Taonga in the Latent Space

A community-scale world model for Indigenous data sovereignty — making taonga, consent, and whakapapa first-class factors in the model an AI uses to foresee its own actions.

Précis. The most-cited researcher in artificial intelligence has left the largest AI lab in the world to bet a company on a single claim: that today's language models cannot reason or plan because they "lack a model of the world," and that the next architecture will not be built on them. He is talking about the physical world — video, robots, the bandwidth of the senses — and that is the part that needs hardware breakthroughs we should not pretend to forecast. But nothing in his definition of a world model requires the world to be made of pixels. A world model is whatever predicts the next state of some world from the current state and a proposed action, so that an agent can see the consequence before it acts. An institution is exactly such a world: discrete, bounded, already written down. This essay argues that the world worth modelling *first* is not the planet but the community; that the Village already holds two of the three pieces such a model needs — a substrate that records the community's state, and a gate that checks an action before it executes — and lacks only the middle piece, the model that predicts what an action would do to the people before it does it. That middle piece is the named gap in our own architecture: sovereignty over data that is not yet sovereignty over cognition. Building it at community scale is achievable in a way that a planet-scale model dethroning the frontier is not. And federated — each community holding its own world model, reaching others' under consent — it composes, bottom-up, into reach no single model could hold, with no centre to capture. Rome was not built in a day, and it was not built by one hand. This is how a sovereign step into world-grounded AI is actually taken: small, governed, and federated.

The bet against the trench

In November 2025 Yann LeCun announced he was leaving Meta, where he had been Chief AI Scientist for seven years and had founded its research lab, to

start a company. His own words were plain: *"I am creating a startup company to continue the Advanced Machine Intelligence research program (AMI) I have been pursuing over the last several years."* The venture, AMI Labs, is a bet — its seed round, reported at roughly $1.03 billion on a $3.5 billion valuation, was described as the largest in European history — and the thing it is a bet *against* is the consensus that has organised the entire industry. At Davos in January he put it without diplomacy: *"The AI industry is completely LLM-pilled." "Everybody is working on the same thing. They're all digging the same trench."*

What he thinks they are missing, he has said many times and most cleanly to MIT Technology Review as the company launched: large language models *"can't truly reason or plan, because they lack a model of the world. They can't predict the consequences of their actions."* And: *"We are going to have AI systems that have humanlike and human-level intelligence, but they're not going to be built on LLMs."* The illustration he reaches for is physical — *"this is why we don't have a domestic robot that is as agile as a house cat, or a truly autonomous car."*

Underneath the rhetoric is an argument about data, and it is worth stating because it is the part most often skipped. A frontier language model trains on a corpus that LeCun reckons would take a human *"170,000 years"* to read *"at eight hours a day."* A four-year-old child has taken in, through vision and the other senses, far more than that — orders of magnitude more — in a handful of years. Text, in his framing, is a thin and late projection of a world the child already understands by living in it. A system trained only on the projection learns the statistics of the shadow, not the structure of the thing casting it. It can be astonishingly fluent and still not know what happens if you let go of the glass.

This is not a fringe position dressed up. It is the considered judgement of the field's most decorated figure, staked publicly and expensively, and it converges — from the machine's side — on something this programme has been arguing from the community's side for two years: that fluency is not understanding, and that an instrument which cannot foresee the consequence of its own action on the people is not yet safe to let act.

What a world model is

LeCun's blueprint is not new; it is set out in detail in his 2022 paper *A Path Towards Autonomous Machine Intelligence*. Strip it to the load-bearing parts. An agent has a **perception** module that estimates the current state of the world; a **world model**, which the paper calls *"the most complex piece of the architecture"* and whose job is to predict *"possible future world states as a function of imagined actions sequences proposed by the actor"* — it is, in his phrase, *"a kind of 'simulator' of the relevant aspects of [the] world."* It has a **cost** module that scores how good or bad a predicted state is, composed of two sub-parts: an *"intrinsic cost, which is immutable (not trainable),"* and *"the critic, a trainable*

module that predicts future values of the intrinsic cost." And it has an **actor** that searches for the action sequence the world model and the cost agree is best, and emits the first step of it.

The crucial distinction is between two ways of acting. LeCun borrows Kahneman's labels. **Mode-1** is reactive: it *"involves no complex reasoning, and produces an action directly from the output of the perception"* — it does not consult the world model or the cost at all. **Mode-2** *"involves reasoning and planning through the world model and the cost,"* and he is exact about what that means: *"it is akin to model-predictive control (MPC),"* the classical engineering discipline of simulating candidate actions forward and choosing the one whose predicted consequences are best. Reasoning, in this frame, *"in a broad sense [means] constraint satisfaction."* You reason by imagining what would happen, and rejecting what violates your constraints, before you move.

A pure language model, on this account, is stuck in Mode-1. It produces the next token directly; it has no separate model of the world to run a consequence through and no cost to weigh the result against. That is precisely why, in LeCun's words, it *"can't predict the consequences of [its] actions."* The whole point of a world model is to make Mode-2 possible.

The other half of his program is *how* such a model should be built, and here he is deliberately heterodox. His architecture — the **Joint Embedding Predictive Architecture**, JEPA — predicts not the next *image* or the next *token* but the next *representation*. It *"is not generative."* It *"performs predictions in representation space,"* learning to ignore the details of the world that are not worth predicting, so that it can predict the things that are. He is blunt that this is a rejection of the dominant approach: *"we advocate against the use of generative architectures."* A good world model, for LeCun, is not one that can paint you the future pixel by pixel; it is one that can tell you what will matter about the future and leave the rest alone. This is not idle theory: by mid-2025 Meta had shipped V-JEPA 2, described as *"the first world model trained on video that enables state-of-the-art understanding and prediction, as well as zero-shot planning and robot control"* — a robot arm planning actions it was never trained on, by imagining their results in a learned latent space.

The world worth modelling first is not the planet

Here the argument turns, and the turn is the whole essay.

LeCun's world is the physical one. His data is video and lidar and proprioception; his proving grounds are robots and cars; his hard problem is the one I was asked to set aside — the sensory bandwidth, the hardware, the long road to a machine as capable in the physical world as a cat. If that were the only kind of world a world model could have, the Village would have nothing to say, because the Village does not build robots.

But read his own definition again with no picture attached. A world model is

the module that *predicts future states as a function of actions*. Nothing in that sentence requires the state to be pixels. It requires only that there be a state, a set of possible actions, and a dynamics that takes one state and an action to the next. LeCun himself concedes the general case: a world model is as well-formed for a discrete, symbolic environment whose state is already written down — a game, a body of code, an institution — as it is for a camera feed. In the physical case you must *learn* the state representation from raw sensors at enormous cost, which is exactly the part that needs the breakthroughs. In the institutional case the state is *already a record*. The expensive half of LeCun's problem does not arise.

This is why the community is the world worth modelling first. A community's state is bounded, discrete, and — in the Village — already serialised: who holds what, under what consent, with what cultural status, in what relationship to whom. The space of an agent's actions on it is small and enumerable. The dynamics — what a given action does to that state — is the kind of thing a community already reasons about in every hui. A community-scale world model is a tractable object. A planet-scale world model that displaces the frontier language models is not, and may not be for a decade or more on LeCun's own timeline. Rome was not built in a day. The mistake would be to wait for the cathedral before laying a single true stone. The achievable first step into world-grounded, consequence-aware AI is not a model of everything; it is a model of one community's own world, held by that community.

The Village already governs that world

The reason this is a *next step* and not a new beginning is that the Village has already built the hard parts around the gap.

Begin with the state. In the sovereign-record architecture, every piece of a community's content is a **sovereign record**: it *"carries its own provenance — who wrote it, who is its kaitiaki (steward), under what tikanga it was shared… and a cryptographic hash binding those fields together,"* its own policy, and *"its own proof chain — every governance boundary it has crossed… recorded with a cryptographic signature."* The proof chain is append-only and replayable; in the te ao Māori register it is read as *whakapapa traceability* of every record. Put the records and their proof chains together and you have something no frontier model is ever given: a faithful, signed, replayable history of a community's actual state and how it has changed. That is the ground truth a world model needs — the data against which a prediction of "what an action will do" can be checked, held in distributed possession, *"without trusting the platform operator."*

Now the gate. The Aotearoa agentic-AI framework specifies a runtime layer of checks that wrap any agent, *"independently of how the underlying agent… is built or trained"* — the framework names *"future JEPA-style architectures"* explicitly among the things it must accommodate. Two of those checks are the ones that matter here. **Boundary enforcement** *"establishes which decision types struc-*

turally require human approval," on the Wittgensteinian principle that *"what cannot be systematized must not be automated"*; its intercept *"fires before action execution… the process cannot bypass the check,"* and its output is not binary but trinary — *allow, deny,* or *escalate-to-human-deliberation.* **Metacognitive verification** *"places a verification gate before action execution"* and evaluates a proposed action against five dimensions, among them *"safety, and consideration of alternatives."* The Village, in other words, already checks an action before it executes. What it does not yet have is much to *check with.*

And it already made *one* kind of cognition sovereign. A Village runs a **Situated Language Model** — in the architecture's own words, *"a model trained on that community's own content, under its own authority,"* per-tenant by construction, on infrastructure inside the community's own jurisdiction, never exchanging model parameters with anyone. When a community withdraws its steering, the model does not fall back to a vendor's default; it falls back to silence. *"Refusal, not substitution, is the guarantee."*

But the architecture is scrupulous about the limit of what it has done, and that limit is the seam this essay opens. Our own *Sovereign-Record Architecture for Community-Scale Platforms* states it without flinching: the design *"preserves community sovereignty over data; it does not by itself preserve community sovereignty over cognition. A community using the situated-language layer to mediate member queries is still using a language model."* Sovereign data, answered by a model whose grasp of the community's *world* is borrowed. The Situated Language Model makes the community sovereign over *what it has said.* It does not make the community sovereign over *what its world is and how its world changes.* That is a world model's job.

The missing middle: a community-state world model

So name the build precisely. Between the substrate that holds the community's state and the gate that checks an action before it executes, insert the piece LeCun says intelligence cannot do without: a model that, given the community's current state and a proposed agentic action, predicts what that action would do to the state — and feeds the prediction into the gate, to be weighed before anything happens.

Map it onto his architecture, because the mapping is exact. His **perception** is the Village's encoder of sovereign records into the community's current state. His **world model** is the community-state model: it predicts the next state of *this* community under a proposed action — not its pixels, but the things that matter about it. Following the recent research literature on latent world models, that state is not an undifferentiated vector but a structured one, with distinct factors a community can name and govern: the operational state of the matter at hand, and alongside it provenance, consent, jurisdiction, cultural status, the relational whakapapa of who is connected to whom, and a standing read of collective benefit and harm. (Work like dynamics-aligned latent imagination, hi-

erarchical latent planning, object- and relation-centric world models, and latent multi-agent collaboration shows that latent state can be deliberately shaped to carry exactly such named factors rather than opaque features — the technical front is live, and the Village's contribution is not the latent state but what governs it.) His **cost** module is the Village's critic stack: the constraint critics that already exist, asking of a predicted state whether it crosses a boundary, stigmatises a group, breaches consent, routes data across a forbidden jurisdiction, or touches what a community has marked tapu. His **actor** is the agent proposing the action. And his Mode-2 loop — imagine the action's consequence, score it, gate it, and only then act — is the Village's check-before-act layer finally given something to check *against*: not a rule about the action's category, but a prediction of the action's *effect*.

This is the move from Mode-1 to Mode-2 for an institutional agent. Today the Village's agent, like a language model, is largely reactive: it proposes, and the gate asks whether the *kind* of decision is one a human must make. With a community-state world model the gate can ask the deeper question — *what would this actually do to these people?* — and refuse a technically permitted action whose predicted consequence is one the community would never accept. The agent learns to plan by simulating its effect on the community before it moves, which is the entire content of LeCun's claim about what makes an action-taker trustworthy.

Two disciplines from the existing architecture carry over without modification. First, the model is **steered, not retrained**: the Village's settled practice is that weight modification underperformed and was abandoned in favour of corpus discipline, governance packs, and steering vectors applied at inference. A community-state world model conditions on community-authored norm and policy factors; when the community changes a rule, it changes the factor, not the weights, and the change is versioned and visible. Second, the model lives **inside the substrate's constraints, by design**. The threat model already anticipates it. Adversary A7 — the *"misattribution-via-aggregating-agent"* — is exactly a future surface that *"aggregates content across tenants… or produces emergent attributions the underlying records do not warrant,"* and the architecture's standing instruction is that any such surface must *"inherit the refusal property as a design-time invariant rather than a remedial patch."* A community-state world model is an A7-class object. It is built inside tenant isolation, never across it; it predicts within one community's world, and reaches another's only through the same consent-bound, revocable federation envelope that governs every other cross-community flow.

Starting to fill it, with what the community already sees

A fair question, once the shape is clear: where does the content of this latent state come from, and must we wait for it? Part of the answer is that we already train each situated model by giving it more of the community's own material — its documents, its decisions, its rules. The same instinct runs straight past text.

A clip of the Saturday farmers' market, the footage someone took when the stop banks breached and the river came through, the recording of a high-school graduation — these are not sentiment. They are a community telling its world to itself, and they can begin filling the latent space now, with tools that already exist.

This is worth being exact about, because it is where LeCun's own work comes back to the ground. His V-JEPA models — the video ones, open and runnable today — do precisely what his architecture asks: they take video and produce *representations*, not pixels, learning the structure worth keeping and discarding the rest. A community does not need a frontier lab to do this. It needs an open encoder and its own footage. Run it over the flood video and what comes out is not a movie; it is a latent trace of how the water moved when the bank failed — the grounded, consequence-bearing knowledge LeCun says a text-only system can never have, about the one place that matters most to the people who live there.

Here the community scale is the advantage, not the limitation. LeCun's program needs something over a million hours of internet video to learn a *general* model of the physical world. A community needs nothing of the sort. It needs only its *own* world — its awa, its marae, its seasons, its streets — and that is a small corpus, already in its possession, already its taonga. The market, the storm, the graduation are enough to begin. This is the on-ramp the argument has been pointing at: you do not wait for the predictive model to be finished before you start grounding it; you lay the first true stones with the footage the community already holds.

And the footage need not be footage, nor need it be finished. A river gauge reporting its level by the minute, a tide and swell feed off the coast, rainfall and soil-moisture sensors above the stop banks, water-quality probes in the awa — these are sensory streams too, and the more faithful to LeCun's picture for being live. In his architecture perception does not read the world once and stop; it updates the model continuously as the world changes, and the model earns its keep by running the next few hours forward — what the river does if the rain holds, where the bank goes if it does not — before anyone has to act. A community can wire its own live feeds into its own model today; the sensors are ordinary, the data is the community's, and the question the model answers is the local one, not the general: not what rivers do, but what *this* awa is about to do. The Village has already built the smallest version of this — a place-based guardian aide that watches one taonga's own indicators and does nothing but notice, remember, and prompt, on the first principle that the AI assists and does not decide.

Two lines must hold while doing any of this, and neither is optional. Whatever the input — a clip, a frame, a live feed — it enters as a sovereign record, carrying consent, provenance, and kaitiaki status into the latent; where it touches wāhi tapu, a protected stretch of awa, or a person who did not agree to be there, the latent inherits that restriction rather than laundering it away. The data of a

taonga belongs to that taonga's kaitiaki and stays in Aotearoa, on infrastructure the community can reach — a live feed is not a licence to pipe a river's life into someone else's cloud. And the hard one, sharpest where the input is people: this is grounding, not surveillance. The graduation enters the model as an event in a community's life, never as a means to identify and track the faces in it. The rule that the architecture persists no biometric data of any kind is not suspended because the input is now video or a sensor stream; it is most needed exactly there. A community-state model is fed by the community's own seeing of its own world. It must never become a way of watching the people in it.

Who is kaitiaki of the world model?

This is where the Village goes beyond LeCun rather than merely applying him, and the opening was written in our own prior work. The *Plural Values* paper identified, in the frontier optimisation literature, a precise absence: *"no kaitiaki on the artifact."* The machinery that decides which version of a model survives — which improvement is promoted into service — has no seat for the community whose instrument it is. The decision is made by a score. The whole burden of that paper was that *"the benchmark is downstream of the kaitiaki,"* not above it.

A world model sharpens the question to a point: *who is kaitiaki of the model of your world?* LeCun's architecture, read carefully, all but hands us the answer. His cost module has two parts, and the division is telling. The critic is trainable — it learns. But the **intrinsic cost is, in his words, *"immutable (not trainable)"*** — it is where the system's non-negotiable drives live, the part learning is not allowed to touch. LeCun fills it with hard-wired biological analogues: pain, hunger, the basic stuff of an animal. The Village has an immutable layer too, and it is not biology — it is a **constitution**. The architecture's constitutional floor already holds, in code rather than policy, the commitments no party may erode by stealth: certain decisions always require a human; content a cultural authority marks tapu forces refusal or escalation; a data-subject's authority over its data cannot be silently revoked; Te Tiriti sits there *"as a constitutional commitment, not a courtesy line."*

So the Village's community-state world model has something LeCun's specification leaves blank: its intrinsic cost is community-authored law, and its trainable critic sits *under* that law, steered by the community's signed governance pack rather than by a global objective. The gate is the steering pack; the score is downstream of the kaitiaki. The intrinsic cost is the constitution. That is the answer to "who is kaitiaki of the world model" rendered in architecture: the community is, because the immutable part of the model's value is the community's own constitutional floor, and the mutable part answers to the community's own steering authority. The unit a community deliberates over is no longer an opaque model update; it is a *latent diff* — a legible, signed change in the model of its world, which it can inspect, contest, and revert exactly as it inspects, contests, and reverts a record today.

Federation: how it takes root

Rome was not built in a day, and it was not built by one hand. A single community-scale world model is a true stone, not a cathedral. The pathway from the one to the many is the move this programme has already named in *Federate, Don't Align*: hold your own, and reach others' through *"consent-bound, revocable channels, rather than racing toward a centre."*

Take that argument one layer down — from federating records and language inference to federating *world models*. By the time this is real, a Village will not run one situated model but several — a model of its own affairs, and, through bilateral envelopes, the use of a peer community's model where a question crosses into the peer's world. Each Village holds the world model of its own world. None holds a model of everyone's. Where a planning question reaches beyond one community — a shared waterway, a regional health pathway, a cross-iwi naming — the relevant communities federate their models the way they already federate their records: bilaterally, scope-limited, addressed to a named recipient, instantly revocable, with the provenance of every prediction intact. Capability at world-model scale is then gained the way the Village gains every other capability — by *composition*, not concentration. A mesh of community-state models, each sovereign over its own world, can reason at a scale no single one of them could reach, with no central model anyone must trust and no registry anyone can capture.

This is, in a strict sense, LeCun's hierarchy turned on its side. His H-JEPA stacks world models at coarser and coarser abstraction inside one system. A federation stacks them across communities, where the coarser, wider view is not a higher layer of one model but the composition of many sovereign ones under consent. It is the polycentric form — Ostrom's many co-equal authorities — rendered as cognition rather than only as governance. And it inherits the property that made federation the safe choice in the first place: there is no single key whose loss is everyone's loss. A community can lose a peer, a model, even a substrate provider, and recover, because it never handed anyone the model of its own world.

That is the pragmatic pathway for world-grounded AI to take root in the real world of human institutions. Not a billion-dollar planet-model that everyone must query and therefore trust, but a federated mesh of small, governed, community-held world models — buildable now, by exactly the actors the capacity race has already written off, on exactly the substrate the Village already runs.

The one word, used once

I have avoided a word that belongs in this essay, and I will use it now, once, and say exactly what I mean by it and what I do not.

LeCun calls his goal *autonomous machine intelligence*. I think the honest name

for where this leads, when an agent has a model of the world it acts in and can foresee the consequences of its actions on the people, is the first real step toward something I will call **sentient AI**. LeCun does not use that word; the responsibility for it is mine. And I mean by it nothing about consciousness, nothing about feeling, nothing about an inner life or a claim to standing. I mean only this: a machine that *has a world* — a model of the particular community it serves — and that *can be governed in it*, because it can predict what its action will do to the people and must submit that prediction to them before it moves. Sentience here is not an interior; it is an accountability. It is the difference between an instrument that acts blind and one that can be made to answer, in advance, for what its action will cost.

The Village's own Tractatus draws the line that keeps this honest, and it is worth quoting because it forbids exactly the overreach the word invites. Boundary 12.6: *"Agency cannot be simulated, only respected."* The community-state world model simulates the *world the community lives in* — places, relationships, consequences, the dynamics of the community's own state. It must never simulate the community's *agency* — the deciding, the authoring, the choosing, which the Tractatus says are not the machine's to model at all. A world model that models the world to protect the people's agency is the servant this programme has always insisted on. A world model that models the people in order to decide *for* them is the subject it has always refused. The first is the build. The second is the failure mode, and the line between them is not rhetorical; it is the constitutional floor, in code.

This keeps the machine where the corpus has always put it. In Karaitiana Taiuru's framing — *He Tangata, He Karetao, He Ātārangi* — the AI is person-like in interaction, puppet-like in operation, and in its derivation a shadow, *"cast from human language,"* not an autonomous newcomer to being. A world model does not change the metaphysics. It makes the shadow more useful and more answerable. It does not make it a person, and the moment anyone proposes to seat it as one, the discipline is being misused and the right response is to stop. Worth taking seriously. Not worth putting on a throne.

The limits, stated plainly

None of this is shipped, and the same discipline that governs the rest of the architecture governs the claim. There is no community-state world model running in the Village today. It is the next architectural step, named — not a system in production, and I will not describe it as one, on the same principle by which the project refuses to commission a situated-model cohort before its first real community exists: no aspirational building, and no aspirational writing about building. The substrate it would stand on is real and running; the federation mechanism beneath the federated-mesh vision is shipped and adversarially tested, but no live federation between independent communities yet exists; the world model itself is design, presented as design.

The honest limits go further. This closes governance sovereignty, not substrate sovereignty: a community-state world model would still run on open weights authored elsewhere and on hardware no small community built, and that dependency is exactly where the capacity race still has its grip. The hardest constraint is the one I am least entitled to adjudicate. This is written from a consciously limited, non-Māori standpoint — a sketch from outside, answerable to those whose knowledge it draws near. Where it invokes kaitiakitanga, whakapapa, tapu, it is borrowing a relational vocabulary, and the most important caution in the whole programme applies in full: to copy that vocabulary without devolving genuine authority is to reproduce the symbolic shell and keep the substance. A community-state world model is meaningful only if the community genuinely holds the constitution that is its intrinsic cost and the steering that governs its critic — only if, in Te Kāhui Raraunga's terms, Māori authority over Māori data is fully realised rather than gestured at. What counts as appropriate use of such a model in te ao Māori contexts is for tangata whenua to determine, and nothing here pre-empts that. This essay, like the ones before it, was drafted with frontier tools running on infrastructure outside the jurisdictions it argues for. That is not an embarrassment to be hidden. It is the argument, demonstrated: the dependency is real, which is exactly why building the layer a community *can* hold is worth doing now.

What this asks

A language model can answer from the world's text. A situated model can answer from a community's own text, under its own authority. But an agent that would *act* on a community needs more than an answer — it needs to foresee what its action will do to the people, and to submit that foresight to them before it moves. That is what a world model is for, and it is the piece our own architecture has named as missing. LeCun has told the whole field, at the cost of a billion-dollar company, that this piece is the difference between fluency and intelligence. The Village can say something he cannot: that the first place to build it is not the planet but the community, that the community must be kaitiaki of the model of its own world, and that federated rather than concentrated, it can take root without a centre to capture.

A model can answer. An agent can act. But only an agent that can imagine what its act will do to a community — and let the community be the one who decides — has earned the right to act at all. The world model is how it imagines. The kaitiaki is who decides. And the line that keeps the first from becoming the second is the one thing the Village has refused to compromise from the start: the machine models the world; the people remain its authors.

The Village platform and the Tractatus framework are an attempt to make governance achievable for communities at human scale — relocating authority to where it can rightfully be held, and letting communities federate that authority

rather than surrender it.

Sources. Yann LeCun, *A Path Towards Autonomous Machine Intelligence*, v0.9.2, 27 June 2022 (OpenReview `BZ5a1r-kVsf`) — all paper quotations verbatim. LeCun on founding AMI Labs, LinkedIn, 18 November 2025. "Yann LeCun's new venture is a contrarian bet against large language models," *MIT Technology Review*, 22 January 2026 ("lack a model of the world"; "not going to be built on LLMs"; house-cat/autonomous-car). "LLM-pilled" / "digging the same trench," AI House, World Economic Forum (Davos), reported in *Fortune*, 23 January 2026. Bandwidth and "house cat" framing, *Lex Fridman Podcast* #416, 7 March 2024. *V-JEPA 2*, AI at Meta, 11 June 2025 (arXiv:2506.09985). On the live research front in latent world models: Röder et al., *Dynamics-Aligned Latent Imagination* (NeurIPS 2025, arXiv:2508.20294); *Hierarchical Planning with Latent World Models* (arXiv:2604.03208); Zou et al., *Latent Collaboration in Multi-Agent Systems* (arXiv:2511.20639); Locatello et al., *Object-Centric Learning with Slot Attention* (NeurIPS 2020, arXiv:2006.15055); Zeng & Dong, *Latent World Models for Automated Driving* (arXiv:2603.09086). On the contested questions, engaged not waved away: Xing et al., *Critiques of World Models* (arXiv:2507.05169); Gurnee & Tegmark, *Language Models Represent Space and Time* (ICLR 2024, arXiv:2310.02207); Melanie Mitchell on world models in LLMs. Village sources, agenticgovernance.digital: *Sovereign-Record Architecture for Community-Scale Platforms* (v4); *A Civil-Society Proposal for Sovereign and Federated Agentic AI in Aotearoa New Zealand* (v1.2); *Held in kōrero, not collapsed to a number — plural values, living organisations, and AI*; *Federate, Don't Align*; *Kaitiaki Intelligence and the Mokopuna Recorder*. Te ao Māori grounding: Te Kāhui Raraunga, *Māori Data Governance Model* (2023); Karaitiana Taiuru, *Kaupapa Māori AI Framework — He Tangata, He Karetao, He Ātārangi* (2026). Intellectual lineage of proof-carrying records: G. C. Necula, "Proof-Carrying Code," POPL 1997.