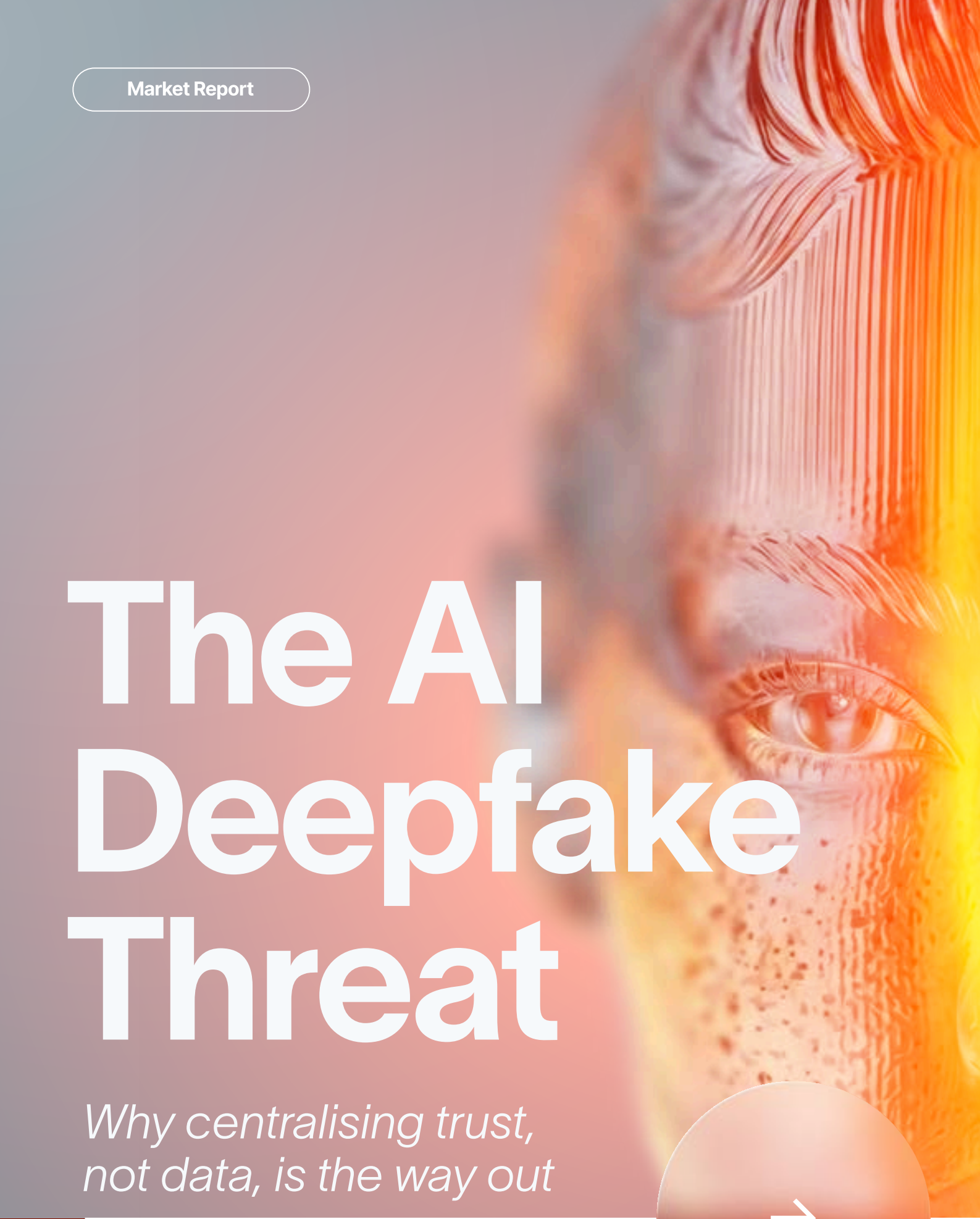
Market Report



The AI Deepfake Threat

Why centralising trust, not data, is the way out

BROUGHT TO YOU BY AlphaX

Executive Summary

Detection asks: is this face real? Architecture asks: how much is a fake face worth?

The Threat

Synthetic media has moved from a novelty risk to a **quantified, line-item cost** across financial services, government, and civic life. Documented global fraud losses tied to deepfakes and AI-enabled impersonation are now estimated in the **billions of dollars annually**, and multiple industry trackers report year-over-year increases well above 100%, with some categories surging by an order of magnitude or more.

Africa is not a peripheral market for this threat, it is one of the fastest-moving fronts. South Africa, Nigeria, and Kenya are each reporting sharp increases in deepfake-driven identity fraud. **The underlying enabler is the same everywhere:** identity systems that re-verify people with cameras and documents at every step, backed by centralised biometric databases that were never designed for an era of cheap, convincing, AI-generated impersonation.

Our Position

The standard response to deepfakes is better detection at the point of capture: sharper liveness checks, smarter forensics. Detection is a race, and the fakes are gaining. The alternative is architectural: verify a person once, at one supervised and well-defended enrollment, anchor that event cryptographically, and let every subsequent check be a proof rather than a face. Detection asks: is this face real? Architecture asks: how much is a fake face worth?

What the numbers say

$3.7B+

Documented global deepfake fraud losses, 2025–H1 2026

(Surfshark / AI Incident Database)

1 in 5

Biometric fraud attempts globally now involve a deepfake.

(Surfshark / AI Incident Database)

269%

Year-on-year jump in South African deepfake incidents (2026)

1. The Many-Moments World

To understand why deepfakes work so well against identity systems, follow one citizen through a typical digital life.

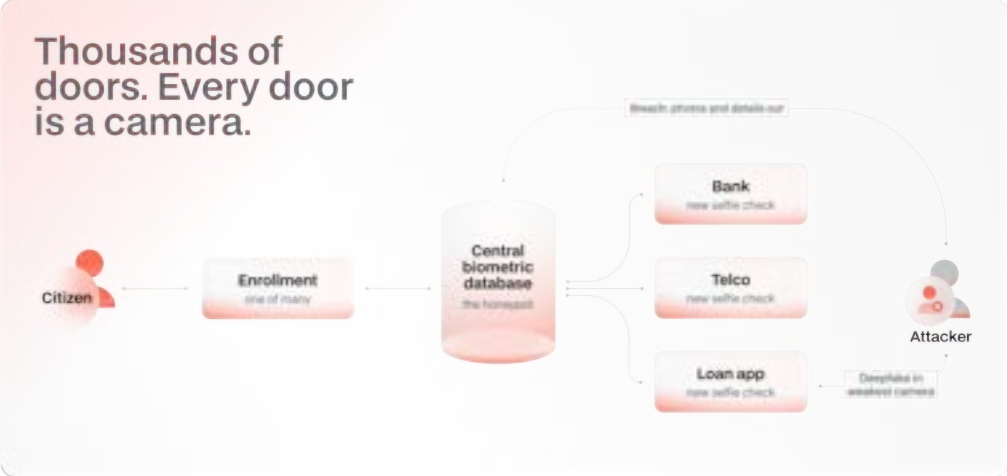
She enrolls with the national identity authority once. **But that enrollment does not travel.** When she opens a bank account, the bank runs its own face capture and matches it against her documents or a database lookup. When she buys a SIM, the telco does the same. A mobile loan, the same again. Each of these is a fresh camera moment, **run by a different organisation,** on different equipment, with different quality checks.

An attacker does not need to beat the strongest of these checks. He needs to beat the weakest selfie flow in the country, and he can try them all. In a many-moments system there are thousands of doors, and every door is a camera.

1.1 The honeypot problem

Behind those doors sits a second structural weakness. To make all this re-verification possible, the system keeps everyone's biometric and document data piled up in central databases. That pool has two properties that attract attackers: it is the single point whose compromise breaks everything, and it is the raw material factory for the attack itself. Breached photos and personal details are exactly what modern generative tools need to produce a convincing fake. The database that verifies you is also the warehouse that arms your impersonator.

The result is a loop. Central data gets breached, breached data feeds deepfakes, deepfakes beat the weakest camera, and every successful pass creates records that look genuine everywhere else.



> **The database that verifies you is also the warehouse that arms your impersonator.**

2. What Is Happening Now

2.1 Scale and trajectory

No single, universally agreed global loss figure exists for deepfake fraud. Trackers measure different things: complaints, verification rejections, blockchain flows, survey responses, and combining them would double-count. Read together, however, the datasets converge on the same direction of travel: fast, accelerating growth from a low base.

- The FBI's Internet Crime Complaint Center logged 22,364 complaints with an AI-related descriptor in 2025, totaling roughly $893 million in adjusted losses, the first year the Bureau broke this out as its own category.
- Surfshark, drawing on public incident databases, put cumulative documented global deepfake fraud losses at upwards of $3.7 billion, with about 89% of that recorded in 2025 and the first half of 2026, a sign of how recent and how steep the acceleration has been.
- Entrust's 2026 Identity Fraud Report, based on more than a billion identity verifications across 195 countries, found that deepfakes now account for roughly one in five biometric fraud attempts worldwide.
- Gartner found 62% of organisations surveyed had already experienced a deepfake-related incident; Deloitte's fraud-expert survey found sizable shares had personally encountered synthetic identity fraud, voice deepfakes, and video deepfakes.
- Congressional and industry analysis suggests fewer than 5% of voice-clone fraud victims ever file a report, meaning every official figure above should be read as a floor, not a ceiling.

2.2 The African front

Global coverage is uneven, but the data that does exist points to Africa as a region where deepfake-enabled identity fraud is growing faster than fraud overall, a signal that attackers are actively shifting tactics rather than riding a general fraud wave.

- South Africa: Smile ID's 2026 Digital Identity Fraud Report found South Africa has the highest share of deepfake-driven fraud on the continent, with AI-generated impersonation implicated in 22% of cases, and nearly nine in ten rejected verification attempts in the region now linked to AI-assisted impersonation or spoofing. Separate industry reporting recorded a 269% year-on-year jump in deepfake incidents even as overall fraud rates fell.
- Nigeria: instant-payment volumes have roughly doubled since 2022, and fraud losses on that rail jumped 603% in Q1 2025 alone, driven in part by deepfake-enabled impersonation used to defeat know-your-customer checks at fintechs and banks.
- Kenya: deepfake-related attacks now account for close to 10% of all fraud attempts, even as the overall fraud rate has been falling, again indicating a shift in method rather than a shift in overall criminal activity.
- Continentally: INTERPOL's Africa Cyberthreat Assessment and related reporting put mobile-money fraud losses, a category increasingly intertwined with AI-enabled impersonation and SIM-swap attacks, at more than $4 billion a year. Cross-border criminal networks exploit the gap between national fraud-intelligence systems; only 14 of 55 African Union member states have ratified the continental cybersecurity convention meant to close it.

What Is Happening Now

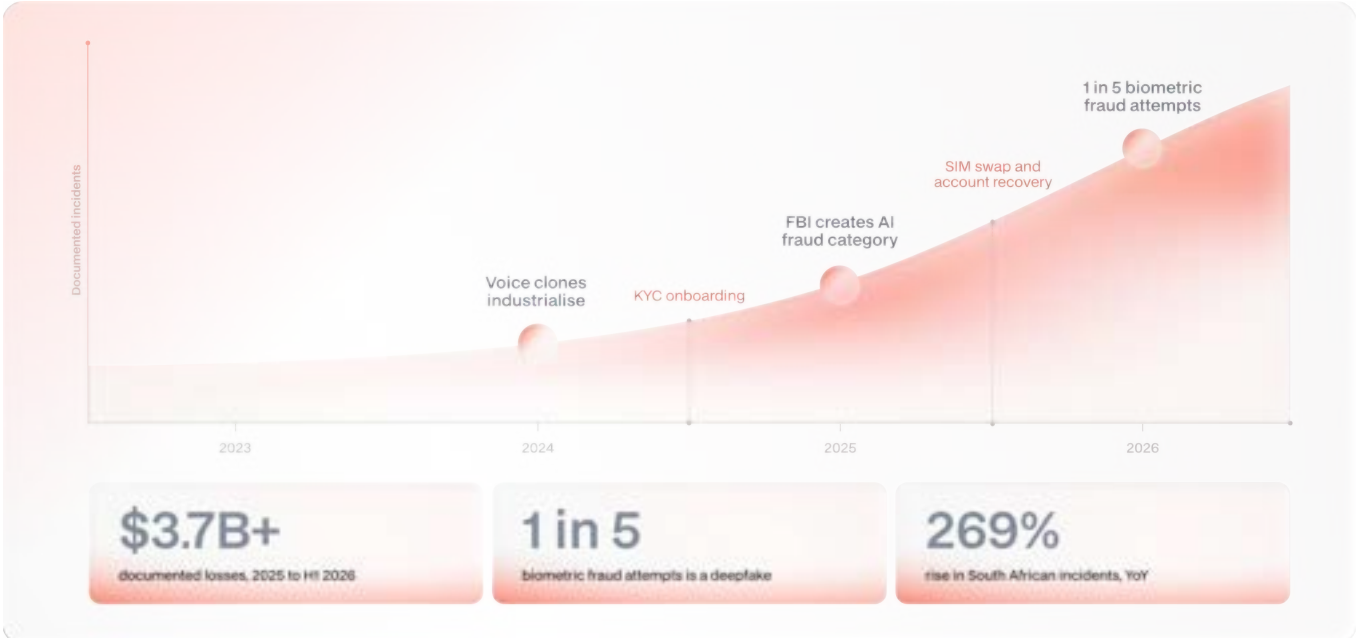
2.3 The human cost

This is not an abstract technology risk. Nigerian fact-checkers have documented AI-generated videos impersonating well-known business and public figures, including a prominent business executive and the WTO's director-general, to promote fake investment schemes aimed directly at ordinary savers. Globally, deepfakes are increasingly used in non-consensual intimate imagery and romance-scam profiles; over 70% of reported romance-scam profiles in one FTC-linked analysis now involve AI-generated or manipulated imagery.

Survey research adds a regional twist. Roughly three quarters of respondents in the Middle East and Africa say they trust AI, largely because it is seen as a tool for economic growth and access to services, against under a third in North America and Western Europe. That optimism is a double-edged sword: the populations most open to AI-enabled services are also among the most exposed to having that trust weaponised through impersonation, because verification habits have not caught up with the pace of adoption.

Uganda's 2026 general election illustrates the civic stakes. In the run-up to the January vote, civil-society groups warned publicly that deepfakes of candidates could incite unrest or impersonate election officials, and the Electoral Commission issued a public advisory about fabricated videos circulating online.

> **Every official figure should be read as a floor, not a ceiling.**



3. Assessing the Risk

Overall assessment: HIGH and rising. Deepfake-enabled identity fraud sits directly on top of the verification infrastructure that national ID, financial inclusion, and land registration programmes depend on. It warrants sustained strategic focus, not a one-off response.

3.1 Risk by dimension

Outcome	Horizon	Risk level
Financial fraud (banking, fintech, mobile money)	Active now	CRITICAL
Identity/KYC bypass (biometric and document verification)	Active now	CRITICAL
Land and property registry impersonation fraud	Emerging	HIGH

Outcome	Horizon	Risk level
Political disinformation / election integrity	Active, cyclical	HIGH
Personal reputational and intimate-image harm	Active now	HIGH
Enterprise business-email-compromise (voice/video)	Active now	MEDIUM-HIGH
Erosion of general trust in digital and video evidence	Structural, long-term	MEDIUM

3.2 Possible futures: short term (12 to 18 months)

Outcome	Risk level
Continued surge in KYC and onboarding fraud outpaces detection tooling at smaller institutions	HIGH
First large-scale, publicly attributed land-title or benefits-fraud incident using deepfakes in Africa	HIGH
Patchwork national regulation creates compliance complexity for cross-border fintech and govtech	MEDIUM
Citizen-reported personal-harm cases outpace legal remedies	HIGH

3.3 Possible futures: long term (3 to 5 years)

Outcome	Risk level
Bifurcation: verified digital identity becomes a precondition for participation in finance, land, and voting systems	STRUCTURAL
Cryptographic provenance and proof-based verification become the default trust layer, not an add-on	OPPORTUNITY
Sustained erosion of default trust in unmediated video and audio evidence across courts, media, and elections	HIGH
A small number of identity-infrastructure providers become critical national infrastructure for African governments	STRATEGIC

Read together, the short-term pathways are mostly defensive: contain damage, patch gaps. The long-term pathways are structural: identity verification becomes foundational infrastructure rather than a bolt-on control. Which future arrives depends less on detection tooling than on the architecture decisions national programmes make in the next two years.

4. The One-Moment World

There is a different way to arrange the same system. Picture the ideal state. A citizen is verified once, at one supervised, well-defended enrollment: hardened capture equipment, strong liveness checks, biometric deduplication. That single event is anchored cryptographically, so it can never be quietly inserted, altered, or deleted afterwards. The citizen receives a credential, held on her phone or a smart card, locked to keys only she controls. From then on, when a bank, telco, or agency needs to know she is a real, valid, government-verified person, her credential produces a cryptographic proof that says exactly that, and only that. No selfie, no document photo, no database lookup, no citizen data transferred.

A deepfake is a fake face. In the many-moments world, a fake face is useful thousands of times a day, at every weak camera in the country. In the one-moment world, a fake face is only useful at one place: enrollment, the single most defended, supervised, deduplicated moment in the whole system. Everywhere else, there is no face to fake. You cannot deepfake math.

4.1 What this does and does not solve

Honesty matters here, because the audience for this argument is technical. Proof-based architecture does not solve the camera. The one enrollment moment still depends on liveness detection, presentation-attack defence, and deduplication, and a sufficiently good fake could still get through it. Those defences are a separate discipline, and the organisations best at them are the incumbent biometrics vendors.

What the architecture changes is the economics. In a many-moments system, a successful fake scales: it works at every weak camera, invisibly, repeatedly. In a one-moment system, a successful fake is single-use and traceable: it yields one anchored credential that can be found and revoked, there is no central biometric pool to mine for the next attack, and no record can be added or edited after the fact without breaking the anchor. Fraud does not become impossible. It stops scaling.

4.2 The role of zero-knowledge proofs

Two things make the one-moment model workable, and zero-knowledge proofs are the second of them. The first is portability: the result of one good verification must be reusable everywhere, so that downstream institutions no longer need cameras. The second is safety of that reuse. Without ZK proofs, portable identity would mean transmitting citizen data to every verifying institution, a privacy and sovereignty failure that no government should accept. A ZK proof lets the citizen demonstrate validity while the data stays put: the verifier learns that the claim is true, and nothing else. Trust becomes portable. Data does not move.

> **You cannot deepfake math.**

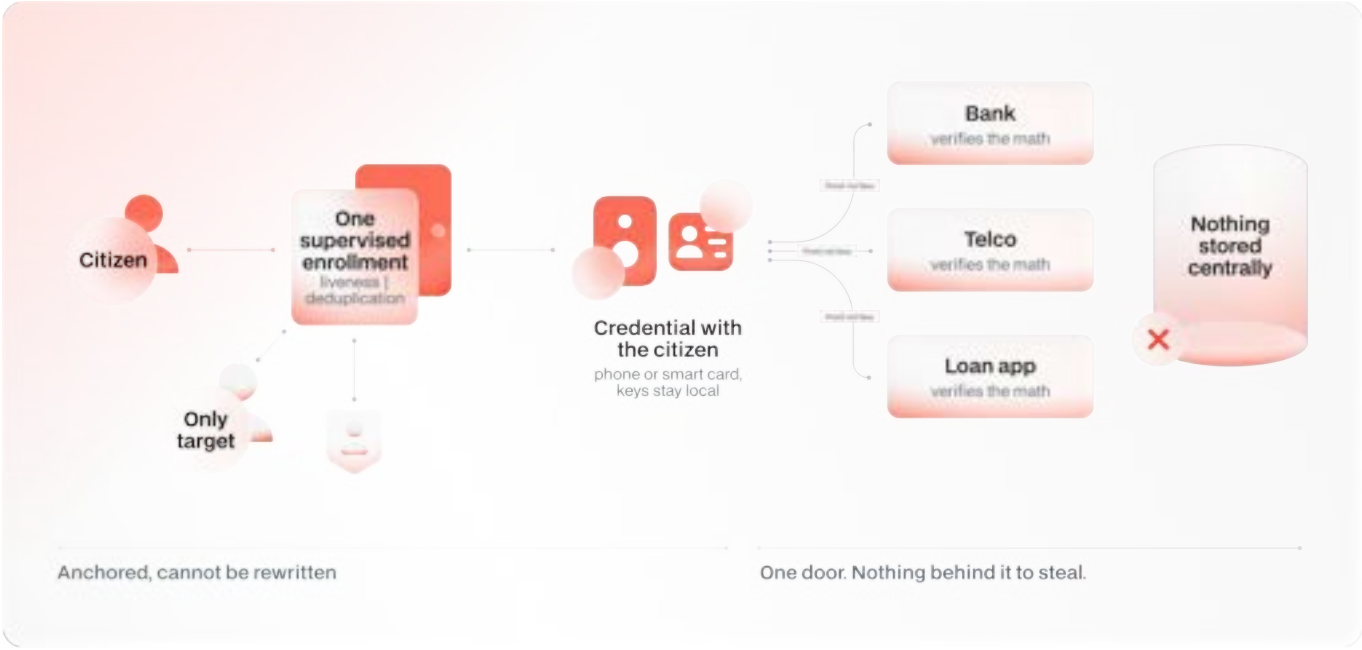
The One-Moment World

4.3 Anchoring, and why after-the-fact fraud dies

Anchoring means that when someone enrolls, a small cryptographic fingerprint of that event, not the person's data, is written to an append-only base layer. Because that layer cannot be rewritten, nobody can later insert an enrollment that never happened or quietly edit a real one. Ghost identities, the insider-driven half of synthetic identity fraud, fail every check because they have no matching anchor. A faster second layer handles the day-to-day verification traffic, including in low-connectivity environments, so proof checks work in a rural district office as well as a capital-city bank.

4.4 Is this feasible for a state?

The objections are practical, and each has an answer visible in live deployments. Smartphone coverage: the same credential lives on a chip card for citizens without phones. Loss and theft: a credential is useless without its local unlock, and it can be revoked centrally in minutes, unlike a stolen paper ID that works at any counter until it expires. Connectivity: proof verification is designed to work offline, because the proof carries its own validity. Institutional adoption: banks and telcos have every incentive to accept proofs, since each proof-based check removes a compliance cost and a fraud channel they currently own. None of this requires a leap of faith; every component is running somewhere today, as the next section shows.



5. Who Is Trying This

The one-moment world is not a thought experiment. Pieces of it are running at national and continental scale today. What separates the deployments that work from those that struggle is not technology maturity, it is where they chose to centralise.

Aadhaar (India) *scale proven, architecture aging*

Launched in 2009, Aadhaar proved the core value proposition at 1.3 billion people: enroll once, use everywhere. It also demonstrated the cost of the first-generation design. Aadhaar centralises both trust and data: one authority, one national biometric database, and verification that queries that database, often with a fresh fingerprint or face capture at the point of service. The result is the honeypot problem at maximum scale, plus a camera moment at every counter. Tellingly, researchers have since built zero-knowledge layers on top of it (Anon Aadhaar) so people can prove they hold a valid Aadhaar without exposing anything. Even the pioneer is being retrofitted toward the newer model.

EU Digital Identity Wallet (eIDAS 2.0) *the new blueprint*

Designed from 2021 onward, the EU wallet is the lesson of Aadhaar learned in law and architecture. Citizens verify once against their national ID, then hold credentials in a wallet and present cryptographic proofs of specific claims: valid, over 18, resident, without exposing the data behind them. No central lookup per transaction, selective disclosure by design, and member-state implementations are actively adopting zero-knowledge techniques. It centralises trust in issuing authorities while distributing the data to citizens' own devices.

MOSIP (Philippines, Ethiopia, Morocco, Togo) *sovereignty, open-sourced*

MOSIP is not a product but an open-source foundation for national ID, developed at IIIT Bangalore. A government takes the codebase, deploys it on its own infrastructure, and runs enrollment, deduplication, and issuance itself: own the code, own the servers, no vendor lock-in. Its newer modules move toward exactly the credential-and-proof model described above. MOSIP matters to this report for two reasons: it is the African reference point, and it demonstrates that sovereignty-first architecture is a deliberate design choice, not a premium feature.

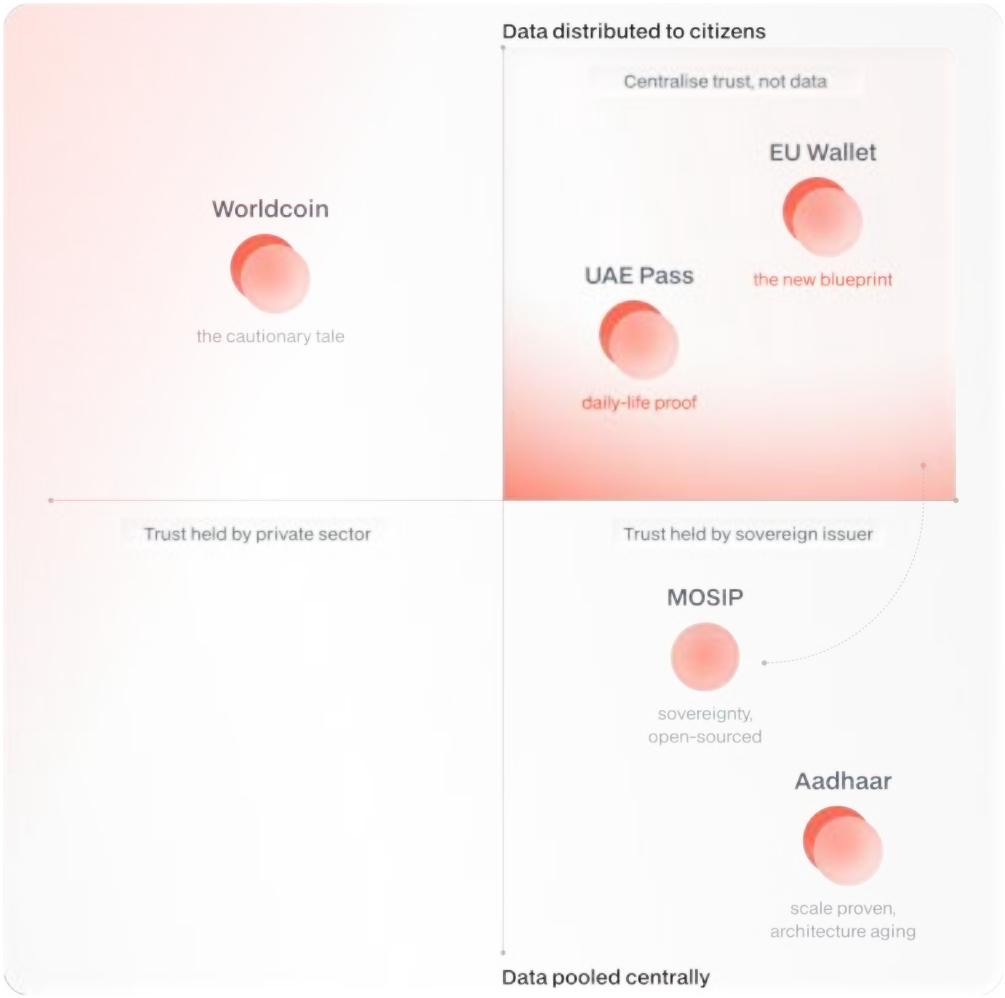
UAE Pass (United Arab Emirates) *proof it works in daily life*

The everyday demonstration. A resident verifies identity once when registering, and from then on opens bank accounts, signs documents, and accesses government services by approving a request on their phone. No counter, no fresh document check, no new photo. UAE Pass shows that the verify-once model survives contact with daily life at national scale, and that citizens adopt it readily when it removes friction rather than adding it.

Who Is Trying This

5.1 Why some work and some do not

Line the five up and one variable explains the outcomes. Aadhaar centralised trust and data, and inherited the honeypot. Worldcoin centralised trust in a private company, and inherited the legitimacy crisis. The EU wallet, MOSIP's direction of travel, and UAE Pass centralise trust in a sovereign issuer while keeping data distributed, and they are the ones compounding. The pattern has a name, and it is the thesis of this report: centralise trust, not data. One office checks who you are, once, properly. After that there is one source of truth, and no jackpot to rob.



> One source of truth, and no jackpot to rob.

6. What Should Happen

These are directional priorities, not point solutions, sized to the risk pathways in Section 3.

6.1 For nations and governments

- **Direct new infrastructure** toward verifiable identity and asset registries built on cryptographic proof rather than centralised biometric databases. This reduces the fraud surface and the mass-surveillance risk in the same move, and the window matters: systems specified today will face materially worse attack conditions within 18 to 24 months.
- **Share fraud intelligence** across borders. The absence of real-time coordination between African mobile-money operators and regulators is repeatedly identified as the single largest structural gap, and it is a policy fix, not a technology one. Ratifying and operationalising the Malabo Convention, currently at 14 of 55 member states, is the concrete first step.
- **Fund digital-literacy programmes** alongside any technical rollout. Survey data consistently shows verification habits lag adoption, especially in the markets with the highest AI optimism. Infrastructure without literacy leaves the front door hardened and the citizens soft.

6.2 For companies: banks, fintechs, govtech vendors

- **Move beyond single-factor liveness** around any high-value transaction or registration. Layer cryptographic attestation and out-of-band confirmation on top of face-match checks; a callback to a known number remains the single most effective control against the current generation of attacks.
- **Treat deepfake incident response** like a security discipline, not a fraud-ops afterthought: staff training, simulated attack drills, clear escalation paths. This is what separates organisations that contain losses from those that do not.
- **Build for the audit trail,** not just the transaction. Regulators are moving toward mandatory provenance and labeling, from the EU AI Act's content obligations to updated African KYC guidance. Systems that can demonstrate how an identity was verified will hold a compliance and trust advantage.

7. The Takeaway

Every serious response to deepfakes so far has been a **better camera**: sharper liveness detection, smarter forensics, faster model updates. That work matters, and it will keep mattering at the one moment where a camera must still stand guard. **But as a national strategy it concedes the terms of the fight.** The generator improves, the detector catches up, the generator improves again. A race, run forever, against an opponent whose costs fall every year.

The architectural response changes the question. It does not ask whether a fake face can be detected. It asks how much a fake face is worth, and then drives that answer toward zero: one door instead of thousands, nothing stored behind the door worth stealing, and no way to write history after the fact. No country is fully there yet. But every system examined in this report that is compounding rather than struggling is a partial assembly of the same blocks, and the blocks exist today.

Detection is a race. Architecture is a choice.

Centralise trust, not data.