Governance That Can't Be Quietly Undone

Tamper-evident community and kāhui Māori governance — and the AI rules of Aotearoa New Zealand and Australia

Précis. New Zealand and Australia have both, deliberately, declined to pass prescriptive AI legislation. What they have instead is *principle*: charters, frameworks, voluntary standards, and the general weight of existing privacy law. Principles are only as good as their enforcement, and the usual enforcement is a promise in a policy document — the kind that can quietly drift. This essay describes a different answer, now substantially built: a community-governance platform in which the principles those rules ask for — transparency, human oversight, accountability, auditability, data sovereignty — are **enforced in the architecture**, not asserted in prose. It then shows what that makes possible: how a governance village, and a kāhui Māori village, can actually run their board deliberations, meetings, and votes on a substrate where every decision is signed, every authority is answerable, and nothing important can be changed without leaving a trace. Implemented and in-development features are kept distinct throughout.

The rules are softer than they look

It is tempting, reading the headlines, to picture Aotearoa and Australia legislating hard limits on government AI. Neither has. The precise legal picture matters, and overstating it is a temptation worth resisting.

In New Zealand, the governing instruments are principle-based and, for the most part, *not binding law.* The **National AI Strategy** ("Investing with Confidence," MBIE, July 2025) is an adoption-focused, principles-led roadmap, explicit that it prefers guidance to "a new prescriptive regulatory regime." The **Public Service AI Framework** that sits within it is, in its own words, encouragement: "Agencies are encouraged to align with the direction set by this Framework, however it's not binding." It carries five principles — inclusive and sustainable development; human-centred values, with "human oversight throughout the AI lifecycle"; transparency and explainability, including public disclosure "when AI systems are used"; safety and security; and accountability, meaning "oversight by accountable humans … at every stage," with "re-

porting, auditing and/or independent reviews." The **Algorithm Charter for Aotearoa New Zealand** is a *voluntary commitment* by agencies to use algorithms fairly and transparently. The one hard-law anchor is the **Privacy Act 2020** (amended in 2025), which applies to AI as to anything else.

In Australia, the trajectory is even more telling. In September 2024 the government *proposed* ten mandatory guardrails for AI in high-risk settings — built around testing, transparency, and accountability — alongside a **Voluntary AI Safety Standard**. Then, in December 2025, after consultation, the **National AI Plan shelved the mandatory regime** — declining, "at this time," to create AI-specific obligations and falling back on existing technology-neutral law and voluntary guidance. The hard-law anchor, again, is general: the Privacy Act and the laws that already apply regardless of the technology.

So the summary is this: in both countries, the *expectations* are real and converging — they track the OECD AI Principles — but the *enforcement* is mostly soft. An agency or a community is asked to be transparent, to keep a human accountable, to keep records, to be fair, to respect privacy. Whether it does is, for now, largely a matter of good faith and self-report.

That gap — between a principle and a guarantee — is exactly the space this work occupies.

The wider legal context (Aotearoa, 2023–2026)

AI itself remains unlegislated in Aotearoa — but the law around the *integrity* of information and decision-making has moved quickly since 2023, and a paper on tamper-evident governance should say so plainly.

The **Privacy Act 2020** — the hard-law anchor named above — was amended by the **Privacy Amendment Act 2025**. A new principle, **IPP 3A**, in force since **1 May 2026**, requires an agency that collects personal information *indirectly* — from someone other than the person concerned — to take reasonable steps to make that person aware of it. It is, in effect, a transparency-of-provenance rule, and it sits naturally with an architecture that already records where each record came from. The **Biometric Processing Privacy Code 2025** tightens the rules on facial recognition and other biometric processing, with organisations already using it required to comply by **3 August 2026**; the Village is not a biometric processor, but the direction of travel is unmistakable.

The sharpest development is the **Crimes (Countering Foreign Interference) Amendment Act 2025** (in force **27 November 2025**). It created two new offences — foreign interference, carrying up to **14 years**, and an offence committed to benefit a foreign power, up to **10 years** — and widened the existing offences for wrongful communication, retention, or copying of official information to cover **local government and the Offices of Parliament**. This is criminal law aimed chiefly at official information, not a regulation of AI or of community data, and it would be a misreading to present it as either. But

it sets the surrounding context, alongside New Zealand's **National Security Strategy 2023–2028**, which names foreign interference and espionage among its core concerns and asks that resilience be built not only in government but across business and communities. An architecture whose records are tamper-evident, whose authority is answerable, and whose operator cannot read across tenants narrows the surface for covert repurposing as much as it resists a foreign legal order. The integrity controls that were merely "good governance" are increasingly the posture a national-security strategy asks of communities, too.

What good governance is actually being asked for

Strip the instruments down and the same handful of expectations recur, on both sides of the Tasman:

- **Transparency** — disclose when and how AI is used.
- **Human oversight and accountability** — a named human decides the consequential things; the system does not.
- **Record-keeping and auditability** — decisions can be traced, reviewed, and independently checked.
- **Fairness and contestability** — those affected can understand and challenge a decision.
- **Privacy and data protection** — under the Privacy Acts, and, in Aotearoa, under Te Tiriti for Māori data.
- **Risk-proportionality** — heavier scrutiny where the stakes are higher.

None of these is controversial. The difficulty is that, written as policy, each can be honoured on Monday and eroded by Friday — not by any decision anyone defends, but by drift. The question worth asking is not "does the policy say the right thing?" but "what in the system makes the right thing *hold*?"

Make the principle structural: the tamper-evident layer

The platform described here — the Village — answers that question by moving the load-bearing commitments out of policy and into architecture. The following are **implemented and in production** unless marked otherwise.

Every record carries its own provenance. Content records are *sovereign records*: each carries embedded metadata about origin, policy, and an append-only, signed **proof chain**. The chain is guarded in the data layer — external attempts to `$pull` or `$set` it through the application's models are rejected, and only the plugin's own signed append is allowed. Each entry is signed with the tenant's own per-tenant Ed25519 key, whose public half is published in the tenant's DID document, so a decision's history is reconstructable — and independently checkable — from the community's own data without trusting the platform operator. This is *tamper-evident*, the word the platform is careful to use — not "immutable" or "court-proof": the guard lives in the application's persistence layer rather than inside the database engine, and the signatures are

the tenant's own, not a third-party notarised timestamp.

A constitutional floor no one can override. A `BoundaryEnforcer` service keeps the AI inside a small set of boundaries enforced in code, not policy: it is built to ensure the AI "never makes values decisions without human approval," so value-laden and governance questions are routed to the responsible humans, and the AI presents options rather than deciding. Above it sits a universal rule layer whose principles, in the platform's own words, "cannot be overridden by any tenant configuration." The stated design goal is governance "structures that operate independently of the AI and cannot be overridden by it" — which is the structural meaning of "a human decides," the thing both countries' frameworks ask for.

The AI cannot reach the rules that bind it. Community-defined instructions live in a separate persistence layer the model cannot access or modify; outputs are checked after generation, and conflicts resolve in favour of the stored instruction. Every AI response passes a six-stage pipeline (classification, boundary enforcement, pressure/uncertainty monitoring, metacognitive verification, cross-reference validation, and pluralistic deliberation). The guardians that enforce rules are **deterministic** — explicit rules and thresholds, not learned models — which makes them reproducible and auditable, and closer to deterministic rule-checkers than to the probabilistic systems AI regulation is chiefly aimed at.

Auditability as a property, not a feature. A `GovernanceAuditLog` records governance decisions — which rules were checked, the outcome, the service, the timestamp — and rule changes are logged with before-and-after state. The durable, long-term decision trail lives in the signed proof chain on the record itself; the audit log adds a queryable enforcement history on top of it. Either way, a member's or regulator's "how was this decided?" has an answer that does not rest on anyone's memory. This is the "reporting, auditing and/or independent reviews" of the NZ Framework, and the record-keeping theme of the Australian guardrails, made native.

Data sovereignty by construction. Hosting is **EU/NZ only** — OVH (France) and Catalyst Cloud (Auckland) — with zero US data-processing footprint, and AI inference run locally rather than sent to external providers. Content is encrypted with per-record keys; plaintext is unwrapped only inside the tenant's own request and never cached or exported. **Cryptographic deletion** destroys the per-record key so the ciphertext is unrecoverable even by the operator with full database access — and the deletion itself leaves a signed tombstone, so erasure is *evidenced, not erased without trace.* Because every query is automatically scoped to the current tenant — a framework-level filter that fails closed when no tenant context is present — and because the content is encrypted under per-tenant keys the operator cannot read, an operator served a foreign legal order cannot be compelled to disclose what it cannot read. That is the Privacy Act's cross-border concern, and Te Tiriti's data-sovereignty concern, answered mechanically rather than promised.

Set against the six expectations above, each maps to a shipped mechanism rather than a clause. That is the whole point: where the rules are principles, the differentiator is making the principles **un-undoable**.

How a governance village actually runs

The wider question is what this substrate makes possible day to day — how a community that exists *to govern* (a board, a committee, a council, a joint deliberation between constituencies) actually conducts its business on it.

A constitution before anything else. Every tenant must populate its sovereign constitution sections — its conflict-resolution policy, its values, its federation posture — before it can create content. This is hard-gated in code: until the constitutional journey is complete, content creation is refused with a 403 that names the missing sections. Governance is not bolted on afterward; it precedes the infrastructure, which is then built to obey it.

Deliberation as a signed record. Deliberation is a sovereign record in its own right: each contribution and state change carries a signed, append-only proof-chain entry, so the sequence of a debate and its eventual outcome are reconstructable from the record itself, in order. Visibility is scoped — a deliberation can be confined to a named subgroup, enforced at the read boundary, not by convention.

Voting with the integrity built in. Polls are sovereign records too, and support three attribution modes — **attributed, anonymous, and roll-call** — so a board can take a secret ballot on a personnel matter and a named vote on a constitutional change. Quorum is fixed against a membership snapshot captured and locked when the poll opens, and immutable thereafter — so that members added or removed while a vote is open cannot shift the threshold under it. The snapshot is schema-enforced: a governance poll cannot be created without it, and it cannot be altered once captured. A pluralistic-deliberation advisory layer can flag procedural problems (for example, a main motion voted before its amendments) and recommend a threshold for the matter at hand — but it *recommends*; the chair decides. *In development:* full minutes export (motion → second → amendment → tally → decision → action items as signed PDF/Markdown), motion-sequencing enforcement, and conflict-of-interest prompts are designed and partly built, not yet general.

A governance queue with deadlines. Consequential decisions move through an explicit lifecycle — create → acknowledge → decide → enact (or reject) — with deadlines enforced by a scheduled task, so nothing consequential sits unowned forever.

Authority that is plural and answerable. Roles are owner, moderator, and member, with approval workflows where a change is proposed and then ratified. Above the tenant, governance is **polycentric**: multiple authorities — platform, iwi, community trust, tenant — each publish their steering, and

any can **withdraw** it; when an authority withdraws a rule, the system must stop relying on it, and the withdrawal is logged. No authority is sovereign over the others; legitimacy is composed, never concentrated. Rules can be authored, edited, translated (DE/FR/NL/MI), shared between communities with attribution preserved, and deleted — every action audited.

This is not a separate "governance" product so much as a configuration any village can switch on: the **committee** and **governance** demonstrators are live (a small-group committee with voting and minutes; a multi-stakeholder deliberation across constituencies — the worked example is a council and a school board deliberating jointly). The machinery beneath them — deliberations, polls, the governance queue, the proof chain, the audit log — is the shipped part; the richer meeting apparatus is being finished on top of it.

Kāhui Māori villages

For a kāhui Māori village the same substrate carries a heavier load, and it is built to. I describe what the platform *supports*; it is scaffolding for Māori-led governance, not the platform speaking for anyone.

Genealogical records — whakapapa — are treated not as a database field but as **taonga**, to which iwi hold authority under Article 2 of Te Tiriti. Each such record carries mandatory stewardship metadata: who recorded it, *who is its kaitiaki*, and the tikanga under which it was shared. Disclosure is governed by that tikanga rather than by platform policy: a record can be marked whānau-internal, confined to a named hapū or marae-rōpū, restricted to recorder-and-kaitiaki only, shared with a named iwi under a bilateral agreement, or never shared at all — and an attempt to read outside the named scope is refused at the route boundary. Content a community's own cultural authority marks as restricted escalates to a human rather than being answered by the AI.

Cross-iwi sharing is **bilateral only** — two iwi who choose to share a bounded interaction do so under a bilateral federation agreement, each retaining full revocation control. There is no platform-wide federation graph, no Crown-mediated network, no central register of who shares what with whom. And because every query is tenant-filtered, the operator structurally cannot read across iwi — the same property that defeats a foreign legal order also defeats inadvertent cross-iwi exposure. Te reo Māori is carried in the platform's vocabulary system rather than bolted on as translation, governance rules can be authored and held in te reo, and training data is governed under Karaitiana Taiuru's Kaupapa Māori AI Framework. The **kāhui Māori** demonstrator — a multi-rōpū federation under a shared kaupapa — is live as a demo; production iwi-to-iwi federation awaits a counterparty agreement, which is as it should be: that is a decision for iwi, not a platform.

A standing discipline runs through this: the generic governance tooling is kept culturally distinct from the Māori track. Tikanga-specific surfaces (a hui minutes template, for instance) are gated behind cultural-authority sign-off and are

not shipped without it.

What this meets, precisely

What the rules ask	Where it lives in NZ / AU	The Village mechanism	Status
Disclose when/how AI is used	NZ Framework (transparency); AU guardrails (transparency)	Per-inference provenance; members see which authorities shaped an output	Shipped
A human decides the consequential things	NZ (human oversight, accountability); AU (accountability)	BoundaryEnforcer constitutional floor; value-laden queries routed to humans	Shipped
Records, audit, independent review	NZ (reporting/auditing); AU (record-keeping)	Signed proof chains; GovernanceAuditLog; reconstructable decision trail	Shipped (full minutes export in development)
Fairness, contestability	NZ (human-centred values); AU (contestability)	Deterministic, reproducible guardians; decision provenance a member can inspect	Shipped
Privacy & data protection	Privacy Act 2020 (am. 2025, IPP 3A); Privacy Act (Cth)	Per-record encryption; cryptographic deletion; EU/NZ hosting; tenant isolation; recorded provenance	Shipped
Māori data sovereignty	Te Tiriti Art. 2; Māori Data Sovereignty	Whakapapa-as-taonga; kaitiaki attribution; tikanga-scoped disclosure; bilateral federation	Shipped (formal legal opinion pending)

What the rules ask	Where it lives in NZ / AU	The Village mechanism	Status
Risk-proportionality	AU high-risk framing; EU AI Act	EU AI Act self-assessment: limited-risk, zero high-risk systems; heavier gating where stakes rise	Assessed (v1.0, Mar 2026)

The limits, stated plainly

This would not be worth reading if it pretended to be finished. Several of the governance-meeting features — full signed minutes export, strict motion-amendment sequencing, conflict-of-interest recusal prompts, a formal constitutional-amendment workflow, vote revocation — are designed and partly built, not general. The Te Tiriti compliance position is published for feedback (v0.2) with a formal legal opinion still to come. Post-quantum cryptography and hardware-backed keys are roadmap, not running. And "tamper-evident" is the correct, modest claim: the proof chains are signed with the tenant's own keys, which is strong against an operator and reconstructable by a regulator, but is not the same as a third-party notarised, court-proof timestamp — a gap the platform names rather than hides.

None of that undercuts the core claim. In a regulatory environment that has chosen principles over hard law, the thing that distinguishes real governance from governance theatre is whether the principles are *enforced where they cannot quietly be undone*. Aotearoa and Australia are asking communities and agencies to be transparent, accountable, auditable, and respectful of data sovereignty. This platform's answer is to make those properties of the architecture — so that meeting the rules is not a quarterly attestation but the default behaviour of the system the community already runs on.

That is the quiet difference. Where a policy can drift, a proof chain cannot be silently rewritten; where a promise can lapse, a constitutional floor holds; where an operator could be compelled, one that cannot read the data cannot disclose it. The rules may be soft. The governance need not be.

Sources

New Zealand

- New Zealand's Strategy for Artificial Intelligence: *Investing with Confidence* (MBIE, July 2025). https://www.mbie.govt.nz/
- Public Service AI Framework (NZ Digital Government) — non-binding guidance; five principles. https://www.digital.govt.nz/standards-and-

guidance/technology-and-architecture/artificial-intelligence/public-service-artificial-intelligence-framework
- Algorithm Charter for Aotearoa New Zealand (voluntary commitment). https://www.data.govt.nz/use-data/data-ethics/government-algorithm-transparency-and-accountability/algorithm-charter/
- Privacy Act 2020. https://www.legislation.govt.nz/act/public/2020/0031/latest/LMS23223.html
- Privacy Amendment Act 2025 (No 53) — IPP 3A (notification of indirect collection), in force 1 May 2026. https://www.legislation.govt.nz/act/public/2025/0053/latest/whole.html
- Crimes (Countering Foreign Interference) Amendment Act 2025 (No 71) — in force 27 November 2025; foreign-interference offences (14 yrs / 10 yrs) + wrongful-communication offences extended to local government and the Offices of Parliament. https://www.legislation.govt.nz/act/public/2025/0071/latest/LMS1003049.html
- Biometric Processing Privacy Code 2025 (Office of the Privacy Commissioner) — existing-processing compliance by 3 August 2026. https://www.privacy.org.nz/privacy-principles/codes-of-practice/biometric-processing-privacy-code/
- New Zealand's National Security Strategy 2023–2028, *Secure Together — Tō Tātou Korowai Manaaki* (DPMC, 2023). https://www.dpmc.govt.nz/our-programmes/national-security/new-zealands-national-security-strategy
- Te Mana Raraunga — Principles of Māori Data Sovereignty (2018). https://www.temanararaunga.maori.nz/tutohinga

Australia

- Introducing mandatory guardrails for AI in high-risk settings — proposals paper (Australian Government, Sept 2024); and the Voluntary AI Safety Standard. https://www.homeaffairs.gov.au/reports-and-publications/submissions-and-discussion-papers/introducing-mandatory-guardrails-for-ai-in-high-risk-settings
- National AI Plan (Dept of Industry, Science and Resources, Dec 2025) — mandatory AI-specific guardrails not proceeding "at this time"; reliance on existing technology-neutral law + voluntary guidance. https://www.industry.gov.au/
- AI Policy Corner: from proposed mandatory guardrails to the National AI Plan (Montreal AI Ethics Institute). https://montrealethics.ai/ai-policy-corner-from-proposed-mandatory-guardrails-to-the-national-ai-plan-ai-governance-in-australia/

Implementation detail in this essay is drawn from the Village platform's own architecture and governance documentation; features are marked implemented or in-development accordingly.

The Village platform and the Tractatus framework are an attempt to make governance achievable for communities at human scale — by building the principles in, where they cannot be quietly undone.