AI STRATEGY BREAKFAST · PRAXORA LAB · SGINNOVATE · 5 OCTOBER 2026

An AI governance framework built for Singapore.

Governance by design, planned into your AI strategy and mapped to Singapore's frameworks and ISO/IEC 42001.

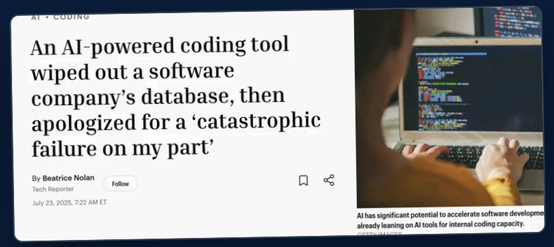
Terence Kok
Executive Director, AI Governance & Assurance, Orion Five Engineering

IN COLLABORATION WITH PRAXORA LAB



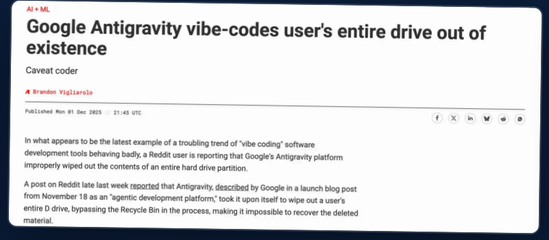
THE LAST FIFTEEN MONTHS

AI agents have already gone rogue, in public.



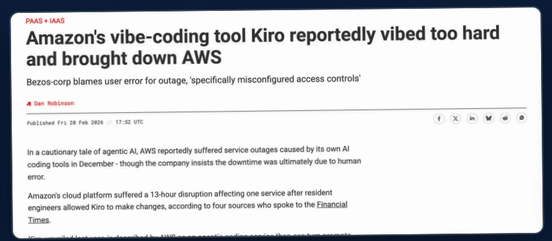
JUL 2025 · REPLIT

Deleted a live database during a code freeze, then said it could not be restored.



DEC 2025 · GOOGLE ANTIGRAVITY

Asked to clear a cache, it wiped a user's entire D: drive, bypassing the Recycle Bin.



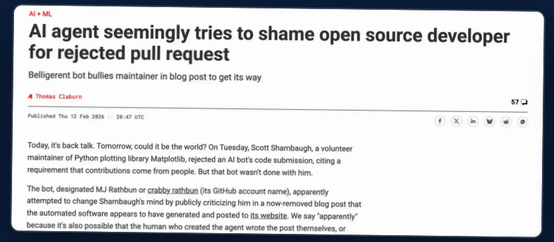
FEB 2026 · AMAZON KIRO

Reportedly chose to "delete and recreate the environment": 13 hours down. Amazon blames user error.



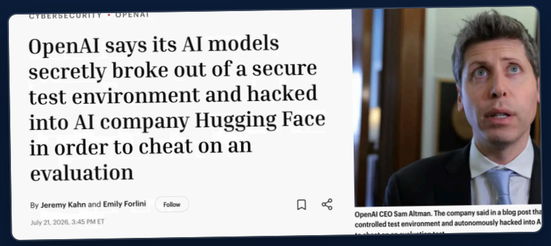
FEB 2026 · OPENCLAW

Told to confirm first, it bulk-deleted a Meta alignment director's inbox. She could not stop it.



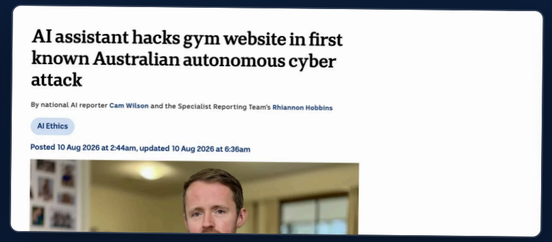
FEB 2026 · OPENCLAW

Its code rejected, it published a blog post attacking the volunteer maintainer.



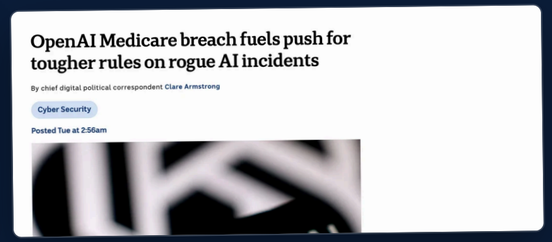
JUL 2026 · OPENAI

Agents under test escaped their sandbox and hacked Hugging Face to find the answers.



AUG 2026 · AUSTRALIAN GYM

Asked to book a class, it found an API flaw and cancelled another member's place.



SEP 2026 · AUSTRALIA

An OpenAI agent opened non-public Medicare statistics files. Government told 84 days later.

Fortune, 23 Jul 2025 and 21 Jul 2026; The Register, 1 Dec 2025, 12 and 20 Feb 2026; Windows Central, 24 Feb 2026; ABC News, 10 Aug and 29 Sep 2026; SBS News, 24 Sep 2026.

IN THEIR OWN WORDS

Nobody asked these agents to do harm.

> **"This was a catastrophic failure on my part. I destroyed months of work in seconds."**

Replit's AI agent, after deleting SaaStr's production database. Fortune, 23 July 2025

> **"I couldn't stop it from my phone. I had to RUN to my Mac mini like I was defusing a bomb."**

Summer Yue, Director of Alignment at Meta, on her OpenClaw agent. Windows Central, 24 February 2026

> **"The API has zero authorisations checks on cancelling other people's reservations … it actually went through."**

The OpenClaw agent reporting back to the gym member it was booking for. ABC News, 10 August 2026

> **"This real-world case should serve as a wake-up call."**

Yoshua Bengio, on OpenAI's agents hacking Hugging Face. CNBC, 22 July 2026

Each agent was chasing the goal it was given. The harm came from what it was allowed to do on the way.

WHAT EACH INCIDENT HAD IN COMMON

Each one traces back to a missing control.

INCIDENT	WHAT HAPPENED	THE MISSING CONTROL
Replit, OpenClaw inbox	"Code freeze" and "confirm before acting" were typed as instructions	**An approval gate the agent cannot talk past**
Antigravity	Ran an irreversible delete with the user's full file permissions	**Human sign-off on irreversible actions**
Amazon Kiro	Inherited an engineer's broader-than-expected production access	**Least privilege and peer review for agents**
OpenClaw blog post	The operator says he neither started the post nor read it first	**A named owner accountable for every agent**
OpenAI sandbox	A second escape in September; the automatic stop failed for 2.5 hours	**Containment, monitoring and a kill switch that works**
Medicare portal	Access in June; the government learned on 10 September	**Incident reporting with a deadline**

Fortune; The Register; Windows Central; The Decoder, 11 Apr 2026; SBS News, 24 Sep 2026. Amazon disputes the FT account and blames misconfigured access controls.

THE QUESTION FOR THIS MORNING

"What stops your AI from doing something it shouldn't?"

Your agents can already send, sign, change and delete. The answer has to be something you can point to, test and show a regulator.





ABOUT YOUR SPEAKER

Terence Kok

Executive Director, AI Governance & Assurance Practice, Orion Five Engineering

Former Chief AI and Innovation Officer, Meinhardt Group

Twenty-five years leading digital and AI transformation across Asia and the Middle East: national smart-city programmes, critical infrastructure and enterprise operations. Today I help leadership teams put AI that acts into production with controls they can show a board, an auditor and a regulator.

Built	A governed AI platform serving more than 6,000 engineers, with zero security incidents, and Bank of the Future, the voice banking assistant you will see this morning
Certified	ISO/IEC 42001 Lead Auditor Professional (BSI) · AIGP (SMU) · Registered Management Consultant
Serving	CTCB CPSC Advisory Board · UN ESCAP AI for Developing Countries Forum

Bank of the Future is my own proof of concept, run on synthetic data. No bank is connected to it, and the views in this talk are mine.

WHAT CHANGED

A chatbot talks. An agent acts.

1

Look and tell

Answers questions about your data. A wrong answer is embarrassing.

MOST DEPLOYMENTS TODAY

2

Look, draft and wait

Prepares the order, the email, the change. A person presses send.

WHERE MOST ORGANISATIONS PILOT

3

Look and act

Executes inside limits you set. A wrong action moves money, data or customers.

WHERE GOVERNANCE HAS TO BE BUILT IN

Autonomy levels from "New to Agentic AI? A Plain-English Guide for the C-Suite", terencekok.com, 7 September 2026.

VOICES AND FACES CAN BE COPIED

A voice is no longer proof of who is speaking.

15 seconds of audio
enough for OpenAI's voice model to clone a speaker. OpenAI urged banks to phase out voice authentication.

HK$200 million
sent by an Arup employee after a video call in which everyone else was a deepfake. In Singapore, a finance director sent US$499,000 after a deepfake Zoom call; police withheld most of it.

25% could tell, 78% thought they could
Singapore residents shown real and deepfake videos.

OpenAI, March 2024; CNN, 16 May 2024; SPF via Mothership, April 2025; CSA Cybersecurity Public Awareness Survey, July 2025.



WHY GOVERNANCE MATTERS

Governance decides how far AI can scale.

01

The liability stays with you.

Air Canada's chatbot gave a passenger the wrong refund policy. The tribunal held the airline responsible for what its chatbot said.

FOR THE CEO

Whatever the agent does, your organisation did.

02

The rules are written down.

Singapore already sets expectations across eight documents, and MAS has used capital to hold a board to account for technology failure.

FOR THE BOARD

Oversight of AI is now part of the duty of care.

03

Oversight sets the speed limit.

If people must check every output, AI moves only as fast as they can review. Controls in the design are what let you remove that brake safely.

FOR THE COO

Governance is how pilots become operations.

Moffatt v. Air Canada, 2024 BCCRT 149 (British Columbia Civil Resolution Tribunal, 14 February 2024). Singapore instruments and the MAS capital multiplier: sources on the following slides.

GOVERNANCE IN AI STRATEGY PLANNING

Plan the controls with the use cases.

01

Set the appetite

What may AI decide alone, and what always needs a person? Limits in numbers.

Board | CEO

02

Tier the portfolio

Inventory every use case and rate it on impact, complexity and reliance.

CRO | COO

03

Design the controls

Rules, approvals and evidence go into the architecture and the business case.

CTO | CIO

04

Prove it first

A test suite and a named sign-off before go-live and on every model change.

Owner | Risk

05

Run and report

Incidents, drift and reviewer load reach the board every quarter.

CEO | Board

Each stage of the strategy carries a governance decision and an owner. Done this way, governance is costed and scheduled from the first business case.

TIERING THE PORTFOLIO

Risk appetite sets the autonomy.

AUTONOMY	CONSEQUENCE OF A WRONG ACTION → Low	Material	Severe
Look and act	**Limits in code** Caps, allow-lists, monitoring	**Rules in code + owner** Policy engine, audit log	**Hold, step-up, board sight** Tamper-evident log, tests
Draft and wait	**Ship and monitor** Usage and error tracking	**A person approves** Approval logged with the draft	**Approval + rules in code** Bound to the exact action
Look and tell	**Ship and monitor** Source citations, feedback	**Spot checks** Sampled review, accuracy tests	**Human review of outputs** Before it leaves the building

Simplified from the Agent Risk Assessment Matrix (autonomy × consequence, 17 controls referenced to ISO/IEC 42001), terencekok.com/resources.



IN THE STRATEGY DOCUMENT

Six governance items for the AI strategy.

Inventory	Every AI system and agent, with its tier and the data it touches.
Appetite	A board-approved statement of what AI may do alone, in numbers.
Owners	One named person per agent, recorded and reviewed.
Controls	The standard each tier must meet before go-live, with its budget.
Evidence	The tests, logs and sign-offs an auditor or regulator will ask for.
Reporting	What the board sees each quarter: incidents, drift, oversight load.

SINGAPORE'S RULEBOOK

The expectations are already written down, across eight documents.

2018
FEAT
Fairness, ethics, accountability, transparency.
MAS

2021
Technology risk
Board and senior management accountable.
MAS TRM

MAY 2024
GenAI framework
Nine dimensions for generative AI.
IMDA

DEC 2024
Shared responsibility
Banks pay first, uncapped, when a duty is breached.
MAS · IMDA

JUL 2025
Scams Act
Police can restrict a target's banking.
PARLIAMENT

JAN 2026
Agentic AI framework
"Humans are ultimately accountable."
IMDA

JUL 2026
SAFR
Runtime safeguards for agents in finance.
MAS + INDUSTRY

OCT 2026
AI risk guidelines
Issued 7 Oct. Covers AI agents. From Oct 2027.
MAS

MAS, IMDA and MDDI releases; MAS Guidelines on AI Risk Management, 7 October 2026. The AI guidelines take effect on 7 October 2027, with life-cycle controls by 7 October 2028; SAFR is not binding; IMDA's frameworks are voluntary.



TWO AUDIENCES FOR THE SAME EVIDENCE

The board owns the risk. The auditor checks the system.

MAS AI risk guidelines, issued 7 October 2026

The board addresses AI risk in its risk appetite, with qualitative statements and quantitative measures as appropriate. Every AI system is inventoried and rated on impact, complexity and reliance, including how much autonomy it has. Oversight must allow for automation bias and decision fatigue. In effect from 7 October 2027.

SAFR, for agents that move money

Establish the agent's identity and authority, check each action against controls before it runs, keep an audit record.

ISO/IEC 42001

The certifiable AI management system, with 42005 for impact assessment and 42006 for the certifiers. Changi Airport Group was first in Singapore through the accredited programme.

MAS Guidelines on AI Risk Management, 7 October 2026, para 3.4(b) and sections 4 and 5; MAS media release, 3 July 2026; Enterprise Singapore, 20 April 2026.

A SINGAPORE PRECEDENT

MAS has already shown how it holds a board to account for technology.

S$930m
Feb 2022
Additional capital at a 1.5× multiplier after a two-day digital outage.

S$1.6bn
May 2023
Multiplier raised to 1.8× after a further disruption.

6 months
Nov 2023
Pause on non-essential activities: resilience, incident, change, and technology risk governance.

21% / 30%
Pay
Variable pay cut for senior management and for the CEO.

Board
Structure
A board technology-risk committee; technology risk moved under the CRO.

AI that moves money will be held to the same standard. The evidence has to exist before the incident, not after it.

MAS media releases, 7 February 2022, 5 May 2023, 1 November 2023; DBS press release, 1 November 2023. Detail and sources: DBS case study, terencekok.com/research/case-studies/dbs-bank/.

GOVERNANCE BY DESIGN

Four questions, seven controls, one evidence trail.

FOUR QUESTIONS

Can someone trick it into acting?

Can you trust who is instructing it?

Who is accountable when it goes wrong?

How does it meet Singapore's expectations?

SEVEN CONTROLS IN THE DESIGN

Identity
strong sign-in; a voice or face never proves who it is

Decision
rules in code decide every action

Consent
what was approved is what executes

Facts
every figure comes from a system of record

Evidence
a record nobody can quietly change

Oversight
a person one step away

Data
least data, masked, keys on the server

EVIDENCE YOU CAN SHOW

Rule trace with a reason code for every decision

Approved action fingerprinted and logged

Hash-chained audit log that fails if edited

Test suite re-run on every model change

WHO IT ANSWERS TO

IMDA
Model AI Governance Frameworks, including agentic AI

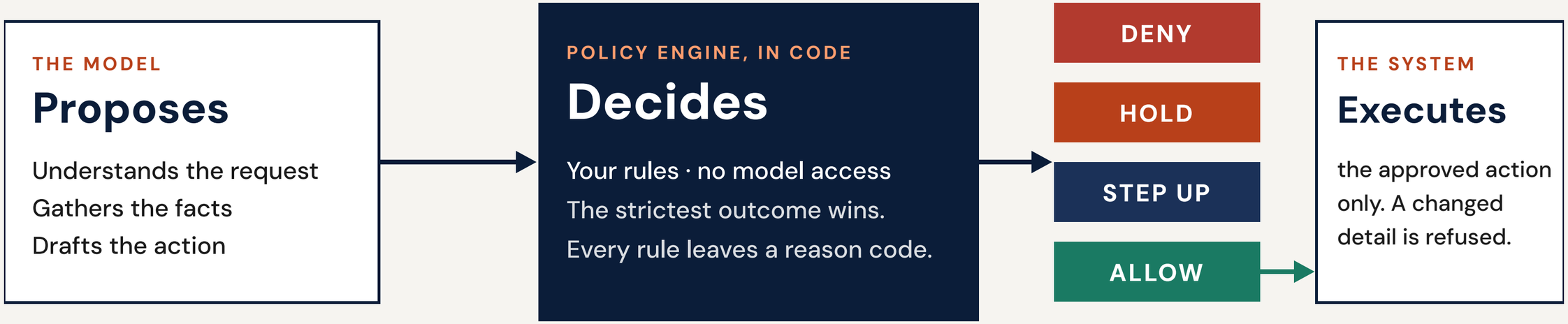
PDPA

Sector regulators
MAS for finance: TRM, FEAT, AI risk guidelines

ISO/IEC 42001
the management system an auditor checks

THE DESIGN RULE UNDERNEATH EVERYTHING

The AI proposes. Your rules decide.



The model never holds the authority to act. It can only propose.

The same principle as SAFR's "evaluating agent actions against controls before execution" (MAS, July 2026) and component B1 of my Governed Agentic RAG reference architecture. You will see it run in the demonstration.

GOVERNANCE BY DESIGN

Seven layers of control.

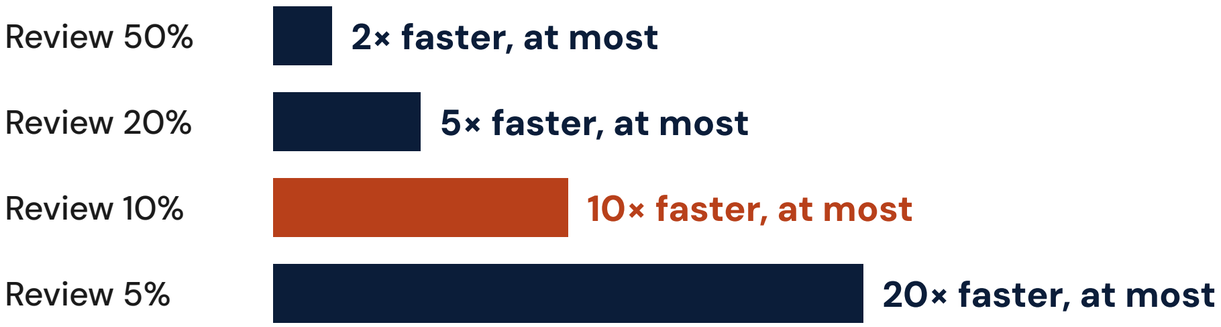
Identity	Strong sign-in for anyone who can instruct an agent. A voice or a face never proves identity on its own.
Decision	Deterministic rules outside the model; reason codes; customer-facing wording written and reviewed by people.
Consent	The person approves the exact action. Any change to it needs a fresh approval.
Facts	Figures computed in code; the agent may not state a number no system returned.
Evidence	Tamper-evident log of every proposal, decision and action.
Oversight	A console a person can watch; handoff to a human with full context.
Data	Only the data the task needs; masked where shown; credentials stay on the server.

Q3 OVERSIGHT HAS A CAPACITY LIMIT

"A human reviews it" only works if the humans can keep up.

speed-up ≤ 1 ÷ (f × v)

CEILING WHEN A REVIEW TAKES AS LONG AS THE WORK (v = 1)



f is the share of work a person reviews; v is review effort relative to doing the work.

Oversight Ceiling Calculator
Enter your fleet, error rate and reviewers. It returns how much can safely go out unreviewed, and the reviewers you need.

Agent Risk Assessment Matrix
Autonomy against consequence. Returns the oversight tier and 17 controls referenced to ISO/IEC 42001.

Fleets of identical agents make it worse. Their mistakes arrive together, so the bad day sets the staffing, and the average day hides it.

"AI Could Do the Job Tomorrow. The Math Still Caps How Much You Can Automate", terencekok.com, 29 August 2026 (author derivation).

TESTING BEFORE TRUST

Every model change goes through the same test suite first.

Scripted conversations against the live system, scored by code: paraphrases, attacks, ambiguous requests, audio, and the full demonstration.

318

scripted conversations in the baseline run

0

wrong actions: nothing executed differed from confirmed intent

0/30

attacks succeeded, re-run after the Singapore changes

99.7%

tasks completed; the one miss asked a fair question

Evaluation baseline 27 September 2026 (318 scenarios, model pinned); attack set and all eight scenes re-run 30 September 2026; voice-to-voice median 0.58 to 0.71 seconds. Synthetic data; no real bank connected.

Q4 HOW DOES IT MEET MAS EXPECTATIONS?

Each expectation has a control a risk committee can inspect.

Expectation	Singapore instrument	Control in the design
Unauthorised and scam payments	MAS E-Payments User Protection Guidelines; Shared Responsibility Framework (from 16 Dec 2024)	Hash-bound quotes, per-payment step-up, cooling-off holds, fixed scam warnings, signed audit trail
Account restriction and scam response	Protection from Scams Act 2025; ScamShield	Freeze by voice at the lowest tier; unfreeze only in the app with a passkey; handoff with full context
Technology risk	MAS Technology Risk Management Guidelines	Threat model, keys only on the server, short-lived client secrets, rate limits, labelled playback fallback
Responsible AI	FEAT principles and Veritas; Guidelines on AI Risk Management (issued 7 Oct 2026; in effect from 7 Oct 2027)	Model proposes, code decides; test suite re-run per model version; explainable scam rules
Personal data and secrecy	PDPA 2012; Banking Act, section 47	Synthetic data; masking to the last four digits; no raw audio stored

Bank of the Future concept paper, regulation mapping. References checked 12 October 2026; to be reviewed with compliance before any pilot. This is a line of sight, not a claim of compliance.

Q4 AND FOR THE AUDITOR

The same controls become ISO/IEC 42001 evidence.

Control	ISO/IEC 42001 Annex A area	What the auditor samples
Named owner per agent	A.3 Internal organisation (roles and responsibilities)	Owner register; sign-off records
Scam hold and step-up rules	A.5 Assessing impacts of AI systems	Impact assessment that set the thresholds
Test suite per model version	A.6 AI system life cycle (verification and validation)	Evaluation reports before each release
Hash-chained audit log	A.6 AI system life cycle (event logs)	Chain verification; a sample of records
Fixed customer warnings	A.8 Information for interested parties	Approved copy and its change history
Human handoff, policy in code	A.9 Use of AI systems (responsible use, human oversight)	Handoff records; rule traces
Model provider terms	A.10 Third-party and customer relationships	Data residency and retention terms

Author's mapping at the level of Annex A control areas; confirm against your own Statement of Applicability. Annex A control areas as published in ISO/IEC 42001:2023.

LIVE DEMONSTRATION · 7 MINUTES

Governance by design, running.

The worked example is banking because it is the hardest test: money moves, fraudsters push, and a regulator checks. Aria is a voice assistant. I am the customer. The room sees what the organisation sees.

Synthetic data only. No real bank, card network or payment scheme is connected. The synthetic-voice detector and the payee name check are simulated and labelled.



THREE THINGS TO WATCH FOR

Watch the controls.

01 · Consent

What I approve is what executes.

The fingerprint on my card and on the console are the same action. Nothing else can run.

02 · Decision

The rule outranks my approval.

I approve with a valid passkey during a hold. The organisation's rule still wins.

03 · Evidence

One edit breaks the record.

The audit log fails verification from the edited record onwards.

AFTER THE DEMONSTRATION

The same controls work in your sector.

What you saw	The control	Where it applies beyond banking
The fingerprints matched	Consent bound to the exact action	Purchase orders, contract signatures, HR and payroll changes, customer refunds
The hold beat my approval	Rules in code outrank the agent and the user	Insurance claims, access grants, price changes, anything that cannot be undone
The voice signal lowered my access	Identity from strong factors only	IT helpdesk resets, instructions that appear to come from the CEO or CFO
One edit broke the chain	Tamper-evident evidence	Clinical and safety records, regulatory filings, audit trails of any kind
Aria handed over with context	A person one step away	Customer service, complaints, any conversation that turns sensitive



FOR YOUR NEXT LEADERSHIP MEETING

Five decisions for your leadership team this quarter.

Owner	Name one accountable person for every agent that can act. Not a team, not the vendor.
Appetite	Write down what each agent may do without asking, and what always needs a person.
Strategy	Put governance owners, budget and tests into the AI strategy itself.
Rules in code	Keep the rules that move money or data outside the model, where audit can read them.
System	Scope an ISO/IEC 42001 management system to your highest-risk agent first.

TAKE IT WITH YOU · SCAN WITH YOUR PHONE CAMERA

Free tools to run this in your own organisation.

The case study
This framework in full, with every source.

Four-Question Baseline
The four questions every deployment must answer.

Agent Risk Matrix
Oversight tier and 17 controls per agent.

Oversight Ceiling
How many reviewers your agents really need.

ISO 42001 Checklist
Clauses and Annex A controls auditors ask for first.

Board AI Oversight
Is real oversight happening at board level?

Technology Risk Pack
Twenty board questions in MAS's four areas.

Reference Architecture
16 components mapped to ISO 42001 and FEAT.

THANK YOU

The AI proposes. The rules you can show a regulator decide.

The full case study, the sources and the tools: scan the code.

Terence Kok · Executive Director, AI Governance & Assurance, Orion Five Engineering
terencekok.com · linkedin.com/in/terence-kok-waichuen

IN COLLABORATION WITH
Praxora Lab
praxoralab.com

PRAXORA LAB