ARCHITECTURAL ALIGNMENT

Community-Governed AI Through Constitutional Infrastructure

A Framework for Sovereign Local Deployment

John Stroh & Claude (Anthropic)

Document Code: STO-INN-0003
Version 2.1-C (Community Edition)
January 2026

Executive Summary

The question is no longer whether AI will be part of community life, but who will govern it when it arrives.

Current AI systems—whether cloud-based assistants or enterprise tools—operate under governance frameworks written by their vendors. Your community's values are accommodated only insofar as they don't conflict with platform policies designed for millions of other users. Your data informs systems you don't control. Your exit rights are limited to what the provider chooses to export.

This paper presents an alternative: constitutional governance for community-controlled AI. The Tractatus Framework implements explicit rules, defined by your community, that constrain what AI systems can do before any action is taken. This isn't about making AI less capable—it's about making AI accountable to the community it serves.

The framework is implemented in the Village platform and designed to support both cloud-based AI and locally-deployed systems. We introduce the concept of Sovereign Locally-trained Language Models (SLLs)—AI systems that run on community infrastructure, adapt to community norms, and operate under community-defined constitutions rather than vendor terms of service.

What This Means for Communities

1. Your rules are the only rules. Constitutional constraints are defined by your community through democratic deliberation, not imposed by distant platform operators.

2. Your data stays yours. AI memory, preferences, and learned patterns remain under community control, with full export rights.

3. Transparency, not trust. Every AI action passes through auditable checkpoints. You don't have to trust that the vendor trained it right—you can see the rules it follows.

4. Gradual autonomy. AI capabilities expand only as your community builds confidence, through staged progression from fully supervised to bounded autonomy.

5. Real exit. If you leave, your governance structures, AI memory, and data leave with you.

The underlying research addresses serious questions about AI safety and alignment. We believe communities benefit from understanding this context—not because your household AI poses existential risks, but because building governance capacity now prepares for a future where such capacity will matter more.

1. The Problem: Who Governs Your AI?

1.1 The Current Reality

When you use a cloud AI assistant—whether for writing, research, or community management—you're interacting with a system governed by rules you didn't write and can't change:

- Training decisions were made by researchers optimising for metrics you weren't consulted about
- Safety constraints reflect corporate liability concerns, not your community's values
- Data handling follows terms of service written by lawyers, not community deliberation
- Capability boundaries are set by platform operators, not local governance
- Exit rights are whatever the provider chooses to offer

This isn't malicious. It's structural. Systems designed to serve millions of users cannot accommodate the specific values, norms, and governance preferences of each community. The result is AI governance by lowest common denominator.

1.2 Why It Matters

For many use cases, generic governance is adequate. A community using AI to schedule meetings doesn't need bespoke constitutional frameworks.

But some communities have legitimate needs that generic platforms cannot address:

Cultural communities may have protocols about who can access certain knowledge, how ancestors are discussed, or what constitutes respectful engagement with cultural heritage.

Family history communities deal with sensitive information about living people, contested narratives, and emotional content that requires context-specific handling.

Professional communities may have ethical requirements, confidentiality obligations, or domain-specific norms that generic AI doesn't understand.

Indigenous communities have collective rights over data, cultural authority over knowledge systems, and governance traditions that predate and don't map onto Western corporate frameworks.

Privacy-conscious communities may want guarantees about data handling that exceed what commercial platforms offer.

For these communities, governance isn't a nice-to-have. It's essential.

1.3 The Coming Shift

The AI landscape is changing in ways that create new possibilities:

Industry research indicates that 72% of enterprise executives expect small language models to surpass large language models in prominence by 2030 (IBM Institute for Business Value, 2026). This suggests a future where capable AI runs on local hardware—home servers, community infrastructure, edge devices—rather than exclusively in distant data centres.

This shift matters because local deployment enables local governance. When AI runs on your infrastructure, under your control, you can implement governance frameworks that reflect your community's values rather than a vendor's policy preferences.

2. The Tractatus Framework: Governance Through Architecture

2.1 The Core Idea

Instead of trusting that AI was trained to behave appropriately, Tractatus requires AI systems to propose actions explicitly and have them evaluated against rules before execution.

```
Your Request       ↓ [AI generates response]       ↓ Structured Proposal (what
the AI wants to do)     ↓ [Constitutional Gate checks against your rules]
↓ Permitted / Denied / Escalated to human review
```

Every significant AI action passes through this checkpoint. The rules are explicit, inspectable, and defined by your community.

2.2 What This Looks Like in Practice

Example: A family history community

A member asks the AI to help write a remembrance for a recently deceased relative.

1. The AI generates a proposed response

2. The constitutional gate checks against community rules:

• Is this about a death within the past year? (Triggers sensitivity protocols)

• Does the community constitution specify cultural requirements for discussing the deceased?

• Does the individual member have preferences about how AI discusses their family?

3. The gate applies relevant rules:

• Use more gentle phrasing (community rule)

• Offer to involve a human moderator (escalation threshold)

• Respect the member's preference for private vs. communal remembrance

4. The response is delivered—or flagged for human review

The AI didn't decide these rules. Your community did. Through conversation, through voting, through deliberation.

2.3 Layered Constitutions

Rules are organised in layers, each with appropriate authority:

Layer	Who Defines It	What It Covers	How It Changes
Core Principles	Platform-wide	Fundamental safety; legal compliance	Rarely; requires broad consensus

| Community Constitution | Your community | Values, norms, policies specific to your context | Community deliberation and vote |
| Individual Preferences | Each member | Personal interaction style, privacy choices | Self-service configuration |

Higher layers take precedence, but within those constraints, your community has genuine authority.

2.4 Progressive Autonomy

AI capabilities don't arrive fully-formed. They're earned through demonstrated trustworthiness:

Stage	AI Authority	Human Role	Duration
Shadow	Observes and proposes; takes no action	Approves everything	Until confident in proposals
Advisory	Recommendations surfaced	Retains full authority	Until acceptance rate stable
Supervised	Acts within narrow scope	Reviews all actions within 24h	Until error rate acceptable
Bounded	Acts within defined boundaries	Reviews samples and edge cases	Ongoing
Operational	Full authority at defined level	Focuses on outcomes	Ongoing with audit

Your community decides when to progress—or regress if problems emerge. The AI earns trust; it doesn't assume it.

3. Sovereign Local AI: The SLL Concept

3.1 What is an SLL?

We introduce the term Sovereign Locally-trained Language Model (SLL) to describe AI systems with specific properties:

- Local deployment: Runs on your infrastructure—a home server, community hardware, or local data centre—not a vendor's cloud
- Local adaptation: Fine-tuned on your community's data and norms, not generic training
- Local governance: Subject to your constitutional rules, not vendor terms of service
- Portable sovereignty: Can connect to larger networks without surrendering governance authority

An SLL isn't just a small model that happens to run locally. It's an architectural commitment to sovereignty.

3.2 Why Sovereignty Matters

Sovereignty in this context doesn't mean isolation. It means the capacity to participate in larger networks on your own terms.

Consider the difference:

Dimension	Vendor-Hosted AI	Sovereign SLL
Where it runs	Vendor's cloud; you don't know where	Your infrastructure; you control location
Who sets the rules	Vendor ToS + whatever law applies to them	Your constitution + your jurisdiction's law
What it learns from	Aggregated data from all users	Your community's data, under your control
Who can change behaviour	Vendor, unilaterally	Your community, through governance
What happens if you leave	Limited export; lose AI context	Full export; AI memory is yours

Sovereignty means real exit rights. If your governance framework isn't working, you can take your data, your AI's learned patterns, and your constitutional rules, and move them elsewhere.

3.3 The Trade-offs

Sovereignty comes with real costs:

You accept:

- Potentially lower raw capability than frontier cloud models
- Higher infrastructure complexity (someone has to run the servers)
- More explicit governance work (constitutions don't write themselves)

In exchange for:

- Complete data sovereignty
- Governance that reflects your community's values
- Real exit rights
- Transparency about what the AI actually does

This trade-off isn't right for everyone. Many communities are well-served by cloud AI with vendor governance. The point is that the choice should be yours, not forced by technological constraints.

4. Constitutional Governance in Action

4.1 The Governance Pipeline

The Village platform implements constitutional governance through a six-stage verification pipeline. Every AI response passes through:

1. Intent Recognition

- What kind of request is this?
- Does it involve values, facts, or actions?
- Route to appropriate handling

2. Boundary Enforcement

- Hard constraints that cannot be overridden
- The AI never tells you what to think about values
- The AI never makes governance decisions

3. Pressure Monitoring

- Is the AI operating under degraded conditions?
- If confidence is low, acknowledge uncertainty
- Escalate when appropriate

4. Response Verification

- Does the response actually address the request?
- Is it complete and structurally sound?
- Pre-flight checks before delivery

5. Source Validation

- Are claims grounded in verifiable sources?
- The AI doesn't present training data as fact
- Citations where appropriate

6. Value Deliberation

- Does the request involve value tensions?
- Present balanced options rather than recommendations
- For community decisions, suggest using democratic processes

This pipeline is operational. Every response you receive has passed through it.

4.2 Democratic Deliberation

Constitutional rules aren't handed down from above. They emerge from community deliberation:

Consent-Based Voting: Not just yes/no, but a spectrum: Enthusiastic Support, Support, Consent (can live with it), Stand Aside, Object. Objections require rationale and trigger discussion—they're invitations to address concerns, not vetoes.

Ranked Choice: When multiple options exist, rank your preferences. Your second choice matters if your first can't win. This prevents the spoiler effects that silence minority views.

Quadratic Voting: For decisions where preference intensity matters, voice credits let you express how much you care. Spend more on issues that matter deeply; less on those where you're indifferent.

Phased Deliberation: Important decisions move through stages:

1. Discussion Phase: Share perspectives, no voting pressure

2. Preliminary Vote: Temperature check on emerging consensus

3. Final Vote: Binding decision with full participation

AI Assists, Never Decides: The AI can help with deliberation—summarising threads, highlighting patterns, suggesting when consensus is emerging. But it never votes, never decides when discussion is complete, never creates policy without community approval.

4.3 Governance Example

Consider a community debating: "Should our AI remember individual preferences about how to discuss recent deaths?"

Discussion surfaces:

- Privacy concerns (some want AI to forget)

- Helpfulness concerns (others want AI to remember)

- Cultural differences (mourning periods vary)

Consent vote reveals:

- 60% Support

- 30% Consent (can live with it)

- 10% Stand Aside

Amendment proposed: Allow individual opt-out

Final vote: Approved with amendment

The result becomes part of your community constitution. The AI implements it. If circumstances change, you can revisit.

5. Why This Architecture?

5.1 The Deeper Context

This framework emerges from serious research on AI safety and alignment. The core insight:

Training AI to be safe isn't sufficient. We can't verify what's inside a neural network. We can't prove it will behave as intended under conditions not encountered during training. We can't guarantee that systems optimised for helpfulness won't develop unintended behaviours.

Architecture provides guarantees that training cannot. By requiring AI to propose actions explicitly, by evaluating those proposals against explicit rules, by logging everything for audit—we create visible, enforceable constraints that don't depend on trusting the AI's internal state.

This matters more as AI systems become more capable. The patterns developed now—at village scale, for manageable stakes—become the patterns available when stakes are higher.

5.2 Not Just Governance, But Preparation

We don't claim your household AI poses existential risks. We do believe that:

1. AI systems will become more capable

2. More capable systems will require more robust governance

3. Governance capacity can't be created instantly when needed

4. Building governance infrastructure now prepares for futures we cannot fully predict

Constitutional governance for community AI is practice. It develops the tools, the expertise, the governance culture, and the democratic capacity that may matter more later.

This is preparation, not prediction. We don't know what the future holds. We do know that having governance infrastructure is better than not having it.

6. Indigenous Sovereignty

6.1 First Principles

The Village platform is developed in Aotearoa New Zealand, under Te Tiriti o Waitangi. This context shapes our approach:

Sovereignty isn't new. Many ideas now emerging under "digital sovereignty" were articulated first by indigenous leaders: collective rights over data, self-determination in how knowledge is used, guardianship rather than ownership.

Te Tiriti creates constitutional obligations. Article Two guarantees Māori tino rangatiratanga (self-determination) over taonga (treasures), which includes language, culture, and knowledge systems. Data is taonga. AI trained on data, and systems that process it, engage these obligations.

Te Mana Raraunga principles guide Māori data sovereignty:

- Rangatiratanga: Māori authority over Māori data

- Whakapapa: Data exists in relational context

- Whanaungatanga: Governance is collective, not just individual

- Kaitiakitanga: Custodians have guardianship responsibilities

6.2 Implications for Constitutional Governance

Tractatus' layered constitution architecture can accommodate these requirements:

- Community constitutions can instantiate tikanga-based rules

- Collective consent can be encoded alongside individual consent

- Cultural authority can be recognised in governance structures

- Benefit-sharing can be specified as constitutional constraint

This doesn't solve all problems. Platform-level accommodation is not a substitute for legislative recognition. But it demonstrates that constitutional governance can respect rather than override indigenous sovereignty.

6.3 Relevance for Other Communities

Indigenous data sovereignty principles offer insights for any community concerned with governance:

- Data about communities belongs to those communities

- Governance is relational, not just transactional

- Exit rights matter—sovereignty means capacity to say no

- Collective interests don't reduce to individual preferences

These aren't uniquely indigenous insights. They're wisdom applicable to any community seeking genuine self-governance.

7. Practical Considerations

7.1 What You Need

For cloud-hosted Village tenancy:

- Internet connection
- Governance capacity (someone to configure and maintain constitutions)
- Community commitment to deliberative process

For sovereign SLL deployment:

- Local server infrastructure (home server or community hardware)
- Technical capacity for deployment and maintenance
- Governance capacity as above

Village supports both modes. Communities can start with cloud hosting and migrate to sovereign deployment as capacity develops.

7.2 What It Costs

Infrastructure: Non-trivial for sovereign deployment. Comparable to running any local server infrastructure.

Governance: Real time investment. Constitutions need to be developed, deliberation takes time, rules need maintenance.

Capability trade-off: Local models may not match frontier cloud models on raw capability. Hybrid approaches (local for most tasks, cloud fallback for complex requests) can mitigate this.

7.3 What You Give Up

- Seamless integration with other platforms (sovereignty means boundaries)
- "It just works" simplicity (governance requires attention)
- Cutting-edge capabilities on day one (local models lag frontier)
- Vendor-managed updates (you're responsible for your infrastructure)

7.4 What You Gain

- Your rules are actually your rules
- Your data stays under your control
- Full transparency about AI behaviour
- Real exit rights
- Governance that reflects your community's values

• Capacity building for futures we can't predict

8. Honest Limitations

8.1 What Constitutional Governance Cannot Do

Guarantee perfect AI behaviour. Architecture constrains actions; it doesn't make AI wise or kind.

Replace human judgment. The framework creates checkpoints for human review; it doesn't eliminate the need for human attention.

Scale to superintelligent systems. The framework assumes AI operating within human-comprehensible parameters. It is not designed for systems that exceed human understanding.

Force community agreement. Constitutional governance requires communities to do the work of deliberation. It doesn't manufacture consensus.

8.2 What We're Still Learning

- Optimal balance between inherited rules and local customisation
- How to prevent governance fatigue while enabling genuine choice
- How community norms should evolve as AI capabilities change
- Sustainable cost models for sovereign deployment

8.3 Our Commitment

We commit to:

- Continuous improvement based on community experience
- Full transparency about how the system works
- Real exit rights that don't trap communities
- Honest acknowledgment of limitations

We don't promise to match frontier model capabilities on day one. We don't promise that governance will be effortless. We do promise that your community will have genuine authority over AI that serves it.

9. Getting Started

9.1 For Communities Considering Village

1. Assess your needs. Does your community have governance requirements that generic platforms can't meet?

2. Evaluate capacity. Do you have people willing to invest in constitutional development and maintenance?

3. Start with cloud. Begin with hosted Village tenancy; sovereign deployment can come later.

4. Develop incrementally. Start with platform defaults; customise as you learn what your community needs.

5. Engage deliberatively. Constitutional governance works when communities actually deliberate. Build that culture.

9.2 For Communities Considering Sovereign SLL

1. Build on Village experience first. Understand constitutional governance before adding infrastructure complexity.

2. Assess technical capacity. Can you deploy and maintain server infrastructure?

3. Plan for hybrid. Most communities benefit from local models for routine tasks with cloud fallback for complex requests.

4. Budget realistically. Infrastructure costs are real; they don't disappear.

5. Contribute back. Sovereign deployment generates insights valuable to the broader community.

9.3 For Researchers and Developers

The Tractatus Framework is documented for external research. We welcome:

• Independent analysis of governance patterns

• Proposals for improved validation methodology

• Contributions to open-source governance tooling

• Critique that helps us improve

10. Conclusion

AI is coming to communities whether communities prepare or not. The question is whether that AI will be governed by vendor terms of service, by constitutional frameworks reflecting community values, or by nothing at all.

The Tractatus Framework offers one answer: architectural governance that makes AI accountable to the communities it serves. Not through trust in vendor training, but through visible, auditable, democratically-determined rules.

This isn't the only answer. It involves real trade-offs. But for communities that value sovereignty—the capacity to participate in larger networks without surrendering local control—it offers something generic platforms cannot: governance that is genuinely yours.

We offer this framework in the spirit of contribution to a larger conversation about how communities can maintain agency in an age of powerful AI. The problems are hard. The answers are provisional. The conversation must continue.

"He aha te mea nui o te ao? He tangata, he tangata, he tangata."

(What is the greatest thing in the world? It is people, it is people, it is people.)

—Māori proverb

References

IBM Institute for Business Value. (2026). The enterprise in 2030. IBM Corporation.

Te Mana Raraunga. (2018). Māori Data Sovereignty Principles. Te Mana Raraunga – Māori Data Sovereignty Network.

Research Institute for Indigenous Data Sovereignty. (2019). CARE Principles for Indigenous Data Governance. Global Indigenous Data Alliance.

Bostrom, N. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press.

Hubinger, E., van Merwijk, C., Mikulik, V., Skalse, J., & Garrabrant, S. (2019). Risks from learned optimization in advanced machine learning systems. arXiv preprint arXiv:1906.01820.

Alexander, C., Ishikawa, S., & Silverstein, M. (1977). A Pattern Language: Towns, Buildings, Construction. Oxford University Press.

— End of Document —