

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 with full access to model weights, rather than consuming AI through commercial APIs.

Such sovereign deployments permit direct modification of the model's internal representations at inference time through steering vectors — interventions that determine how the model represents kinship, place, authority, grief, and spiritual practice. These are instruments of norm enforcement. The paper asks who should govern them, and argues that the answer should not default to the platform operator.

The CARE Principles provide the paper's normative foundation. The proposed architecture attempts to operationalise each:

Collective Benefit is addressed through polycentric governance: indigenous and community authorities maintain co-equal jurisdiction over model behaviour alongside the platform operator, ensuring that corrections to cultural representation serve the communities whose knowledge they encode.

Authority to Control is addressed through "taonga steering registries" — governed repositories, operated by indigenous institutions, that maintain independent control over the creation, versioning, access conditions, and withdrawal of steering packs encoding cultural knowledge. The platform integrates with these registries but cannot encapsulate, fork, or redistribute their contents.

Responsibility is addressed through mandatory steering provenance: every inference records which steering packs were active, from which authorities, at what magnitude, and under what governance terms. This makes norm enforcement attributable and contestable, rather than opaque.

Ethics is addressed through the rejection of a single bias ontology. The paper argues that different governance authorities will define bias differently — and that those definitions may legitimately conflict. The architecture supports multiple bias frameworks simultaneously without requiring reconciliation into a single schema.

The paper is grounded in concepts from te ao Maori — taonga (treasured possessions subject to kaitiakitanga), tikanga (customary practice), tino rangatiratanga (self-determination) — as these are the indigenous governance frameworks most relevant to our context in Aotearoa. The architectural model, however, is intended to be adaptable beyond that context.

Two questions arise on which your perspective would be particularly valuable. First, the paper's "steering provenance" bears an obvious resemblance to what your IEEE 2890-2025 standard addresses at the standards level. We may have arrived independently at a narrower application of the same principle, or we may have missed structural requirements that the standard identifies. Second, whether the polycentric governance architecture — co-equal authorities, governed registries, right of non-participation — transfers meaningfully to indigenous governance contexts beyond Aotearoa, or whether its Maori-specific framing limits its applicability.

The paper is a draft written by a non-indigenous author in collaboration with an AI assistant, without indigenous peer review. We are concurrently seeking review from Professor Tahu Kukutai and Associate Professor Maui Hudson at the University of Waikato. We recognise that the concepts we draw upon carry authority and obligation beyond what we can represent, and we invite correction.

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>