A Civil-Society Proposal for Sovereign and Federated Agentic AI in Aotearoa New Zealand

John G. Stroh / My Digital Sovereignty Ltd

Civil-Society Proposal · v1.2 May 2026 draft | Constructive parallel to CAC 2026 Implementation Guidelines →

v1 (superseded) → Email feedback

v1.2, 2026-05-16: revised per Ted Howard's correspondence on v1.1, focused on the substrate-vs-runtime distinction. §0(i) primitives section adds a clarifying paragraph naming the substrate-layer (PKI, federation, portability) as architecturally distinct from the runtime-layer primitives the six §0(i) services constitute; §0(i) boundary enforcement adds the four-category fallibility framing and the trinary router-output (allow / deny / escalate); Appendix A obj-3 and obj-5 amplify the substrate-vs-runtime separation and the survival-posture independence from agent containment. The substantive architectural detail is developed in *Architectural Alignment* §3.3, §3.4, §3.5, §7.4, §7.5 and *Paper A* §5.3. v1.1 remains accessible at its URL for historical reference.

v1.1, 2026-05-14: revised same-day per Dr Karaitiana Taiuru's feedback on v1. §0(iii) now cites Taiuru's 20 Sep 2025 critical analysis of Te Mana Raraunga; cites Te Kāhui Raraunga as the currently recognised operative body; adds explicit gap analysis. v1 remains accessible at the v1 URL for historical reference. Comments engaging specific sections welcomed. Please cite section numbers (e.g. §III item 5). The author replies personally; allow one to two weeks.

A Civil-Society Proposal for Sovereign and Federated Agentic AI in Aotearoa New Zealand

v1.2 May 2026 — draft research paper (revised per Ted Howard's correspondence on v1.1; substrate-vs-runtime distinction, four-category fallibility, trinary router output). Constructive parallel to the People's Republic of China's 2026 Implementation Guidelines on Intelligent Agents.

John G. Stroh / My Digital Sovereignty Ltd

2026-05-16

A Civil-Society Proposal for Sovereign and Federated Agentic AI in Aotearoa New Zealand

v1.2 May 2026 — draft research paper (revised per Ted Howard's correspondence on v1.1; substrate-vs-runtime distinction, four-category fallibility, trinary router output; see About this paper)

A civil-society proposal from My Digital Sovereignty Ltd, presented to New Zealand policymakers, community organisers, and sector practitioners. Constructed as a constructive parallel to the People's Republic of China's 2026 Implementation Guidelines for the Standardised Application and Innovative Development of Intelligent Agents, hosted in English translation at /research/translations/.

About this paper

This is the **v1.2 May 2026** draft of a civil-society proposal from My Digital Sovereignty Ltd, offered to New Zealand policymakers, community organisers, and sector practitioners. Comments are welcome via the standing paper-comments channels on agenticgovernance.digital; revisions in response to comments will be published as v2.

v1 → v1.1 changelog (2026-05-14): v1 was published earlier on 2026-05-14 and immediately reviewed by Dr Karaitiana Taiuru, who flagged that v1's foundational citation of Te Mana Raraunga's 2016-2018 Māori Data Sovereignty principles is superseded for AI contexts (per his 20 September 2025 *Critical Analysis of Te Mana Raraunga Data Principles*). v1.1 revises §0(iii) to cite Taiuru's critical analysis directly; to cite **Te Kāhui Raraunga** (the currently recognised operative body for Māori data governance in Aotearoa NZ, established 2019) and its published Māori Data Governance Model and Māori AI Governance Framework as the

current articulations; to adopt Taiuru's preferred grounding terms (mana motuhake, rangatiratanga) where appropriate; and to add an explicit gap-analysis subsection naming what this proposal does and does not yet do in the te ao Māori dimension. Items 4, 23, 37, §I principle 2, and §VI carry the same citation update. The architecture this proposal specifies is unchanged from v1. v1 remains accessible at `/papers/aotearoa-nz-agentic-ai-framework-v1-may-2026.html` for historical reference; this URL serves v1.2.

v1.1 → v1.2 changelog (2026-05-16): v1.1 was reviewed by Ted Howard in correspondence; he flagged that v1.1's framing of the six §0(i) primitives as architectural safety mechanisms is most defensible when the substrate-vs-runtime distinction is made explicit. v1.2 adds a clarifying paragraph after the §0(i) framework-primitives list naming the substrate layer (PKI / federation envelopes / portable records, developed in *Paper A* and *Architectural Alignment* §3.4) as the architectural sibling to the runtime layer the six §0(i) services constitute. §0(i) boundary enforcement gains the four-category fallibility framing (the categories are community-negotiated and appealable, not fixed essences) and the trinary router-output note (allow / deny / escalate-to-human). Appendix A obj-3 (route-around) and obj-5 (runtime-service-exploit) gain cross-references to *Architectural Alignment* §3.4 substrate-vs-runtime, §7.4 survival posture independent of agent containment, §7.5 social-layer attack surface (open frontier), and Paper A §5.3 PQC migration horizon. The architecture this proposal specifies is unchanged from v1.1; the revisions are clarifying framings that surfaced through reviewer dialogue. v1.1 remains accessible at `/papers/aotearoa-nz-agentic-ai-framework-v1.1-may-2026.html` for historical reference.

v1.2 same-day operationalisation pass (2026-05-16 evening): §III(II) Item 10 (blast-radius mechanism), §III(III) Items 11–12 (polycentric governance grounding), §III(IV) Item 13 (federation mechanics), §I Principle 4 (architecturally-enabled adoption evidence), and §IV (Adoption-Driven Development) — all nineteen sector items — given primitive-capability grounding so policy actors can advocate for the substantive sovereignty claims in committee. Architecture-generic register: names **cryptographic provenance**, **federation envelopes**, **member-driven portability**, and **boundary enforcement** by capability rather than MDSL implementation. §IV gains a leading paragraph naming the four substrate primitives once; each sector item then operationalises the sector-specific manifestation only. Substance unchanged from v1.2 morning; the upgrade makes the sovereignty claims operationally legible to readers who would advocate for them in committee rooms.

v1.1 same-day clarity revision (2026-05-14 evening): §0(i) framing paragraph revised to lead explicitly with the system-level / model-level distinction (system-level primitives, code-level runtime checks, substrate-agnostic across transformer-LLMs / JEPA-style / hybrid architectures). BoundaryEnforcer paragraph wording "by architecture rather than by hope" → "by runtime intercept rather than by hope" to reduce ambiguity for engineer-class readers familiar with the LLM-alignment debate. Substance unchanged; this is a wording revision for accessibility. Triggered by a technical reader pattern-matching the §0(i) primitives to model-level alignment claims they aren't making.

v1.1 Level 2 same-day clarity revision (2026-05-15): each of the six §0(i) primitives received a concrete engineering analogy (OS kernel privileged syscalls / circuit breaker / runtime check vs training-time alignment / configuration vs runtime arg / verification gate at runtime boundary / coordination service). Metacognitive verification primitive reframed from "requires agents to check their own reasoning" to "places a verification gate before action execution" to remove the residual model-level read. Substance unchanged; wording revisions for accessibility.

v1.1 Level 3 same-day clarity revision (2026-05-15): Appendix A added — "Common technical objections + responses" — six objection-response pairs covering LLM-enforcement skepticism, substrate-agnosticism, agent-bypass, prompt-engineering equivalence, runtime-service-exploit, and values-evolution. Same-day extension of L1 + L2 substantive work; no

claims beyond §0(i) primitive specifications.

This paper is **not** New Zealand Government policy. It is **not** Crown-endorsed. It is **not** Treaty-grounded in any formal sense. It is a civil-society proposal from My Digital Sovereignty Ltd, offered to NZ stakeholders as a basis for adoption, adaptation, or rejection. Where its principles are useful to the adopter's own work, they are free to use under permissive open-source licences; where they are not, they remain on the page.

The mirrored source structure is published in English translation at `/research/translations/china-cac-implementation-guidelines-2026.html` and originally as the Cyberspace Administration of China's 2026 *Implementation Guidelines* in Mandarin.

Abstract

This paper proposes a sovereign, federated framework for the application and development of intelligent agents in Aotearoa New Zealand, offered as a civil-society contribution by My Digital Sovereignty Ltd. The proposal is structured as a constructive parallel to the People's Republic of China's 2026 *Implementation Guidelines for the Standardised Application and Innovative Development of Intelligent Agents* — six sections, fourteen sub-sections, thirty-eight numbered items — with a new §0 "Philosophical Foundations" chapter prepended. §0 draws on three lineages: the Tractatus AI Safety Framework's six runtime services (boundary enforcement, context pressure monitoring, cross-reference validation, instruction persistence classification, metacognitive verification, pluralistic deliberation orchestration); the CARE Principles for Indigenous Data Governance (Carroll et al. 2020) and the global Indigenous Data Sovereignty movement that produced them, including Te Tiriti-grounded scholarship from Te Mana Raraunga and Dr Karaitiana Taiuru; and the international AI-standards landscape coordinated through ISO/IEC JTC 1/SC 42 (22989 terminology, 23053 lifecycle, 23894 risk management, 42001 management systems). The proposal advocates committee formation under a suitable umbrella organisation — candidates including the Royal Society Te Apārangi, the Standards New Zealand SC42 mirror committee, and the New Zealand AI Forum — to develop NZ-context recommendations and to engage in international dialogue. This is the v1 May 2026 draft; comments are welcome via the standing paper-comments channels on agenticgovernance.digital.

Preamble

Intelligent agents — intelligent systems capable of autonomous perception, memory, decision-making, interaction, and execution — are accelerating their integration with the records, infrastructure, and social processes of Aotearoa New Zealand. This proposal offers a civil-society contribution to how that integration should be governed: a sovereign, federated architecture in which every operation of an intelligent agent against a record produces an attributed, cryptographically-signed entry against the record's holder, and in which coordination between sovereign installations occurs through bilateral federation. The proposal mirrors the structure of the People's Republic of China's 2026 *Implementation Guidelines* so the architectural choices on each side appear in constructive parallel, opening dialogue with the authors of that framework, with international peers, and with New Zealand policymakers, community organisers, and sector practitioners. My Digital Sovereignty Ltd offers this as a starting point — for adoption, adaptation, and revision — under permissive open-source licences. It is offered as civil-society contribution and makes no claim of Crown policy status. Where its principles are useful to the adopter's own work, they are free to use; where they are not, they remain on the page.

§0. Philosophical Foundations

We open with foundations because architecture follows from philosophy. The recommendations that follow in §I-§VI are not arbitrary technical choices; they are implications of philosophical commitments that this section names explicitly. Three lineages converge here: the Tractatus AI Safety Framework's structural account of how intelligent agents may safely operate against records held by sovereign entities, developed and published openly at agenticgovernance.digital; the global Indigenous Data Sovereignty movement, which articulates that data about people belongs to those people and the communities they belong to; and the international AI-standards work coordinated through ISO/IEC JTC 1/SC 42, which provides the formal vocabulary in which architectural recommendations become implementable in organisational practice. Naming all three at the outset is part of the constructive contribution this proposal makes to dialogue — with the authors of the Cyberspace Administration of China's 2026 *Implementation Guidelines*, with New Zealand policymakers and community organisers, and with international peers working on parallel questions.

(i) Tractatus framework primitives as named foundations

The Tractatus framework consists of six **system-level primitives** that together specify the architectural conditions under which intelligent agents may safely operate against records held by sovereign entities. **They are not model-level alignment techniques**; they are code-level runtime checks that wrap the agent, independently of how the underlying agent (current transformer-LLMs, future JEPA-style architectures, hybrid systems) is built or trained. They intercept and verify behaviour at the runtime boundary — the same architectural shape as filesystem capability scoping or OAuth scope checks. A working demo of the boundary-enforcement primitive is at /demos/boundary-demo.html. `[CITATION: Stroh, J. (2026). Tractatus AI Safety Framework — Core Values and Principles, and Core Concepts of the Tractatus Framework. Agentic Governance Digital. https://agenticgovernance.digital — both works CC BY 4.0.]`

Boundary enforcement establishes which decision types structurally require human approval. The foundational claim — adapted from Wittgenstein and named explicitly in the Tractatus framework — is that "what cannot be systematized must not be automated." Values decisions, cultural-context judgments, irreversible consequences, and unprecedented situations are not delegable to autonomous agents; the framework blocks such delegation by runtime intercept rather than by hope. The intercept fires before action execution, the same architectural shape as an OS kernel intercepting privileged syscalls — the process cannot bypass the check. The four categories above are treated as **fallible classifications negotiated by the community in practice**, not as fixed essences — classification errors are themselves recorded, appealable, and used to revise the router's behaviour over time (see *Architectural Alignment* §3.5). The router's output is trinary, not binary: *allow*, *deny*, and *escalate-to-human-deliberation*. The third state is load-bearing — it carries the architectural acknowledgement that a substantial fraction of significant decisions are not binary at the time of decision (see §3.3).

Context pressure monitoring recognises that an agent's context window is a finite resource and that pressure on capacity is a governance signal. Agents operating near capacity make more errors, and the framework intervenes before failure rather than after. Same shape as a circuit breaker: the breaker trips on measured load before the system damages itself; the framework throttles or routes-to-human-approval based on measured context utilisation before output quality degrades.

Cross-reference validation verifies an agent's proposed actions against the canonical in-

struction history, catching cases where training-time patterns override explicit user direction. The illustrative case is the "27027 incident": a user specifies a non-default database port, and the agent — despite the explicit instruction — defaults to the trained-on port number. Validation catches the override; without it, the override would silently corrupt operations. The validator is a runtime check on each proposed action, not a training-time alignment of the model itself.

Instruction persistence classification distinguishes transient instructions from durable governance state. Not all instructions are equally important; treating them as if they were degrades both safety (critical directives forgotten) and usability (trivial preferences over-enforced). Same shape as a configuration-vs-runtime-arg distinction in software: config values persist; CLI args are per-invocation; the classifier tags each instruction by class so downstream services treat it appropriately.

Metacognitive verification places a verification gate before action execution. The gate evaluates each proposed action against five dimensions — alignment, coherence, completeness, safety, and consideration of alternatives — and confidence thresholds determine whether actions proceed, proceed with caution, require review, or are blocked. The check is at the runtime boundary, not a behavioural ask of the model.

Pluralistic deliberation orchestration facilitates multi-stakeholder deliberation when boundary enforcement flags a values conflict. It does not adjudicate between moral frameworks; it structures the deliberation so that values held by different stakeholders are documented, accommodated where possible, and explicitly named when they cannot be reconciled. Foundational pluralism — the view that moral frameworks are irreducibly different and that no supervalue resolves them — is the philosophical commitment that makes pluralistic deliberation a structural primitive rather than a procedural nicety. It runs as a coordination service that documents and surfaces stakeholder positions; the orchestration does not ask the agent to mediate values internally.

These six services are the structural skeleton of this proposal. Every architectural recommendation that follows can be traced to one or more of them.

Substrate-vs-runtime distinction. The six §0(i) primitives above are the framework's *runtime* layer — code in the agent's host process that checks proposed actions against configured decision-class rules at the moment of decision. The architecture has a second layer that does not depend on runtime cooperation: *substrate* mechanisms — cryptographic provenance, federation envelopes, and member-driven portability of records — that live in distributed possession independent of the agent. A deeper network reasoning around the runtime layer cannot reason around the substrate layer: the substrate's safety follows from mathematics (signatures, distributed replication, exit without permission) rather than from the agent's cooperation. The substrate layer is developed in Paper A, *Sovereign-Record Architecture* (*Paper A*), and the distinction is argued in *Architectural Alignment* §3.4. The two layers are complementary: the runtime catches what it can; the substrate ensures that what the runtime misses still leaves the community holding verifiable records, federation pathways, and exit options. Survival posture is layered, not single-mechanism (see *Architectural Alignment* §7.4 amplification on agent-containment-independence).

(ii) The CARE Principles for Indigenous Data Governance

The CARE Principles for Indigenous Data Governance, published in 2020 by an international team of Indigenous data scientists under the auspices of the Global Indigenous Data Alliance, articulate four commitments: **Collective benefit** (data ecosystems should advance Indigenous self-determination and collective benefit); **Authority to control** (Indigenous Peoples' rights and interests in their data must be recognised); **Responsibility** (those working with Indigenous data have a responsibility to share how that data is used to support Indigenous

Peoples' self-determination); and **Ethics** (Indigenous Peoples' rights and wellbeing should be the primary concern at all stages of the data lifecycle). [CITATION: Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. Data Science Journal, 19, 43. https://doi.org/10.5334/dsj-2020-043]

CARE is positioned as a complement to FAIR (Findable, Accessible, Interoperable, Reusable). FAIR optimises for data circulation and reuse; CARE optimises for the rights and wellbeing of those the data is about. The two are not antagonistic. They address different questions: FAIR asks how data should flow; CARE asks under whose authority data flows are governed. A well-designed sovereignty architecture answers both.

We adopt CARE as a foundational reference. Where the recommendations that follow specify that agents must operate against attributed, provenance-anchored records held by their sovereign holders, that specification is operationalising the Authority-to-control commitment. Where the recommendations specify federated coordination rather than central registration, that specification is consistent with Responsibility — those holding data are accountable to those the data concerns.

(iii) Te Tiriti, tikanga, and mātauranga in AI ethics — Aotearoa New Zealand scholarship

Aotearoa New Zealand's Indigenous Data Sovereignty scholarship is among the most developed internationally. The early articulation of Māori Data Sovereignty principles came from Te Mana Raraunga (the Māori Data Sovereignty Network), founded in 2015 with a charter adopted in 2016. Those principles have been substantively reassessed by Dr Karaitiana Taiuru's 20 September 2025 *Critical Analysis of Te Mana Raraunga Data Principles*, which identifies them as not adequately addressing AI, AI bias and algorithmic discrimination, model training and analytics, digital colonialism, or environmental impacts; observes that the 2016 scope was narrow whereas "today Māori data is everywhere"; and finds that despite extensive academic citation the principles are largely not implemented in practice. [CITATION: Taiuru, K. (20 September 2025). Critical Analysis of Te Mana Raraunga Data Principles. https://www.taiuru.co.nz/critical-analysis-mana-raraunga/]

The currently recognised operative body in Aotearoa New Zealand for Māori data governance is **Te Kāhui Raraunga** (established 2019 as a Charitable Trust). Their published frameworks — the **Māori Data Governance Model "Tuia te korowai o Hine-Raraunga"**, structured around eight pou; the **Māori AI Governance Framework** which extends it; and the supporting **Māori AI Governance Summary Report** and **Conceptual AI Use Cases Reference Resource** — provide the current articulation of Māori data and AI governance. Te Kāhui Raraunga describes the Māori AI Governance Framework as "Activated" with public-sector case studies referenced; broad operationalisation outside specific public-service deployments remains an open question that this proposal honours rather than papers over. Te Kāhui Raraunga's Māori AI Governance Framework states that "AI systems must not be implemented in Aotearoa without fully realising Māori authority over Māori data"; this proposal does not displace that requirement. [CITATION: Te Kāhui Raraunga Charitable Trust. Māori Data Governance Model: Tuia te korowai o Hine-Raraunga, https://www.kahuiraraunga.io/maoridatagovernance; Māori AI Governance Framework, https://www.kahuiraraunga.io/maoriaigovernance; full bibliographic detail of dated publications pending primary-source verification.]

Dr Karaitiana Taiuru's published scholarship on Māori ethical frameworks for AI, on tikanga (Māori law and custom) in AI ethics, on Te Tiriti-respectful AI, and on mātauranga (Māori knowledge) protection in AI training data — including the 20 September 2025 critical analysis cited above — has provided foundational language for thinking about agentic AI in te

ao Māori contexts. We adopt his preferred grounding terms where appropriate: **mana motuhake** and **rangatiratanga** rather than prescribed Western conceptual frameworks; responsive and adaptive frameworks grounded in tikanga that can evolve with technological and social change; frameworks tailored to specific organisations and industries developed in partnership with relevant Māori stakeholders. We cite his work as foundational scholarship; we do not represent him — or anyone — as endorsing this specific proposal. What counts as appropriate use of intelligent agents in te ao Māori contexts is for tangata whenua to determine, not for this proposal to specify.

Gap analysis — what this proposal does and does not yet do Honest assessment matters more than aspirational claims for a v1.1 draft addressed to reviewers including Dr Taiuru. The proposal's architectural primitives — sovereignty by attribution, cryptographic provenance, member-portability, bilateral federation — are **compatible with operationalising** Māori authority over Māori data under the Te Kāhui Raraunga framework and Taiuru's preferred grounding terms. The compatibility points include:

- **Tenant isolation as foundational** (Village deployment property, anchored on the Tractatus boundary-enforcement primitive) operationalises by architecture the requirement that AI systems not be implemented without realising Māori authority over Māori data. A tenant operated by a hapū, iwi, or kaitiaki body holds its records under that body's authority by the framework's design rather than by promise.
- **Boundary enforcement** can be configured to require explicit kaitiaki / hapū / iwi authorisation for values-sensitive operations on the tenant's records; the framework enforces by architecture rather than by trust.
- **Cryptographic provenance and member-portable identifiers** support kaitiakitanga across generations: records carry their own audit trail, cannot be silently mutated, and members can migrate to a different sovereign installation under the same architecture.
- **The pluralistic-deliberation primitive** is designed for multi-framework moral deliberation when boundary enforcement flags a values conflict; it is capable in principle of holding kaupapa Māori frameworks alongside other frameworks in structured deliberation.

What this proposal **does not yet do**, and what reviewers should weigh accordingly, is equally important to name:

- **Mana motuhake and rangatiratanga as foundational philosophical grounding.** The Tractatus framework's foundational pluralism is itself a Western philosophical commitment, drawing on Berlin, Rawls, and Ostrom; it accommodates kaupapa Māori as one framework among many; it does not ground in mana motuhake and rangatiratanga as prior commitments. Closing this gap would require the framework's foundations to be re-articulated from a kaupapa Māori starting point — substantive work that cannot honestly be done by the present authors alone.
- **Indigenous-led design partnership.** The Tractatus framework and the Village implementation were developed by My Digital Sovereignty Ltd's principal (Pākehā) with subsequent Claude (Anthropic) authorship contributions. They were not co-designed with Māori stakeholders. Taiuru's recommendation for "frameworks tailored to specific organisations and industries, developed in partnership with relevant Māori stakeholders" is not satisfied at the design-process level. The architecture is *available* to be applied in partnership; the framework itself was not co-developed in partnership.
- **AI bias on cultural and racial dimensions at the training-data layer.** Cross-reference validation catches training-pattern overrides at the instruction-conflict layer; it does not address the deeper biases baked into training data that would not surface as instruction conflicts. The companion Paper B work on Situated Language Layers (per-tenant training discipline, no-weight-modification stance, jurisdiction-bound inference) engages more directly with these concerns than the Tractatus core does.

- **Digital colonialism as a named theoretical and political concept.** The proposal's tenant isolation and architectural sovereignty are partial structural responses to digital colonialism; the proposal does not engage the concept theoretically. The Distributive Equity whitepaper (cross-referenced from this site) engages this more explicitly than the Tractatus core does.
- **Environmental impacts of AI** are largely not engaged. The CPU-fallback inference architecture in the Village deployment is a partial operational response; it is not part of the framework's stated commitments.

The honest implication of this gap analysis is that the proposal offers the architectural primitives that operationalising Māori authority over Māori data would require, while acknowledging that operationalisation in te ao Māori contexts is a substantial separate undertaking that requires kaupapa-Māori-led design work that this proposal has not done. The committee proposal in §II item 4 is one mechanism by which that further work might be advanced; it is offered for consideration rather than as a complete answer.

The Algorithm Charter for Aotearoa New Zealand, signed by Crown agencies in 2020, provides the existing baseline for transparency, partnership with Māori, fairness, accountability, and data protection in Crown algorithmic decision-making. This proposal does not displace the Algorithm Charter; the recommendations that follow are intended to be implementable within and alongside it, and alongside the Te Kāhui Raraunga frameworks. [CITATION: Algorithm Charter for Aotearoa New Zealand (2020). https://www.data.govt.nz/leadership/governance/data-ethics/algorithm-charter/ — current status and any subsequent updates pending verification.]

(iv) The global Indigenous Data Sovereignty lineage

Indigenous Data Sovereignty is an international movement, not a New Zealand idiosyncrasy. Naming the international lineage matters: it places the Te Tiriti-grounded work above within a global conversation rather than as parochial localism, and it opens common ground with the CAC framework's authors as fellow contributors to non-Western framings of how data and AI should be governed.

The **First Nations Information Governance Centre** (FNIGC) in Canada operates the **OCAP** principles — Ownership, Control, Access, Possession — originally articulated in the 1990s in the context of the First Nations Regional Longitudinal Health Survey and now embedded in research-ethics practice across Canadian universities, governments, and First Nations communities. [CITATION: First Nations Information Governance Centre. The First Nations Principles of OCAP®. https://fnigc.ca/ocap-training/]

The **United States Indigenous Data Sovereignty Network** (USIDSN), established in 2016 in connection with the Native Nations Institute at the University of Arizona, has advanced Indigenous Data Sovereignty practice in the United States context, including through engagement with US federal data-policy processes. [CITATION: United States Indigenous Data Sovereignty Network. https://usindigenousdata.org/]

The **Maiam nayri Wingara Indigenous Data Sovereignty Collective** in Australia — the name carries the meaning "Many Voices One Mind" — was established in 2017 and published an Indigenous Data Sovereignty Communiqué in 2018 that has shaped Australian Indigenous-data practice. [CITATION: Maiam nayri Wingara Indigenous Data Sovereignty Collective. (2018). Indigenous Data Sovereignty Communiqué.]

The **Global Indigenous Data Alliance** (GIDA) coordinates across these and other national-level Indigenous Data Sovereignty networks internationally; it is the auspice under which the CARE Principles were published. [CITATION: Global Indigenous Data Alliance. https://www.gida-global.org/]

That so much of the philosophical heavy lifting in this proposal traces to Indigenous scholarship is not coincidental. The recurring questions — under whose authority do data and the agents operating on it act? to whom are accountability and provenance owed? what is the proper scale at which collective interests are weighed against individual ones? — are questions Indigenous Data Sovereignty has been working on for decades. Resistance to extractive big-tech architectures, and the articulation of architectural alternatives grounded in collective authority, is one of the international movement's most generative contributions. The recommendations that follow draw on this lineage and are addressed to it in dialogue.

(v) ISO/IEC JTC 1/SC 42: the international AI-standards landscape

International AI standards work, coordinated through ISO/IEC JTC 1/SC 42, provides the formal vocabulary and management-system frameworks in which recommendations of this kind become implementable in organisational practice. Four standards are particularly relevant.

ISO/IEC 22989:2022 specifies the concepts and terminology of artificial intelligence. We use ISO/IEC 22989 terminology where compatible — for example, the term "AI system" carries its 22989 definition. Terminology consistency makes this proposal readable to standards-rigorous reviewers and implementable alongside other 22989-aligned work. [CITATION: ISO/IEC 22989:2022. Information technology — Artificial intelligence — Artificial intelligence concepts and terminology. International Organization for Standardization / International Electrotechnical Commission.]

ISO/IEC 23053:2022 establishes a framework for AI systems using machine learning, mapping the components of a machine-learning-based AI system and the relationships between them. Recommendations in this proposal that concern lifecycle, provenance, or component-level attestation can be implemented alongside 23053 lifecycle stages. [CITATION: ISO/IEC 23053:2022. Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML). ISO/IEC.]

ISO/IEC 23894:2023 provides guidance on AI risk management. It is the standards-body counterpart to the framework's per-installation risk-monitoring and incident-handling recommendations. [CITATION: ISO/IEC 23894:2023. Information technology — Artificial intelligence — Guidance on risk management. ISO/IEC.]

ISO/IEC 42001:2023 specifies requirements for an AI management system. It is the AI counterpart to ISO/IEC 27001 (information security management) and ISO 9001 (quality management). We position the recommendations in this proposal as implementable within an ISO/IEC 42001-style management system; organisations adopting any part of this proposal are likely to be those already operating, or planning to operate, ISO/IEC 42001-aligned governance. [CITATION: ISO/IEC 42001:2023. Information technology — Artificial intelligence — Management system. ISO/IEC.]

The committee work that produces these standards involves national-mirror committees in numerous jurisdictions, including the United Kingdom (via the British Standards Institution) and other national-standards bodies internationally. Aotearoa New Zealand's participation in SC42 work — through Standards New Zealand or a mirror committee constituted for that purpose — is one of the venues in which the constructive contribution this proposal advocates would naturally take place. [NOTE: existence of a current NZ SC42 mirror committee to verify before paragraph drafts of items 4, 12, 14, 35, 38.]

Closing

These five lineages — Tractatus, CARE, Te Tiriti-grounded Indigenous Data Sovereignty scholarship, the global Indigenous Data Sovereignty movement, and ISO/IEC SC42 — converge in

the architectural choices the rest of this proposal specifies. Sovereignty as attribution; bilateral federation as coordination; polycentric governance as authority structure; cryptographic provenance as audit infrastructure: none of these are invented for this proposal. Each has roots in one or more of the lineages named above. What this proposal contributes is a particular arrangement of these primitives, adapted to the Aotearoa New Zealand context, offered as a constructive parallel to the framework with which it shares its structure.

---

§I. Basic Principles

We propose four basic principles for the sovereign and federated development of intelligent agents in Aotearoa New Zealand. Each parallels one of the four principles that opens the Cyberspace Administration of China's 2026 *Implementation Guidelines*; in each case we affirm the principle's underlying intent and offer a constructive parallel grounded in the §0 foundations.

Sovereignty and attribution. Every operation of an intelligent agent against a record is attributable to a sovereign holder of that record; provenance is cryptographic; safety arises from the record-holder's authority over their own records. We affirm the CAC framework's commitment to safety and controllability as foundational. We propose, as a constructive parallel, that for the Aotearoa New Zealand context — where Te Tiriti partnership, the existing Privacy Act 2020 framework, and the CARE Authority-to-control commitment converge — attribution-based sovereignty is well-suited to operationalising those same safety concerns. The Tractatus boundary-enforcement primitive provides the architectural mechanism; cryptographically signed records provide the audit trail; and the legitimate authority over both is the holder of the records, by jurisdiction and by partnership obligations. *(Parallels CAC §I principle 1 "safety and controllability".)* [CITATIONS: Tractatus boundary enforcement (Stroh 2026, CC BY 4.0); CARE Principles, Authority to control commitment (Carroll et al. 2020); Privacy Act 2020 (NZ), information privacy principles.]

Bilateral and federated. Coordination between sovereign installations occurs through bilateral federation and open international standards. We acknowledge the merit of the CAC framework's commitment to standardised and orderly development; standardisation and order are necessary conditions for any large-scale agentic-AI deployment, and the CAC framework's coordinated standardisation programme is one credible approach. We propose, for the Aotearoa New Zealand context — smaller scale, well-established Māori Data Sovereignty principles, existing bilateral institutional arrangements across Crown agencies, hapū, iwi, civil-society organisations, and the private sector — that a federated approach to coordination is well-suited. Federation between sovereign installations is well-supported by existing W3C, IETF, and ISO/IEC SC42-aligned standards. We offer bilateral federation for consideration as a parallel architecture that may interoperate with central-registration approaches in other jurisdictions, and we invite committee-formation under suitable umbrella organisations to develop the interoperability dialogue. *(Parallels CAC §I principle 2 "standardised and orderly development".)* [CITATIONS: W3C Decentralized Identifiers (DIDs) v1.0 (W3C Recommendation, 2022) and W3C Verifiable Credentials Data Model v1.1; ActivityPub (W3C Recommendation, 2018); ISO/IEC 42001:2023 management systems; Te Kāhui Raraunga Māori AI Governance Framework + Taiuru critical analysis (see §0(iii)).]

Pluralistic-deliberation, polycentric. Multiple value frameworks coexist within and across sovereign installations; deliberation between them is procedural and structured; innovation arises from local adaptation under local authority. We affirm the CAC framework's commitment to innovation-driven development. We propose, as a constructive parallel, that polycentric governance — multiple loci of authority, multiple value frameworks held in productive tension, with structured deliberation when conflicts arise — is well-matched to the Aotearoa

New Zealand context of Te Tiriti partnership, and is well-supported by international scholarship on polycentric governance (notably Elinor Ostrom's foundational work). The Tractatus pluralistic-deliberation primitive provides the architectural mechanism for facilitating multi-stakeholder deliberation when boundary enforcement flags a values conflict; foundational pluralism is the philosophical commitment that makes this a structural feature of the framework. *(Parallels CAC §I principle 3 "innovation-driven development".)* [CITATIONS: Tractatus pluralistic deliberation primitive (Stroh 2026, CC BY 4.0); Ostrom, E. (2010). Beyond Markets and States: Polycentric Governance of Complex Economic Systems. American Economic Review, 100(3), 641–672. https://doi.org/10.1257/aer.100.3.641 — full bibliographic detail to be verified before v1 publication.]

Adoption-led, evidenced. Applications of intelligent agents are evidenced by deployment in communities that have adopted them; for a civil-society proposal, the appropriate evidential basis is real-world deployment. We affirm the CAC framework's commitment to application-led development. We propose, as a constructive parallel, that for a civil-society proposal originating from a single company, deployment evidence must precede recommendation. Where this proposal cites Aotearoa New Zealand deployment examples (in §IV and §V) — in parish and hapū / iwi contexts, in iwi and diaspora family-history contexts, in small-business contexts — those citations are to actual deployments, with concrete deployment data (counts, start dates, scope) to be added before v1 publication. Where the proposal advances recommendations into sectors in which MDSL has not yet deployed, those recommendations are framed as sovereignty-architecture conditions for any agent deployment in that sector, addressed to whoever may wish to apply the architecture there. What deployment evidences is not just adoption-count but architectural availability — that the substrate primitives (cryptographic provenance, federation envelopes, member-driven portability, boundary enforcement) operate as specified in real conditions against the data of communities that have authorised the trial, rather than in artificial benchmarks. *(Parallels CAC §I principle 4 "application-led approach".)* [CITATIONS: MDSL deployment evidence — Village (parish + community contexts), family-history (iwi + diaspora contexts), sydigital (small-business contexts); specific deployment data (counts, start dates, tenancy scope) pending operator-verified figures before v1 publication.]

---

§II. Foundations for Sovereign Development

Where the Cyberspace Administration of China's framework consolidates technological foundations under a state-coordinated standardisation programme, we offer foundations rooted in cryptographic sovereignty and bilateral protocols. The two sub-sections that follow — strengthening the sovereignty foundation, and establishing bilateral protocols — together specify the architectural primitives on which the rest of the proposal builds.

(I) Strengthening the Sovereignty Foundation

Item 1. Build sovereign primitives for agents. Cryptographically signed records, member-portable identifiers, and attributed provenance form the foundation for agent operations on sovereign data. They are architectural primitives that constitute sovereignty at the level of the record itself. We propose sustained investment in open-source cryptographic primitives — digital signing, verifiable credentials, content-addressed storage with provenance — and in portable-identity standards usable across any sovereign installation in any sector. We propose that these primitives be developed and maintained as common infrastructure, available under permissive open-source licences (Apache 2.0, EUPL-1.2, or compatible) allowing adoption, modification, and redistribution by any party. The Tractatus boundary-enforcement primitive, the cross-reference-validation primitive, and the instruction-persistence-classification primitive together specify the runtime mechanics;

cryptographic signing and verifiable-credential infrastructure provide the underlying audit trail. *(Parallels CAC item 1 "strengthen R&D in foundational technologies".)* [CITATIONS: Tractatus framework (Stroh 2026, CC BY 4.0 text / Apache 2.0 code); W3C Decentralized Identifiers (DIDs) v1.0 (W3C Recommendation, 2022); W3C Verifiable Credentials Data Model v1.1; CARE Principles, Authority-to-control commitment (Carroll et al. 2020).]

Item 2. Refine the sovereign toolchain. Open-source reference implementations of agent frameworks — including the Tractatus framework's six services — should be available for adoption by any sovereign installation under permissive open-source licences that permit installation-local operation. We propose that the toolchain for developing, testing, deploying, and maintaining sovereignty-architected agentic systems be developed in the open, with contributions encouraged from any sovereign installation. The current MDSL implementations — the Tractatus framework distributed under Apache 2.0 (with documentation under CC BY 4.0); the Village and community codebases migrating toward EUPL-1.2 in phases as of mid-2026 — are offered as one set of reference implementations among potentially several. Security tooling — adversarial-input detection, behavioural-anomaly detection, attestation tooling for builds and dependencies — is the appropriate technical complement to the Tractatus boundary-enforcement and metacognitive-verification primitives. *(Parallels CAC item 2 "refine the agent toolchain".)* [CITATIONS: Tractatus framework reference implementation (Stroh 2026), Apache 2.0 (code), CC BY 4.0 (text and figures); EUPL-1.2 (European Union Public Licence); Apache 2.0 (Apache Software Foundation).]

(II) Establishing Bilateral Protocols

Item 3. Federated bilateral protocols. Interoperability between sovereign installations occurs through bilateral agreements and open international standards. We acknowledge the merit of the CAC framework's commitment to a standardised interconnection programme — the proposed Intelligent Agent Interconnection Protocol (AIP), foundational interface standards across software, services, and hardware peripherals, and mandatory standards in sensitive sectors. We propose, for the Aotearoa New Zealand context, that interoperability between sovereign installations is well-supported by the existing international standards landscape: W3C Decentralized Identifiers and Verifiable Credentials for identity; ActivityPub and related W3C federation protocols for inter-installation communication; IETF protocols for authentication, transport, and content addressing; and ISO/IEC SC42 work for AI-specific terminology, lifecycle, risk, and management-system alignment. We propose that Aotearoa NZ contribute to international interoperability standards as a peer participant in those existing forums. *(Parallels CAC item 3 "standardisation system" and the proposed AIP interconnection protocol.)* [CITATIONS: W3C DIDs v1.0; W3C Verifiable Credentials Data Model v1.1; ActivityPub (W3C Recommendation, 2018); ISO/IEC 22989:2022 terminology; ISO/IEC 23053:2022 ML framework.]

Item 4. Cryptographic identity; federated dialogue on the Intelligent Internet. Identity is per-installation, anchored in DNS and cryptographic keys; verification between counterparties is peer-to-peer; capability declarations are published by each installation. We acknowledge the merit of the CAC framework's proposal for an intelligent-agent registration platform, which contemplates not only digital identity management and capability declaration but also search and discovery, trusted interconnection, compliant payment, security protection, conflict resolution, IPv6 leverage, and a monitoring indicator system — a substantial and coherent set of inter-related functions. A centralised registration platform with a coordinating authority is one credible architectural approach to these functions.

We propose, for the Aotearoa New Zealand context — where smaller scale, well-established Māori Data Sovereignty principles, and the architectural primitives already represented in MDSL deployments converge — a federated approach in which each Intelligent Internet func-

tion is addressed through bilateral arrangements and open international standards. Identity and capability declaration are served by W3C Decentralized Identifiers and Verifiable Credentials. Search and discovery between sovereign installations can draw on the patterns established by ActivityPub-derived federation, by WebFinger (IETF RFC 7033), and by federation-aware directory protocols such as nodeinfo — though we note federated discovery at scale remains an open engineering problem and acknowledge it as such. Trusted interconnection and security protection operate through bilateral cryptographic attestation. Compliant payment routes through existing financial-regulatory channels. Conflict resolution operates through bilateral mediation and existing dispute-resolution mechanisms, with cryptographic provenance providing the audit trail. IPv6 is an underlying infrastructure choice available to any installation. A monitoring indicator system is realisable through open publication of operational metrics by each participating installation, aggregated by independent observers.

We propose formation of a single committee under a suitable umbrella organisation — candidates include the Royal Society Te Apārangi, the Standards New Zealand SC42 mirror committee (existence to verify), the New Zealand AI Forum, or a joint structure across these — **to develop NZ-context recommendations on agentic-AI architecture in detail, to contribute to ISO/IEC JTC 1/SC 42 work as a peer participant, and to engage in bilateral dialogue with the CAC framework's authors and with international peers. The committee would carry five named workstreams: (i) federated identity for intelligent agents and the broader Intelligent Internet functions named in this item; (ii) federated audit and compliance services (cross-reference §III item 12); (iii) attestation-based reputation systems (cross-reference §III item 14); (iv) industry-coordination patterns including federation versus alliance models (cross-reference §V item 35); and (v) international engagement and bilateral cooperation on agentic AI (cross-reference §V item 38). The committee's contribution to international standards work and to dialogue with the CAC framework's authors is its principal product.** We offer this committee proposal as one contribution to the international conversation; the conversation will benefit from contributions across many architectural traditions. *(Parallels CAC item 4 "intelligent internet architecture" with registration platform; committee-formation pattern is consolidated across items 4, 12, 14, 35, and 38.)* [CITATIONS: W3C Decentralized Identifiers (DIDs) v1.0; W3C Verifiable Credentials Data Model v1.1; ActivityPub (W3C Recommendation 2018); WebFinger (IETF RFC 7033); nodeinfo federation directory; ISO/IEC 22989:2022; Te Kāhui Raraunga (kahuiraraunga.io — Māori Data Governance Model and Māori AI Governance Framework); Taiuru, K. (20 Sep 2025) Critical Analysis of Te Mana Raraunga Data Principles, taiuru.co.nz/critical-analysis-mana-raraunga/; Royal Society Te Apārangi; ISO/IEC JTC 1/SC 42.]

---

§III. Upholding the Sovereignty Baseline

Where the Cyberspace Administration of China's framework establishes a security baseline through product guidelines, behavioural-fencing technologies, tiered governance, and industry self-regulation with credit-rating sanctions, we offer a baseline rooted in the adopter's own jurisdictional framework, cryptographic provenance, polycentric governance arrangements, and federation-based coordination. The four sub-sections that follow — product principles, security risks, governance system, federated coordination — together specify how an intelligent agent's compliance with sovereignty principles can be verified at runtime and audited post-hoc.

(I) Clarifying Product Principles

Item 5. Anchor in the adopter's own laws. Policies, regulations, and ethical standards governing intelligent agents arise from the adopter's jurisdiction. Values are sourced from

local law and local institutional arrangements; the architecture provides the implementation infrastructure in which those values are operative. In Aotearoa New Zealand, the applicable instruments include the Privacy Act 2020 (with the Health Information Privacy Code 2020 and other codes as applicable to specific sectors); the New Zealand Bill of Rights Act 1990 where state actors are involved; the Algorithm Charter for Aotearoa New Zealand for Crown agencies; the Te Tiriti o Waitangi obligations on Crown actors and the partnership obligations they entail; the Official Information Act 1982; the Public Service Act 2020; the Public Records Act 2005; and sectoral statutes including the Reserve Bank of New Zealand Act 2021, the Education and Training Act 2020, the Local Government Act 2002, and the Search and Surveillance Act 2012, applicable to the relevant deployment context. The architecture is implementation-neutral with respect to which jurisdiction's law applies; the proposal is addressed to Aotearoa NZ adopters, and the same primitives serve adopters in any jurisdiction whose values they wish to operationalise. *(Parallels CAC item 5 "policies, regulations and ethical standards".)* [CITATIONS: Privacy Act 2020 (NZ); New Zealand Bill of Rights Act 1990; Algorithm Charter for Aotearoa New Zealand (2020); Health Information Privacy Code 2020; Official Information Act 1982; Public Service Act 2020; Public Records Act 2005; Reserve Bank of New Zealand Act 2021; Education and Training Act 2020; Local Government Act 2002; Search and Surveillance Act 2012 — current legislative versions to verify before v1 publication.]

Item 6. User-final decision authority, cryptographically backed. We affirm the same principle the CAC framework affirms: the user retains the right to be informed of, and final decision authority over, autonomous actions taken by intelligent agents on their behalf. The principle is foundational to the relationship of trust between a person and the agentic systems acting in their name. We propose, as the audit mechanism, per-record cryptographic provenance against the user's own sovereign record: every autonomous action by an agent operating against the user's records produces a cryptographic entry attesting to the action, attributable to the agent and to the user's authorisation framework. The user can inspect, replay, and challenge any agent action against this provenance, and the Tractatus instruction-persistence-classification primitive provides the framework for distinguishing routine actions from those requiring explicit user reconfirmation. *(Parallels CAC item 6 "clarify decision-making authority".)* [CITATIONS: Tractatus instruction persistence classification primitive (Stroh 2026, CC BY 4.0); Privacy Act 2020 (NZ), information privacy principle 6 (access rights); CARE Principles, Authority-to-control commitment (Carroll et al. 2020).]

Item 7. Provenance, complementing behavioural control. We acknowledge the CAC framework's emphasis on rule embedding, behavioural fencing, and blockchain-anchored verification of agent behaviour in critical application scenarios. These are credible architectural approaches for ensuring lawful and compliant behaviour in centrally-coordinated deployments. We propose, as an additional architectural primitive well-suited to bilateral federation, **provenance**: every action by an intelligent agent produces a cryptographic record attributable to the actor. The two approaches complement each other. Behavioural fencing constrains what an agent may attempt at runtime; provenance creates an unforgeable record of what was actually attempted. Both have a role, and the appropriate balance between them is likely context-specific. *(Parallels CAC item 7 "strengthen behavioural control".)* [CITATIONS: Tractatus cross-reference validation primitive (Stroh 2026, CC BY 4.0); W3C Verifiable Credentials Data Model v1.1; ISO/IEC 23894:2023 risk management.]

(II) Mitigating Security Risks

Item 8. Intrinsic security through sovereign primitives. Personal information remains in the holder's installation; cryptographic protection is per-record as well as perimeter-based; attack detection runs locally against the holder's records; access is contract-bound between counterparties. The blast radius of a failure is bounded to the affected installation. We

affirm the CAC framework's commitment to intrinsic security capabilities — data security, personal information protection, cryptographic protection, attack detection, access control, behavioural control. We propose, as a constructive parallel, that for a federated architecture the appropriate locus of these capabilities is the sovereign installation, with bilateral mechanisms for cooperation between installations where threats cross jurisdictional or organisational boundaries. *(Parallels CAC item 8 "intrinsic security capabilities".)* [CITATIONS: Tractatus boundary enforcement primitive (Stroh 2026, CC BY 4.0); Privacy Act 2020 (NZ); ISO/IEC 23894:2023 risk management.]

Item 9. Supply-chain attestation, federated sharing. We propose per-installation full-lifecycle attestation — signed build provenance, dependency manifests, training-data attestation where applicable, security incident response history — published openly by each installation. Supply-chain incidents are shared bilaterally between federated peers and via established international channels including CERT-NZ, CERT-EU, US-CERT, and the CVE coordination system. We acknowledge the merit of the CAC framework's commitment to full-lifecycle security standards and supply-chain information sharing. We propose that for federated coordination, supply-chain transparency is achieved through open publication of attestations by each installation, with bilateral cooperation on incident response. *(Parallels CAC item 9 "supply chain security".)* [CITATIONS: ISO/IEC 23894:2023 risk management; CERT-NZ disclosure procedures; international CVE coordination process; ISO/IEC 42001:2023 management systems.]

Item 10. Bound the blast radius; audit post-hoc. Routine risk identification operates locally to each installation, with cross-installation incidents propagating through federation. The framework's principal contribution to mitigating automated-attack risk, privacy infringement, and false-information dissemination is bounding the scale at which automated harm compounds. We affirm the CAC framework's commitment to risk identification, early warning, intervention, and prevention of agentic AI from being used in illegal activities (automated attacks, privacy infringement, false-information generation and dissemination, online fraud). We propose, as a complementary architectural contribution, that bounding the scale of automated harm — through per-installation operational boundaries and bilateral incident-response cooperation — is a structural complement to detection-and-intervention approaches at the centralised level. The structural mechanism is the **federation envelope**: by default, an installation's records remain within that installation, and cross-installation flow occurs only through envelopes the installation explicitly signs, carrying provenance and recipient binding. Compromise of one installation cannot silently propagate into others, because the substrate has no implicit cross-installation read path — there is no shared registry an attacker can pivot through. Bilateral incident-response then operates on what was deliberately shared, with the federation envelope's provenance enabling forensic reconstruction of the affected scope without requiring a centralised audit aggregator. *(Parallels CAC item 10 "mitigate risks arising from applications".)* [CITATIONS: Tractatus pluralistic deliberation primitive (Stroh 2026, CC BY 4.0); ISO/IEC 23894:2023 risk management; Privacy Act 2020 (NZ); Harmful Digital Communications Act 2015 (NZ) — current legislative versions to verify before v1 publication.]

(III) Improving the Governance System

Item 11. Polycentric governance, in dialogue with tiered approaches. Governance authority over what an intelligent agent may do with a record belongs to the holder of the records. Scenario-permissibility is determined per-installation by the holder's own jurisdiction, supported by sectoral regulators where their authority extends to the relevant subject matter. We acknowledge the merit of the CAC framework's categorised and tiered governance approach for sensitive sectors and key industries, with the Cyberspace Administration of China and relevant industry authorities determining permissible application scenarios and implementing management measures such as filing, testing, and the recall of problematic

products. We propose, for the Aotearoa New Zealand context, that polycentric governance — multiple loci of authority across Crown agencies, hapū / iwi entities, sectoral regulators, professional bodies, and the holders of records themselves — is well-suited to the existing institutional landscape and to Te Tiriti partnership obligations. International scholarship on polycentric governance, notably Elinor Ostrom's foundational work, provides the theoretical grounding for this approach. Polycentricity is operationally supported by the substrate's properties: a single **cryptographically-signed audit chain** per record allows multiple authorities — Crown regulator, hapū / iwi entity, sectoral body, professional college, the record-holder themselves — to each verify the same record under their respective competence, without requiring centralised collation or duplicated registries. Different authorities hold their own copies of the audit chain under their own keys and reach independent compliance judgements on the same underlying record. The architectural primitive that makes polycentric governance operationally tractable is the federation-replicated, signed audit chain; the institutional question of who has authority over which decision class remains political. *(Parallels CAC item 11 "categorised and tiered governance".)* [CITATIONS: Ostrom, E. (2010). Beyond Markets and States: Polycentric Governance of Complex Economic Systems. American Economic Review, 100(3), 641–672. https://doi.org/10.1257/aer.100.3.641; Algorithm Charter for Aotearoa New Zealand (2020); Te Kāhui Raraunga (kahuiraraunga.io — Māori Data Governance Model and Māori AI Governance Framework); Taiuru, K. (20 Sep 2025) Critical Analysis of Te Mana Raraunga Data Principles, taiuru.co.nz/critical-analysis-mana-raraunga/.]

Item 12. Compliance services federated. Risk monitoring, testing, evaluation, audit, and certification services for intelligent agents exist as commercial, community, and academic offerings; mutual recognition between services occurs through open publication and peer review. We acknowledge the merit of the CAC framework's commitment to a compliance service system providing professional services such as risk monitoring, testing and evaluation, consultancy, and certification, with promotion of mutual recognition between accredited providers. **This area is workstream (ii) of the single committee proposed in §II item 4. The committee would develop NZ-context recommendations on a federated audit framework for intelligent agents, contribute to ISO/IEC SC42 work on AI assessment, evaluation, and management systems, and engage in bilateral dialogue with the CAC framework's authors on the interaction between federated and centralised compliance services.** Compliance services federate concretely as follows: each installation publishes its own attestations — build provenance, dependency manifests, training-data attestation where applicable, incident-response history, audit-framework adherence — under its own cryptographic identity. Compliance providers verify against these attestations and publish their findings under their own identities; mutual recognition between providers occurs through cross-citation of cryptographically verifiable assessments rather than through central accreditation. The substrate primitive making this operational is content-addressed publication with cryptographic provenance — anyone can verify any compliance assessment against the specific version of the installation it actually assessed. *(Parallels CAC item 12 "compliance service system"; consolidated committee-formation workstream applies.)* [CITATIONS: ISO/IEC 42001:2023 management systems; ISO/IEC 23894:2023 risk management; Royal Society Te Apārangi.]

(IV) Strengthening Federated Coordination

Item 13. Coordination by federation. Sovereign installations federate bilaterally; coordination on shared concerns — interoperability standards, security incident disclosure, audit framework development — occurs through open publication and consensus among contributing peers. We acknowledge the merit of the CAC framework's commitment to industry self-regulation, with industry organisations and major enterprises jointly formulating self-regulatory rules covering AI functionality compliance, algorithm governance, intellectual

property protection, and fair competition. We propose, for the federated architecture this proposal specifies, that coordination on shared concerns occurs through open publication and consensus among contributing peers; the architectural commitment to bilateral federation extends to the coordination mechanism itself. Federation in this proposal is bilateral by construction: each installation publishes a federation endpoint and chooses which peers it will federate with, on what specific record classes; the **federation envelope** format specifies what records may travel, with what consent scope, to which recipient, with what non-onward-forwarding constraints. The substrate primitives making bilateral federation operational are the federation envelope (recipient-bound, provenance-attached, scope-limited message format), **member-driven portability** (the holder of records can demand egress from any installation to any destination of their choosing), and **cryptographic provenance** (every record carries verifiable origin metadata that survives transit). Coordination on shared concerns — interoperability standards, security incident disclosure, audit-framework development — proceeds bilaterally between federated peers without requiring a centralised registry. *(Parallels CAC item 13 "industry self-regulation".)* [CITATIONS: ActivityPub federation protocol (W3C Recommendation 2018); IETF Request for Comments process; W3C process document.]

Item 14. Reputation by attestation. Sovereign installations publish their own attestations — security posture, audit history, dependency manifests, incident response — and counterparties verify cryptographically. Reputation accrues through history of accurate self-disclosure verified by bilateral counterparties. We acknowledge the merit of the CAC framework's proposal for voluntary credit rating mechanisms for market entities in the intelligent agent sector, with credit assessments for behaviours such as misuse of technology, inducing consumption, false advertising, and concealing information on defects, and sanctions for dishonest conduct in accordance with laws and regulations. **This area is workstream (iii) of the single committee proposed in §II item 4. The committee would develop NZ-context recommendations on attestation-based reputation versus registry-based reputation, contribute to international standards work on AI provenance and attestation, and engage in bilateral dialogue with the CAC framework's authors on interoperability between attestation-based and credit-rating-based reputation systems.** *(Parallels CAC item 14 "credit rating mechanisms"; consolidated committee-formation workstream applies.)* [CITATIONS: W3C Verifiable Credentials Data Model v1.1; ISO/IEC 42001:2023 management systems.]

---

§IV. Strengthening Adoption-Driven Development

Where the Cyberspace Administration of China's framework lists nineteen sectors in which the state directs that "agents shall do X", we mirror the nineteen sectors and reframe each as a question of sovereignty conditions for any agent deployment in that sector. The framework does not direct deployment; it specifies the architectural conditions under which deployment is sovereignty-compatible. The reframe is rhetorically modest but structurally consequential: the state-directed reading positions intelligent agents as instruments of sectoral programmes, while the sovereignty-conditions reading positions them as tools whose use must satisfy attribution, provenance, and member-portability requirements regardless of who deploys them.

The architectural primitives invoked across the nineteen sectors that follow are four. **Cryptographic provenance** attaches verifiable origin metadata to every record — who wrote it, when, against what authority — immutable to silent post-hoc edit (corrections are countersigned and themselves recorded). **Federation envelopes** mediate cross-installation sharing: only the consented subset travels, carrying provenance, recipient binding, and non-onward-forwarding by default. **Member-driven portability** lets the holder of records export their bundle to another installation without the original holder's permission, with provenance intact at the destination. **Boundary enforcement** routes decisions in the four boundary cat-

egories (irreversibility, values-laden, cultural-context-dependent, unprecedented) to human deliberation by default, with the routing itself recorded. The sector items that follow name the sector-specific manifestation of one or more of these primitives; the generic capabilities are constant across sectors. See *Architectural Alignment* §3 for the development of the primitives; *Paper A* for the substrate layer in full.

(I) Scientific Research

Item 15. In research, sovereignty primitives apply. Research environments operate on sovereign datasets — held by participating individuals, institutions, hapū / iwi entities, or research consortia under their respective governance arrangements; provenance accompanies derived results; bilateral federation between institutions provides the interoperability layer where data-sharing is necessary. We acknowledge the merit of the CAC framework's vision of intelligent agents enhancing theoretical deduction, knowledge integration, and integration with scientific instruments and experimental platforms. We propose that for Aotearoa NZ research, those capabilities are deployed under research-ethics governance specific to each institution and to each research project, with the Tractatus pluralistic-deliberation primitive providing the architectural mechanism for scaling research-ethics review across competing value frameworks. The operative primitives are cryptographic provenance (every dataset and derived result carries provenance attesting to its sources, derivations, and the ethical-review framework under which it was produced) and federation envelopes (cross-institutional sharing occurs under explicit data-sharing agreements, with the envelope recording what subset of data travels and under what consent scope). Member-driven portability lets a research participant withdraw their contribution and have provenance updated downstream; boundary enforcement routes the values-laden ethics decisions to the research-ethics committee rather than to autonomous agent action. *(Parallels CAC item 15 "research and exploration".)* [CITATIONS: CARE Principles (Carroll et al. 2020); FAIR Principles (Wilkinson et al. 2016, https://doi.org/10.1038/sdata.2016.18); Te Kāhui Raraunga (kahuiraraunga.io — Māori Data Governance Model and Māori AI Governance Framework); Taiuru, K. (20 Sep 2025) Critical Analysis of Te Mana Raraunga Data Principles, taiuru.co.nz/critical-analysis-mana-raraunga/; New Zealand research-ethics framework via the Health Research Council and Royal Society Te Apārangi; Tractatus pluralistic deliberation primitive (Stroh 2026).]

Item 16. In software R&D, attribution and audit apply. Code-generation agents operate against attributed sources; derived works carry their lineage; CI/CD pipelines verify build attestation and dependency provenance. We acknowledge the merit of the CAC framework's commitment to software-development intelligent agents enhancing requirements analysis, architectural design, code generation, and testing. We propose that all such capabilities operate under attribution and provenance requirements; agentic contributions to code, design, or simulation outputs are attributed both to the agent and to the human or organisational operator on whose authority they were produced. The operative primitive is cryptographic provenance applied to every code artefact — the agent that proposed it, the human reviewer who approved it, the build pipeline that compiled it, the dependency tree's own attestations — forming a verifiable chain from authored line back to authorised commit. The Tractatus cross-reference validation primitive provides runtime verification that proposed code actions are consistent with the canonical instruction history. *(Parallels CAC item 16 "R&D support".)* [CITATIONS: W3C Verifiable Credentials Data Model v1.1; SBOM (Software Bill of Materials) standards via NTIA and OWASP CycloneDX; Tractatus cross-reference validation primitive (Stroh 2026).]

(II) Industrial Development

Item 17. In manufacturing, sovereignty primitives apply. Production data is the manufacturer's sovereign record; agents operating against it are attributed; cross-installation coordination for supply chains is bilateral. We acknowledge the merit of the CAC framework's commitment to production-management agents for scheduling, resource allocation, and process optimisation, and to integration with CNC machine tools, industrial robots, and automated production lines. We propose that all such capabilities operate under the manufacturer's authority, with supply-chain coordination occurring through bilateral agreements between participating manufacturers and counterparties. The operative primitives are cryptographic provenance (every batch carries production-line attestation — sensor readings, agent decisions, human approvals — verifiable backwards from any defect investigation) and federation envelopes (supply-chain coordination occurs through bilaterally-signed envelopes specifying what production data is shared with which counterparty for what purpose). Member-driven portability translates here to the manufacturer's ability to export their full production audit trail to another supply-chain audit body or regulator without the original platform vendor's permission. *(Parallels CAC item 17 "intelligent manufacturing".)* `[CITATIONS: ISO/IEC 42001:2023 management systems; pending lookup for NZ manufacturing data standards and Industry 4.0 NZ initiatives.]`

Item 18. In energy and resources, sovereignty primitives apply. Environmental data, resource catalogues, and dispatch logs are sovereign records of the responsible entities: Crown for some (statutory resources, certain environmental data); hapū and iwi for those where Treaty Settlement allocations apply; private entities for the remainder. Agents operate against the relevant entity's records under that entity's authority. The specific allocations are entity-specific and depend on the relevant Settlement legislation and arrangements. We acknowledge the merit of the CAC framework's commitment to environmental-sensing agents for natural-disaster and pollution-risk early warning, to power-dispatch and grid-maintenance agents, and to resource-exploration applications. We propose that for the Aotearoa NZ context, the relevant authorities arise from the existing institutional and Treaty framework, and the architecture provides the audit and attribution infrastructure within which those authorities operate. Cryptographic provenance attaches to environmental readings, grid-dispatch decisions, and resource-allocation choices, with attribution to the responsible entity (Crown, iwi, or private). Federation envelopes carry only the consented subset of environmental data across entity boundaries — early-warning signals propagate to all relevant entities without requiring central aggregation. Where Treaty Settlement allocations apply, the iwi entity holds its own audit chain under its own keys, independent of Crown agency systems. *(Parallels CAC item 18 "energy and resources".)* `[CITATIONS: Resource Management Act 1991 (NZ); relevant Treaty Settlements legislation (entity-specific, pending verification before v1 publication); Electricity Industry Act 2010 (NZ); Crown Minerals Act 1991 (NZ).]`

Item 19. In transport, sovereignty primitives apply. Vehicle telemetry, traffic data, and infrastructure sensor data are sovereign records of operators, Crown agencies, and road-controlling authorities; coordination between them — Waka Kotahi New Zealand Transport Agency, KiwiRail, maritime authorities, the Civil Aviation Authority, regional councils, and city councils — is bilateral federation across the relevant institutional boundaries. We acknowledge the merit of the CAC framework's commitment to traffic-safety, emergency-dispatch, and vehicle-control intelligent agents. We propose that the Aotearoa NZ context, with its existing bilateral institutional arrangements across transport modes, is well-suited to a federated approach. Federation envelopes specify what telemetry, traffic data, and infrastructure-sensor data is shared across institutional boundaries (Waka Kotahi ↔ regional councils ↔ KiwiRail ↔ Civil Aviation Authority), with cryptographic provenance ensuring forensic reconstruction of any incident is possible at the level of individual agent decisions. Member-driven portability applies at the vehicle and operator level — the operator can demand egress from any platform without lock-in. *(Parallels CAC item 19 "transport".)* `[CITATIONS: Land Transport Act`

1998 (NZ); Land Transport Management Act 2003 (NZ); Civil Aviation Act 1990 (NZ); Maritime Transport Act 1994 (NZ); pending lookup for NZ transport data sovereignty work.]

Item 20. In agriculture, sovereignty primitives apply. Farm data is the farmer's sovereign record; pest, disease, yield, and stocking data may be shared bilaterally with extension services, research institutions, or hapū rōpū where applicable, under the farmer's terms. We acknowledge the merit of the CAC framework's commitment to agricultural-services intelligent agents for technical guidance, pest and disease diagnosis, and integration with smart agricultural machinery and greenhouses. We propose that for Aotearoa NZ — where agricultural data sovereignty is a recognised issue across farm-data co-operatives, sectoral organisations, and increasing engagement with Māori data sovereignty in primary-industry contexts — bilateral data-sharing under the farmer's terms is well-suited. Cryptographic provenance attaches to farm data — sensor readings, agent recommendations, treatment decisions, yield outcomes. Federation envelopes carry only the consented subset (pest and disease data to extension services; aggregated yield data to research institutions; cultural-context-dependent data to hapū rōpū under appropriate tikanga) under the farmer's terms. Member-driven portability lets the farmer move between farm-data co-operatives without losing their historical audit trail. *(Parallels CAC item 20 "agricultural production".)* [CITATIONS: pending lookup for NZ agricultural data sovereignty work and farm-data governance arrangements; Te Kāhui Raraunga (kahuiraraunga.io — Māori Data Governance Model and Māori AI Governance Framework); Taiuru, K. (20 Sep 2025) Critical Analysis of Te Mana Raraunga Data Principles, taiuru.co.nz/critical-analysis-mana-raraunga/ where applicable.]

Item 21. In financial services, sovereignty primitives apply. Customer records, transaction data, and risk signals are sovereign records of the holding institution, subject to Reserve Bank of New Zealand / Te Pūtea Matua prudential requirements, the Privacy Act 2020, and the Anti-Money Laundering and Countering Financing of Terrorism Act 2009. AML/CFT cooperation is bilateral via established channels — the New Zealand Financial Intelligence Unit and the international FATF channels — and AI assistance is attributed and bounded by these existing regulatory arrangements. We acknowledge the merit of the CAC framework's commitment to financial-risk-control agents for credit approval, transaction monitoring, account security, and anti-money laundering monitoring. We propose that for Aotearoa NZ, the existing institutional and regulatory framework is well-suited to attribution-based audit at the level of each financial institution, with bilateral cooperation through established channels for cross-institutional and international coordination. Cryptographic provenance attaches to every transaction, AI assessment, and human approval — verifiable backwards by AML/CFT auditors, the Reserve Bank, FATF inspectors, and the customer themselves under their respective competences. Federation envelopes mediate cross-institution AML/CFT cooperation: only the consented suspicious-activity signal travels, with recipient binding to the New Zealand Financial Intelligence Unit or relevant counterpart. Member-driven portability supports account-portability obligations: the customer's transaction history is exportable to another institution with provenance intact. *(Parallels CAC item 21 "financial services".)* [CITATIONS: Reserve Bank of New Zealand Act 2021; Anti-Money Laundering and Countering Financing of Terrorism Act 2009 (NZ); Privacy Act 2020 (NZ); FATF Recommendations.]

(III) Daily Life

Item 22. In end-user applications, sovereignty primitives apply. Member-portable identifiers replace platform-specific accounts; cross-device coordination is mediated by the member's own keychain or identity wallet. Sovereignty here means the user holds the records — whether the application is built by an Aotearoa NZ vendor or an international one. We acknowledge the merit of the CAC framework's commitment to intelligent agents empowering internet applications and services across mobile phones, computers, vehicles, home appli-

ances, wearables, and consumer-grade robots. We propose that for any application operating against user records, the architectural primitives of attribution and member-portability apply regardless of vendor jurisdiction. The operative primitive is member-driven portability, operationalised via W3C Decentralized Identifiers and Verifiable Credentials: the user's identity is held in their own keychain (or wallet), with cross-device coordination mediated by their own keys. Cryptographic provenance attaches to every agent action taken against the user's records, attributable to the agent and to the user's authorisation framework. Federation envelopes mediate cross-vendor coordination only when the user authorises it. *(Parallels CAC item 22 "end-user applications".)* [CITATIONS: W3C Decentralized Identifiers (DIDs) v1.0; W3C Verifiable Credentials Data Model v1.1; Privacy Act 2020 (NZ), information privacy principle 7 (correction).]

Item 23. In culture and tourism, sovereignty primitives apply. Cultural content is governed by its creators; in the Aotearoa NZ context, kaitiaki obligations over taonga are central to how AI agents may interact with cultural material. Translation agents preserve attribution and cultural context; their outputs do not stand in for the original mātauranga, and what counts as appropriate use in te ao Māori contexts is for tangata whenua to determine. Visitor data handled by tourism services is treated as the visitor's sovereign record. We acknowledge the merit of the CAC framework's commitment to cultural-content-creation agents and tourism-service agents. We propose that for Aotearoa NZ — where mātauranga Māori is taonga under Te Tiriti partnership obligations, and where Dr Karaitiana Taiuru's published work on mātauranga Māori protection in AI training data, alongside Te Kāhui Raraunga's Māori AI Governance Framework, is foundational scholarship — the architectural primitives provide audit infrastructure, and the substantive determination of appropriate use is for the holders of the mātauranga. Cryptographic provenance attaches to cultural material: who created it, under what authority, with what use scope. For mātauranga, the holders' tikanga determines what authority-to-control means in practice; the substrate provides the audit infrastructure so that breach of consent is forensically reconstructable, not merely contractually deniable. Federation envelopes carry only the consented subset of mātauranga across installation boundaries, with non-onward-forwarding by default — translation agents inherit but cannot relicense. Visitor data carries the visitor's identity attestation; member-driven portability means the visitor exports their tourism-data record on departure. *(Parallels CAC item 23 "culture and tourism".)* [CITATIONS: Taiuru, K. — mātauranga Māori protection in AI training data (specific publications pending verification); Te Kāhui Raraunga (kahuiraraunga.io — Māori Data Governance Model and Māori AI Governance Framework); Taiuru, K. (20 Sep 2025) Critical Analysis of Te Mana Raraunga Data Principles, taiuru.co.nz/critical-analysis-mana-raraunga/; CARE Principles (Carroll et al. 2020); Wai 262 (Waitangi Tribunal Report on Indigenous Flora and Fauna and Cultural Intellectual Property).]

Item 24. In commercial services, sovereignty primitives apply. Customer interactions create records; both parties — operator and customer — hold provenance copies; disputes are coordinated bilaterally. Embodied agents in retail, hospitality, aged care, and disability support operate under the deployment-holder's authority and produce auditable records of their actions. We acknowledge the merit of the CAC framework's commitment to 24/7 customer service, embodied intelligent agents for guidance, cleaning, warehousing, and distribution in commercial venues, and embodied agents for domestic help, elderly care, childcare, and disability support. We propose that for Aotearoa NZ, all such applications operate under existing consumer-protection, care-quality, and disability-services regulatory frameworks. Both parties (operator and customer) hold cryptographically-signed provenance copies of each interaction, so disputes can be resolved against a shared verifiable record rather than against unilateral platform logs. Embodied agents in retail, hospitality, aged care, and disability support operate under boundary enforcement: cultural-context-dependent or values-laden decisions (medication-administration overrides, care-plan changes, disability-support escalations) route to human deliberation by default. The federation envelope mediates handover

of care-record-class data between providers. *(Parallels CAC item 24 "commercial services".)* [CITATIONS: Consumer Guarantees Act 1993 (NZ); Fair Trading Act 1986 (NZ); Health and Disability Services (Safety) Act 2001 (NZ); New Zealand Disability Strategy.]

(IV) Public Welfare

Item 25. In education, sovereignty primitives apply. Learning records are the student's sovereign record, with co-stewardship where the student is a minor; teaching materials produced by agents are attributed; institutional records — rolls, assessments, qualification records — follow existing institutional governance under the Education and Training Act 2020. Portability is to the student, with appropriate institutional arrangements for handover at transitions between providers. We acknowledge the merit of the CAC framework's commitment to lesson-material generation, homework marking, learning progress analysis, personalised learning plans, and virtual teaching assistants. We propose that for Aotearoa NZ, these capabilities operate under the Privacy Act 2020 and the Education and Training Act 2020, with student-data sovereignty maintained throughout. Cryptographic provenance attaches to every assessment, every agent-produced teaching artefact, every qualification award — attributable to the agent, the supervising educator, and the institution. The student's sovereign record is portable: at every transition (school-to-school, school-to- university, between providers, between countries), the student carries their full record bundle with provenance intact to the receiving institution. Boundary enforcement routes assessment-altering and credential-affecting decisions to human deliberation by default. *(Parallels CAC item 25 "education and teaching".)* [CITATIONS: Education and Training Act 2020 (NZ); Privacy Act 2020 (NZ); New Zealand Curriculum.]

Item 26. In healthcare, sovereignty primitives apply. Patient records are the patient's sovereign record under the Health Information Privacy Code 2020 and Te Whatu Ora / Health New Zealand stewardship structures; diagnostic agents produce attributed outputs; treatment recommendations carry provenance; coordination between providers occurs via Health Information Standards Organisation (HISO) channels and Te Whatu Ora interoperability arrangements. We acknowledge the merit of the CAC framework's commitment to medical-imaging analysis, disease-diagnosis reasoning, personalised treatment plans, medication management, surgical scheduling, and medical-records management agents. We propose that for Aotearoa NZ, these capabilities operate under the existing health-information governance framework, with patient sovereignty over health records maintained as the architectural baseline. Cryptographic provenance attaches to clinical records, AI diagnostic outputs, treatment recommendations, and medication administration — each entry attributable to the writer, immutable to silent post-hoc edit (corrections are counter-signed amendments, themselves recorded). Federation envelopes carry referrals: only the consented clinical subset travels from the referring provider to the receiving one, with recipient binding and non-onward-forwarding by default. Member-driven portability lets the patient export their record bundle to another provider — public, private, or international — without the original holder's permission, with provenance intact at the destination. Boundary enforcement routes clinically-uncertain and values-laden decisions (end-of-life, contested diagnoses, mental-health competence) to human deliberation rather than autonomous agent action. *(Parallels CAC item 26 "healthcare".)* [CITATIONS: Health Information Privacy Code 2020 (NZ); Pae Ora (Healthy Futures) Act 2022 (NZ); HISO data standards.]

Item 27. In employment and labour, sovereignty primitives apply. Employment records, training certifications, and dispute records are sovereign to the parties; mediation operates under existing Employment Mediation Service governance; AI assistance is attributed and bounded by the existing tripartite (worker / employer / state) structure of NZ labour law. We acknowledge the merit of the CAC framework's commitment to agents for employment promotion, training and assessment of technical personnel, labour-relations services, social insurance, labour-dispute arbitration, and wage-arrears management. We

propose that for Aotearoa NZ, these capabilities operate under the Employment Relations Act 2000 and the associated tripartite framework, with attribution and provenance applied throughout. Cryptographic provenance attaches to employment records, training certifications, and dispute records — attributable to the parties involved. Federation envelopes mediate worker-portability of training credentials across employers without loss of provenance. Boundary enforcement routes hiring, firing, disciplinary, and dispute-mediation decisions to human authority — autonomous agent action against an individual worker's employment status is structurally blocked. *(Parallels CAC item 27 "human resources".)* `[CITATIONS: Employment Relations Act 2000 (NZ); Holidays Act 2003 (NZ); Human Rights Act 1993 (NZ); New Zealand tripartite labour-relations framework.]`

Item 28. In information services, sovereignty primitives apply. Content is attributed to its creators; recommendation agents operate against the user's sovereign profile, which the user can inspect, export, and port; editorial review remains a human function. Where AI agents produce content, attribution is to the agent and to the human or organisational operator on whose authority it acted; AI-generated content disclosure is the baseline architectural commitment. We acknowledge the merit of the CAC framework's commitment to intelligent agents for online-content construction, user analysis, topic planning, editorial processing, distribution and recommendation, content review, opinion guidance, emotional support, and real-time translation. We propose that for Aotearoa NZ, attribution requirements apply to all such applications, with the existing broadcasting-standards and harmful-digital-communications framework providing the regulatory context. Cryptographic provenance attaches to every piece of content: author identity, AI-vs-human attribution, editorial-review chain. Federation envelopes carry distribution decisions: each recommendation surface receives only the content the upstream installation has signed and consented to share, with non-onward-forwarding by default. Member-driven portability gives the user their interaction history and recommendation profile in a portable form — they can move to another service without losing their content history. Boundary enforcement routes editorial decisions to human authority — autonomous agent action on what to amplify or suppress is structurally blocked. *(Parallels CAC item 28 "information services".)* `[CITATIONS: Broadcasting Act 1989 (NZ); Harmful Digital Communications Act 2015 (NZ); Privacy Act 2020 (NZ); pending lookup for AI-content attribution standards.]`

(V) Social Governance

Item 29. In public administration, sovereignty primitives apply. Citizen interactions with the state produce records held by both citizen and state; member-held identity credentials migrate over time toward member control; agentic assistance in approval processes is attributed and bounded by Administrative Law principles. Crown agencies remain accountable through the Public Service Act 2020, the Official Information Act 1982, the Privacy Act 2020, the Algorithm Charter for Aotearoa New Zealand, and the Public Records Act 2005. We acknowledge the merit of the CAC framework's commitment to administrative-approval, policy-consultation, and proactive-service-delivery agents. We propose that for Aotearoa NZ, all such Crown-agency applications operate within the existing accountability framework, with the architectural primitives providing the audit infrastructure consistent with the Algorithm Charter's commitments to transparency and partnership with Māori. Cryptographic provenance attaches to every administrative action: who decided, on what authority, against what citizen record, with what agent assistance. Federation envelopes mediate inter-agency coordination — only the consented subset of citizen records travels from agency to agency, with audit trail. Member-driven portability operationalises Privacy Act information privacy principle 6 (access rights) and the Public Records Act 2005's preservation obligations — the citizen can export their full government-interaction record under their own keys. Boundary enforcement routes administrative-discretion decisions, Treaty-obligation-engaged decisions, and rights-affecting decisions to human authority. *(Parallels CAC item 29 "public administra-*

tion services".) [CITATIONS: Public Service Act 2020 (NZ); Official Information Act 1982 (NZ); Privacy Act 2020 (NZ); Algorithm Charter for Aotearoa New Zealand (2020); Public Records Act 2005 (NZ).]

Item 30. In judicial services, sovereignty primitives apply. Court records, evidence, and legal documents are governed by existing court processes; AI assistance is attributed; evidence chain-of-custody is cryptographic where applicable; access controls follow existing judicial governance. Self-represented-litigant support tools that use AI disclose their use and produce auditable provenance. We acknowledge the merit of the CAC framework's commitment to end-to-end case-handling assistance, legal-document generation, legal-publicity, legal-consultation, and legal-supervision agents. We propose that for Aotearoa NZ, all such applications operate under the Senior Courts Act 2016, the Evidence Act 2006, and the established court rules and practice notes governing AI use in legal proceedings. Cryptographic provenance attaches to every piece of evidence introduced, every AI-assisted legal-document draft, every search result. Evidence chain-of-custody is cryptographically anchored — admissibility questions answerable from the substrate rather than from disputed platform logs. Member-driven portability means the self-represented litigant can carry their full case record bundle between forums (tribunal → court → appeal) with provenance intact. Boundary enforcement routes value-judgement and discretionary-decision content (whether to plead, whether to accept a settlement, whether to take an action that prejudices the litigant) to the human litigant or their counsel — autonomous agent action against the litigant's legal position is structurally blocked. *(Parallels CAC item 30 "judicial services".)* [CITATIONS: Senior Courts Act 2016 (NZ); Evidence Act 2006 (NZ); pending lookup for current court guidance on AI use.]

Item 31. In public safety, sovereignty primitives apply. Surveillance is governed by existing legislation — the Privacy Act 2020, the Search and Surveillance Act 2012, and the Intelligence and Security Act 2017 — and any AI agents operating in public-safety contexts produce auditable provenance under those frameworks. Behaviour-monitoring agents operate within the reach already lawful under those statutes. We acknowledge the merit of the CAC framework's commitment to monitoring and early-warning agents, emergency-response and rescue-coordination agents, and abnormal-behaviour identification and dynamic-prevention applications. We propose, for the Aotearoa NZ context, that the architectural contribution of attribution and provenance is to make agentic AI in public-safety contexts auditable; whether and how such capabilities should be deployed is a values decision for the relevant legislative and policy framework, addressed to Parliament and the responsible Ministers, with the architecture providing the audit infrastructure within which those decisions become tractable. Cryptographic provenance attaches to every surveillance signal, every behaviour-monitoring decision, every emergency-response action — making post-hoc accountability tractable in a way unaccompanied agent action is not. Federation envelopes mediate inter-agency public-safety coordination: what intelligence travels between Police, the Government Communications Security Bureau, and emergency-response agencies is bound by what each signed off as releasable for what specific purpose. Boundary enforcement is load-bearing here: rights-engaging decisions (search, arrest, surveillance authorisation, use-of-force) route to human authority — the architectural floor under which no agent autonomously acts. *(Parallels CAC item 31 "public safety".)* [CITATIONS: Privacy Act 2020 (NZ); Search and Surveillance Act 2012 (NZ); Intelligence and Security Act 2017 (NZ); New Zealand Bill of Rights Act 1990.]

Item 32. In urban governance, sovereignty primitives apply. Urban data — sensor networks, planning data, building consents, infrastructure operating data — is held by councils as sovereign records; agentic systems operating in council functions are attributed and accountable through the Local Government Act 2002 and council governance structures. We acknowledge the merit of the CAC framework's commitment to urban-planning, urban-construction, and urban-governance intelligent agents, including for smart construction, building manage-

ment, and urban infrastructure operation. We propose that for Aotearoa NZ, all such applications operate within existing local-government accountability arrangements, with the architectural primitives providing the audit and attribution infrastructure. Cryptographic provenance attaches to every sensor reading, planning decision, building consent, and infrastructure operating choice — auditable by residents, by Local Government NZ, by the Auditor-General, and by successive councils. Federation envelopes mediate inter-council coordination and central-local data sharing — only the consented subset travels. Member-driven portability applies to resident-level data: a resident can carry their interactions with the council out when they move districts. Boundary enforcement routes Treaty-engaged and culturally-significant decisions (urupā, wāhi tapu, taonga management) to the appropriate tangata-whenua deliberation rather than to autonomous agent action. *(Parallels CAC item 32 "urban governance".)* [CITATIONS: Local Government Act 2002 (NZ); Building Act 2004 (NZ); Resource Management Act 1991 (NZ).]

Item 33. In procurement, sovereignty primitives apply. Tender records, evaluations, and contracts are sovereign records of the contracting entity; agentic assistance in procurement is attributed and bounded by Government Procurement Rules and applicable contract law; transparency is via existing OIA-compliant publication. We acknowledge the merit of the CAC framework's commitment to end-to-end intelligent management of tendering and bidding processes, with intelligence applied to transactions, services, and supervision. We propose that for Aotearoa NZ, the Government Procurement Rules and the existing public-procurement framework provide the appropriate accountability context, with attribution and provenance applied throughout. Cryptographic provenance attaches to every tender record, evaluation, contract amendment, and award decision — published under existing OIA-compliant transparency arrangements with substrate-level integrity properties. Federation envelopes mediate the consortium and supply-chain coordination procurement requires, without exposing competitively-sensitive data outside the consented scope. Boundary enforcement routes the discretionary procurement decisions (award, override, exception) to human authority — autonomous agent action on a contract award is structurally blocked. *(Parallels CAC item 33 "tendering and bidding".)* [CITATIONS: Government Procurement Rules (NZ); Public Records Act 2005 (NZ); Official Information Act 1982 (NZ).]

---

§V. Building a Federated Ecosystem

Where the Cyberspace Administration of China's framework envisions an industry-cluster ecosystem with national-champion projection through international AI conferences, we offer a federated ecosystem where coordination occurs by bilateral federation between sovereign peers and where international alignment is via established standards bodies. The two subsections that follow — promoting federated cooperation, and strengthening bilateral promotion — together specify how an ecosystem of sovereign installations sustains itself and engages internationally.

(I) Promoting Federated Cooperation

Item 34. Open-source under permissive licences. Reference implementations should be available under permissive open-source licences. The current MDSL implementations are one set of references among potentially several: the Tractatus framework is distributed under Apache 2.0 for code and CC BY 4.0 for documentation; the Village and community codebases are migrating toward EUPL-1.2 (European Union Public Licence) in phases as of mid-2026; future MDSL contributions are intended to be EUPL-1.2 where practicable, for sovereignty alignment with European Union sovereignty work and for compatibility with bilateral federation between sovereign installations across multiple jurisdictions. We acknowledge the merit

of the CAC framework's commitment to fostering open-source innovation, including domestic-AI open-source communities, compatibility with open-source chips, operating systems, and large models, and engagement of enterprises, universities, and research institutions in open-source projects. Open-source under permissive licences is bilateral-federation-friendly: each sovereign installation forks the upstream, contributes back via pull request, and takes its own deployment decisions. *(Parallels CAC item 34 "foster open-source innovation".)* [CITATIONS: Apache 2.0 (Apache Software Foundation); EUPL-1.2 (European Union Public Licence); CC BY 4.0 (Creative Commons).]

Item 35. Federation by publication. Where coordination is needed on common technology, interoperability standards, security incident response, or audit framework development, it occurs through open publication and consensus among contributing installations. International alignment is via W3C, IETF, ISO/IEC, and similar established standards bodies. We acknowledge the merit of the CAC framework's commitment to industry-collaboration platforms — including intelligent-agent ecosystem alliances, technology-verification laboratories, and joint R&D arrangements — and to coordination of upstream and downstream supply-chain participants in common technology R&D, standards-setting, and assessment-and-certification work. **This area is workstream (iv) of the single committee proposed in §II item 4. The committee would develop NZ-context recommendations on federation patterns and alliance patterns for industry coordination, contribute to ISO/IEC SC42 work on AI industry-collaboration models, and engage in bilateral dialogue with the CAC framework's authors on the interaction between federated and alliance-based industry coordination.** *(Parallels CAC item 35 "industry collaboration platforms"; consolidated committee-formation workstream applies.)* [CITATIONS: W3C process document; IETF Request for Comments process; ISO/IEC 42001:2023 management systems.]

(II) Strengthening Bilateral Promotion

Item 36. Adoption is bilateral. Each sovereign installation reaches its counterparties directly — partner organisations, peer institutions, federated peers. We acknowledge the merit of the CAC framework's commitment to application-promotion channels, including intelligent-agent software stores, industry supply-demand information platforms, customised product development via tendering and the "unveil-and-take-the-helm" challenge model, and hardware-system and software enterprise development of intelligent-agent products and services. We propose, for the Aotearoa NZ context, that adoption channels arise from the existing commercial, civil-society, and institutional landscape; sovereign installations build their counterparty relationships through ordinary direct engagement, with public procurement following the Government Procurement Rules. *(Parallels CAC item 36 "application promotion channels".)* [CITATIONS: Government Procurement Rules (NZ); pending verification for any current NZ procurement reforms.]

Item 37. Pilot deployment is bilateral and evidence-led. Sovereign installations pilot adoption with willing communities directly. Existing MDSL deployments — Village in parish and hapū / iwi contexts; family-history in iwi and diaspora contexts; sydigital in small-business contexts — are examples; specific deployment data (counts, start dates, tenancy scope) is to be added before v1 publication. We acknowledge the merit of the CAC framework's commitment to driving the opening of intelligent-agent application scenarios in key sectors, with pilots in industrial clusters, key industries, and key sectors building a portfolio of demonstration projects. We propose that for Aotearoa NZ, pilot deployment is bilateral between deploying installations and their willing communities. Where Crown agencies wish to pilot agentic AI, they do so under existing Privacy Impact Assessment processes, the Algorithm Charter for Aotearoa New Zealand, and Te Mana Raraunga / Māori Data Sovereignty Network commitments. *(Parallels CAC item 37 "advance the opening of key scenarios".)* [CITATIONS: MDSL deployment evidence — Village (parish + community contexts), family-history (iwi + diaspora contexts), sydigital (small-business contexts), specific data pending

operator-verified figures before v1 publication; Algorithm Charter for Aotearoa New Zealand (2020); Te Kāhui Raraunga (kahuiraraunga.io — Māori Data Governance Model and Māori AI Governance Framework); Taiuru, K. (20 Sep 2025) Critical Analysis of Te Mana Raraunga Data Principles, taiuru.co.nz/critical-analysis-mana-raraunga/.]

Item 38. International alignment by bilateral federation. Sovereign installations in Aotearoa New Zealand federate bilaterally with sovereign installations in other jurisdictions; international standards engagement occurs through W3C, IETF, ISO/IEC, and similar fora as peer participation. We acknowledge the merit of the CAC framework's commitment to actively cultivating the global ecosystem through international platforms such as the World Artificial Intelligence Conference and the World Internet Conference, promotion of intelligent-agent adaptation by terminal-device and software enterprises, and engagement on overseas compliance and adaptation to local laws, regulations, and cultural customs. **This area is workstream (v) of the single committee proposed in §II item 4. The committee would develop NZ-context recommendations on international AI cooperation, contribute to ISO/IEC SC42 international standards work, and engage in bilateral dialogue with the CAC framework's authors and with international peers on interoperability between bilateral-federation and platform-projection approaches to international cooperation.** We offer this as one contribution to an early-stage international conversation; contributions across many architectural traditions and political contexts will improve the field. *(Parallels CAC item 38 "actively cultivate the global ecosystem"; consolidated committee-formation workstream applies.)* [CITATIONS: ISO/IEC JTC 1/SC 42; W3C international standards process; pending lookup for current NZ bilateral AI agreements and international engagements.]

---

§VI. Safeguarding Adoption

As a civil-society proposer, My Digital Sovereignty Ltd does not coordinate adoption. We name here the bodies whose participation would be required if any part of this framework were to be adopted by Aotearoa New Zealand entities.

Crown agencies whose work this proposal touches include the Ministry of Business, Innovation and Employment for digital strategy; the Ministry of Justice for legal framework alignment; the Office of the Privacy Commissioner for Privacy Act 2020 alignment; Stats NZ and Te Kāhui Raraunga for data-sovereignty alignment (with Dr Karaitiana Taiuru's 20 September 2025 critical analysis as foundational reference); Te Whatu Ora / Health New Zealand for health-information governance; Te Pūtea Matua / Reserve Bank of New Zealand for financial-services prudential alignment; Waka Kotahi New Zealand Transport Agency for transport; the Ministry of Education for education; and New Zealand Police for public-safety contexts. Civil-society evaluation would naturally involve Royal Society Te Apārangi, Internet NZ, NetSafe, the New Zealand AI Forum, and academic researchers across the relevant disciplines. Hapū and iwi consideration is essential where Treaty obligations or Settlement implications arise, and the architecture this proposal specifies is intended to support — and is offered for use under — Māori data sovereignty work as articulated by Te Kāhui Raraunga (Māori Data Governance Model; Māori AI Governance Framework) and by Dr Karaitiana Taiuru's published scholarship — including his 20 September 2025 critical analysis that informs why earlier framings are inadequate for AI contexts.

International dialogue with the authors of the CAC framework and with peer Indigenous Data Sovereignty networks — FNIGC in Canada, USIDSN in the United States, Maiam nayri Wingara in Australia, GIDA internationally — would enrich both directions of the conversation.

My Digital Sovereignty Ltd commits to the architectural openness and licence-openness ele-

ments of the proposal: the Tractatus framework, the Village and community codebases, and future MDSL contributions will remain available under permissive open-source licences, and the reference implementations will be developed in dialogue with adopters. The rest is addressed to those who would decide adoption.

We close with an explicit invitation: to the CAC framework's authors, to international peers, to NZ policymakers and community organisers, and to anyone working on parallel questions — comments on this v1 are welcome via the standing paper-comments channels on agentic-governance.digital.

Appendix A. Common technical objections + responses

This appendix collects the most common technical objections raised against the Tractatus framework, with brief structural responses. Each response refers back to the relevant §0(i) primitive or later section. The framework is research-stage; these responses summarise the architectural reasoning, not engineering completion.

1. "How can an LLM enforce anything?"

Objection. LLMs are statistical pattern-completers; adversarial prompts and emergent capabilities bypass model-level guardrails.

Response. The framework does not ask the model to enforce anything. The Tractatus primitives are runtime intercepts — code in the agent's host process that decides whether each proposed action executes, before the model's output reaches the world. The intercept does not depend on the model's compliance or alignment training; it applies configured decision-class rules at the action boundary, the same architectural shape as OS kernel privileged-syscall interception. See §0(i) Boundary enforcement.

2. "What about JEPA / alternative architectures?"

Objection. The framework appears to assume transformer-based LLMs; architectural shifts (JEPA, hybrid systems) may invalidate it.

Response. The framework is substrate-agnostic. Its primitives operate on agent actions, not on model internals — no access to weights, attention patterns, or reasoning traces required. An agent on any architecture emits proposed actions; the intercept layer evaluates those actions against the same decision-class boundaries. See §0(i) leading paragraph.

3. "Can the agent route around the intercept?"

Objection. A capable enough agent could model the intercept and engineer actions that bypass it.

Response. The agent does not hold the privilege to execute — the intercept does. Bypass requires either (a) operating outside the agent's capability scope, which the runtime catches as privilege violation, or (b) full privilege escalation against the host process, an OS-level concern hardened the same way file-system permissions handle it. The audit trail (sovereign records, cryptographically signed) makes successful bypass attempts forensically visible, removing defection-as-strategy. The deeper structural point — developed in *Architectural Alignment* §3.4 — is the **substrate-vs-runtime distinction**: even where a sufficiently capable agent reasons around runtime intercepts, the substrate mechanisms (cryptographic provenance, federation envelopes, member-driven portability of records) are not reasonable-around because they live in distributed possession independent of the agent. The substrate's

safety follows from mathematics and distributed replication, not from the agent's cooperation. See §0(i) Cross-reference validation; §II item 5; *Architectural Alignment* §3.4 substrate-vs-runtime; §7.5 social-layer attack surface (the surface the substrate does *not* close).

4. "How is this different from prompt-engineering safety?"

Objection. Prompt-engineering and RLHF also constrain model output. The framework appears similar in spirit.

Response. Structurally different. Prompt-engineering and RLHF modify the model's output distribution but leave the same statistical mechanics. The framework's primitives run before model invocation (capability-scoping), after proposed action (boundary enforcement), or alongside invocation (cross-reference validation) — none depend on the model producing the right output. They depend on the runtime layer correctly identifying decision-class membership. See §0(i) Boundary enforcement and Metacognitive verification.

5. "What if the runtime service itself is exploited?"

Objection. Putting trust in a runtime service moves the attack surface rather than removing it.

Response. The framework does not claim runtime services are unhackable. It claims defection has receipts and that the receipts constrain the blast radius.

What a breach looks like. Either the service code is exploited — an attacker gets the intercept to approve actions outside policy — or the policy state the service consults is rewritten — an attacker changes which decision classes route to human approval. In both cases the attacker's goal is to convert a "route to human" decision into an "auto-approve" decision without the operator noticing.

What is at risk. Decisions the framework would otherwise have routed to human approval — values judgements, irreversible operations, cross-tenant data access. The blast radius is bounded by what the intercept was already authorised to approve: the framework grants permissions to approve, not new privileges to perform actions. An exploited intercept cannot exfiltrate data the agent never had access to in the first place; it can only mis-approve actions within the agent's existing capability scope.

Remedies. Three properties combine. (i) Every framework decision is written to the sovereign-record audit trail — cryptographically signed, append-only, federation-replicated to peers — so post-hoc forensic analysis can reconstruct what was approved, by which version of the service, against which policy state. The breach window is bounded in time and scope. (ii) Cross-reference validation (§0(i)) catches divergence between observed approvals and declared policy in near-real-time, surfacing breaches before they normalise. (iii) Federation replication prevents an attacker controlling one node from retroactively erasing records; defection requires colluding with the federation, not compromising a single service.

The trust anchor can fail, but failure is bounded, observable, and forensically reconstructable — the same architectural pattern as Certificate Transparency for the TLS PKI: the trust anchor (a Certificate Authority) can be compromised, but the audit log of issued certificates makes the compromise globally visible.

Survival posture is layered. The framework's runtime services (BoundaryEnforcer, the §0(i) primitives) are *agent-facing*: they constrain what the agent can authorise. The sovereign-records architecture (Paper A) is *substrate-facing*: it ensures the records survive the agent regardless of whether the agent escapes the gate. See *Architectural Alignment* §7.4 (survival posture independent of agent containment) and §7.5 (the social-layer attack surface the substrate does not close — persuasion, mass-coordinated impersonation, synthesised consent —

named here as an open frontier). The PKI signature scheme underpinning the audit chain has a quantum-vulnerability horizon (10-30 years); NIST post-quantum signature standards were finalised August 2024 and the migration path is standards-track, with per-record forgery costly even on a CRQC and federation/portability mechanisms not depending on signature integrity. See *Paper A* §5.3.

See §II item 5; §0(i) Cross-reference validation.

6. "What if values change?"

Objection. Decision-class boundaries lock in current values; communities' values evolve, making the framework brittle.

Response. Decision classes are configuration, not architecture. The framework provides the intercept mechanism; which action classes route to human approval is operator-editable per tenant (§III item 3 — adopter-driven governance). Where stakeholders within a governance scope hold incompatible boundary positions, the pluralistic deliberation orchestration primitive structures the deliberation rather than picking a winner. The framework is designed to host evolving values, not freeze them. See §0(i) Pluralistic deliberation orchestration; §III item 3.

Licence and citation

Suggested citation: Stroh, J. G. (2026). *A Civil-Society Proposal for Sovereign and Federated Agentic AI in Aotearoa New Zealand* (v1.2, May 2026, revised per Ted Howard's correspondence on v1.1). My Digital Sovereignty Ltd. https://agenticgovernance.digital/papers/aotearoa-nz-agentic-ai-framework-v1.2-may-2026.html

Comments and correspondence: Substantive feedback engaging specific sections is welcomed. Please cite section numbers (e.g. §III item 5) so corrections can be traced. The author replies personally; allow one to two weeks. Email: john.stroh@mysovereignty.digital.