

Current-Generation Architecture and Research Status

July 2026 · the maintained view lives at taniwha.ai/engine

This note describes the **current generation** of the Taniwha Engine as of July 2026. It is the companion to the historical first-generation note of April 2026. Where a first-generation mechanism has not yet been migrated and revalidated, this document says so plainly: first-generation results do not automatically transfer.

1 · Four authorities

The first-generation engine integrated cognition, world state, memory, and lifecycle in one system. The current generation separates four kinds of truth, each with its own authority:

Cognition

How perceptions, beliefs, rules, drives, and actions interact: dispatch, reasoning, consolidation, and belief updates. Executable rules are shared machinery kept outside any one agent's beliefs — agents believe content; they do not confuse machinery with knowledge.

The world

Authoritative world state: scene, bodies, actions, events, and perceptions. Agents submit intentions; the world resolves what actually happened and returns only what each agent could perceive. The boundary is one-way in each direction — cognition never reaches into the world's state, and the world never writes beliefs directly.

Durable symbolic belief

Each agent's committed epistemic state: beliefs with provenance, weights, valence, and validation metadata, on a native substrate built for durable persistence, indexed retrieval, and many independent minds over a shared vocabulary.

Associative representation

Fuzzy recognition beside the symbolic graph: similarity, context-conditioned prediction, surprise, and the representational foundations for portraits of other agents.

The consequence is a claim the first generation could not make structurally: **an agent does not define reality by believing it**. The world resolves what happened; the agent receives what it could perceive, and may draw the wrong conclusion. The graph is the mind — but it is not the world, and it is not the rulebook.

2 · One event, end to end

A worked path through the stack. An agent submits an intention — say, demonstrating a technique to a companion. The world resolves the action against its own state and emits events, scoped to what each body present could notice. A witnessing agent's bridging reflex translates the perception into candidate belief commits, gated by the witness's trust in the actor — testimony or instruction from a distrusted

source can be rejected outright. Committed facts join the agent's episodic chain for that scene, and related steps consolidate atomically or not at all. If the companion later applies the technique successfully, the advisor's trust rises — trust moves on observed outcomes. At every step the transition is traceable: asking the agent *why* it believes something returns the committing process, the triggering evidence, and the cycle it happened in. A cognitive transition that cannot be traced is structurally invalid — it is not accepted.

3 · Memory and consolidation

The engine does not preserve every experience equally. Co-located happenings thread into per-agent episodic chains; meaning emerges from the sequence, and chains consolidate as a unit. Consolidation runs in staged passes: contemplation does a shallow recent-window pass mid-session, triggered by prediction error, conflict, or a salient episode closing; sleep runs the full pass plus a dream replay of high-salience chains, with provenance preserved. Long-lived agents' near-identical mundane chains compress into routine schemas while anomalies stay specific — memory mirrors experience. Emotionally salient and surprising experiences are more likely to consolidate; unreinforced beliefs fade.

4 · Trust and social cognition

Trust is directional and signed: each agent holds a stance toward each other agent, from distrust to trust, carried directly in its belief state. It gates what gets in — testimony, instruction, and hearsay are admitted or rejected on the trust the agent has in the speaker — and it updates through outcomes.

Social life is recorded alongside it: bonds accrue incrementally through shared acts; slights deposit at full strength on a single breach; hearsay folds in at reduced confidence scaled by speaker trust; apology can discharge a matching slight; and cultural norms modify how the same act is read by different agents.

Theory of mind is being replatformed rather than copied: the first-generation symbolic portrait and disclosure behaviour has not been revalidated, and its representational foundations are being rebuilt around associative portraits, prediction, and surprise — watching what another agent does, and noticing when they act out of character. The richer first-generation identity-gated disclosure model is being reconsidered as it migrates.

5 · Capability matrix

Where each first-generation capability stands today:

First-generation capability	Current status
Persistent per-agent beliefs and provenance	Carried forward — rebuilt on a native belief substrate built for durable persistence and many independent minds over a shared vocabulary.
Consolidation, forgetting, analogy, abduction	Carried forward and expanded: episodic chains, routine compression, surprise-preserving memory, dream replay with provenance.
Traceable cognition	Strengthened into a runtime invariant — a cognitive transition that cannot be traced is not accepted.

First-generation capability	Current status
Layered trust-skepticism model with identity-gated disclosure	Replaced by directional trust (signed stance, updated through observed outcomes); richer social models evolving.
Identity gate and scarring (“trauma without memory”)	First-generation research; not yet current. Being reconsidered as perception is rebuilt.
Symbolic theory-of-mind portraits and disclosure pipeline	Not ported as-is; portrait foundations rebuilt around associative recognition, prediction, and surprise. End-to-end behaviour not yet revalidated.
Tactical planner and counterfactual action system	Not yet current.
Mortality and developmental phases	First-generation research.
Stasis and federated agent travel	First-generation platform capability; not a current-generation claim.
Containerised multi-tenant deployment topology	Historical first-generation deployment; not a current-generation claim.

6 · Demonstrations and evidence, by generation

Briarwatch — first-generation research line

The long-run validation environment: a 15-agent cohort over multi-week runs, producing the published psychology-anchor results — trait stability $r = 0.71$ against Roberts & DelVecchio's human band of 0.62–0.74, 47% of agents in Bonanno's resilient band after firewall events, and theory-of-mind portrait accumulation — alongside mechanism-attribution ablations. Two of four anchors do not yet validate and are tracked openly. These figures characterise the first-generation line.

Copperhollow — current generation

Current-generation minds on deterministic ground: a mining camp whose every feature can be causally explained, with agents holding individual beliefs, grudges, fears, and economic demands.

Kano and Hīkoi — the ground layer

Not an engine demonstration but proof of the ground: whole worlds grown deterministically from a seed, walkable in first person and queryable by AI — the same kind of ground the current generation's minds stand on.

Current-generation validation results are published separately as mechanisms are migrated and revalidated.

7 · Open migration and validation work

Honest edges, as of July 2026: the full perception pipeline (context weighting, motivational bias, the identity gate) is partially ported, with the gate being reconsidered rather than copied; theory-of-mind behaviour awaits revalidation on the associative foundations; the tactical planner and the mortality and developmental-stage systems remain first-generation research; the out-of-process world-server composition is demonstrated but not yet the default; and current-generation validation is published as it lands, never inherited from the first generation.

Questions about the architecture: hello@taniwha.ai · generation map: taniwha.ai/engine