run rate · auditability uncompromised · still in production today

OPERATIONAL AUTONOMY · KNOW YOUR AGENT

An agent is an actor

Autonomy earns its way in. Operations first, user intent second. And twenty five years taught me never to let an unidentified actor near the money.

Back of house
Mutation testing, telemetry, observability
16,000+ test suite

Zero touch operations
Self hosted ops intelligence across L0, HUB and monitoring
15% MTTR · 20% onboarding

Multi-agent execution
27 agents under a ten level governance hierarchy with consensus voting
Executes complex user intent

188

of 200 technical staff adopted it.

Tell them why the other twelve did not.

KYA · Identity. Scope. Revocation. Audit trail. It is KYC, applied to machines.

THE INTERNET OF FINANCIAL SERVICES

Convergence has a physical floor

Identity, sovereign agents and intelligent middleware converge into self orchestrating finance.

And then thermodynamics sends the bill.

Identity

eKYC, KYA, verifiable actors

Sovereign agents

On-prem, air-gapped, governed

Intelligent middleware

The nervous system, now reasoning

LANDAUER'S PRINCIPLE, APPLIED TO AGENTIC AI

Autonomy is not free.

Every irreversible computation has a thermodynamic minimum cost. Every agent thought is energy.

Agent swarms are not free, and this industry is designing as though they are.

BACK TO THE LINE

Anybody can code.

Engineering is what happens when the system meets real traffic and real business.

The engineering was always for that line. Blessings never saw any of it. That was the point.

Perfection or Nothing

FRAMEWORKS REFERENCED IN THIS TALK

Five lines I have drawn

The Determinism Boundary AI assists. Deterministic rules decide.	Probabilistic intelligence above the line, authorization and settlement below it.
Know Your Agent (KYA) An agent is an actor.	Identity, scope, revocation and audit trail. KYC, applied to machines.
GMPI One budget, not four.	Grow, Maintain, Protect, Improve and Transform, funded as a single conversation.
The GST of Engineering Debt is a tax.	Track it, tax it, amortize it, or it collects silently.
Landauer for Agents Autonomy has a physical floor.	Every agent thought is energy. Swarms are not free.