

Precis: Taonga-Centred Steering Governance

Document Type: Technical Documentation

Generated: February 10, 2026

Tractatus AI Safety Framework

<https://agenticgovernance.digital>

Precis: Taonga-Centred Steering Governance

Polycentric Authority for Sovereign Small Language Models

STO-RES-0010 v0.1 DRAFT — Stroh & Claude (2026)

The paper addresses a governance problem that arises when communities deploy their own language models rather than consuming commercial AI through APIs.

Sovereign small language models — locally hosted, with full access to model weights — permit a class of intervention unavailable to API consumers: direct modification of the model's internal representations at inference time through steering vectors. These interventions determine how the model represents kinship, place, authority, grief, and spiritual practice. They are, in substance, instruments of norm enforcement.

The question the paper poses is not technical but political: who governs these instruments?

The prevailing architecture assumes a single governance root. The platform operator defines what constitutes bias, extracts the corrections, and distributes them to all downstream deployments. Communities may customise within the limits the platform sets, but they cannot contest the root definitions. This is a delegation model: authority flows downward from the platform.

The paper argues that for domains of Maori cultural knowledge — whakapapa, tikanga, kawa, the mana of kaumatua and kuia — this hierarchy is structurally inappropriate regardless of the platform operator's intentions. It proposes a polycentric alternative, drawing on Ostrom and on the principles your work has established through *Indigenous Data Sovereignty* and Te Mana Raraunga.

The core of the proposal is a distinction between delegation and recognition. In the delegation model, the platform accommodates indigenous governance as a feature. In the recognition model, iwi authority exists prior to and independently of the platform, and the architecture either respects that independence or undermines it. The paper argues for recognition, and develops three architectural consequences:

First, steering packs that encode iwi knowledge are treated as taonga — subject to kaitiakitanga, with iwi-controlled lifecycles, access conditions, and constraints on redistribution that the platform cannot override.

Second, iwi governance bodies operate as co-equal steering authorities alongside the platform, not beneath it. Each maintains its own registry, its own bias ontology, its own review processes. The model's activation space is a shared technical substrate, not a constitutional order with the platform at its apex.

Third, iwi hold a right of non-participation. They may refuse to publish steering packs, may withdraw them at any time, and the platform must not fill the resulting space with its own values. The absence of an iwi pack is a governed absence, not a gap.

The paper is honest about what it cannot do. It cannot resolve the tension between platform safety baselines and plural cultural norms — it can only make that tension visible and politically navigable. It cannot create the institutional trust that polycentric governance requires. And it cannot, as a work of non-Maori authorship, speak with authority about the concepts from te ao Maori on which its argument depends.

It is that last limitation which brings us to you.

Reference: Stroh, J. & Claude (2026). Taonga-Centred Steering Governance: Polycentric Authority for Sovereign Small Language Models. STO-RES-0010 v0.1 DRAFT. agenticgovernance.digital

© 2026 Tractatus AI Safety Framework

This document is part of the Tractatus Agentic Governance System

<https://agenticgovernance.digital>