



Aidan Quigley

AI systems engineer focused on knowledge-centric architectures

Builds AI systems that treat knowledge as the first-class citizen — represented faithfully, transformed without loss, transferred cleanly between models and people, composed into judgments better than any single source, then *evaluated hard enough to prove it*. Not chasing bigger models — chasing the principles by which knowledge moves between intelligences.

quigleybits@proton.me · github.com/Quigleybits · quigleybits.work · London, UK

ARCHITECTURE	MODALITIES	INTENDED USE	KNOWN BEHAVIOURS
Artist × scientist × engineer — taste for what's good, method for what's true	Python · TypeScript · Go · Postgres/pgvector · MCP · Next.js/Expo	AI systems architecture · agent platforms · research engineering (ML × systems)	Ships with decision logs · deletes what doesn't earn its place · claims only what evals prove

THE THROUGH-LINE — ONE QUESTION UNDERNEATH EVERY PROJECT

How should knowledge be represented, transformed, transferred, and composed across intelligent systems – and how do we prove it worked?



RESEARCH PROGRAMME — LOGIA · HYPOTHESES UNDER TEST, NOT CLAIMS

H1 · LORA VS PROMPT
Can a LoRA beat retrieval plus a strong prompt from the same corpus? Stable added value, or it doesn't earn its cost.

H2 · CONCEPTUAL ADAPTATION
Stable conceptual relationships — or just vocabulary and style? Novel questions, held-out sources, identical evidence.

H3 · THE INTERFACE
What representation carries learned expertise to a stronger model? The interface controls fidelity — it is not plumbing.

Compositional experts with explicit interfaces — not "which expert activates?" but "how is expertise composed once selected?" Evals precede training; a negative result is a valid finding.

CAPABILITIES — STRONGEST FIRST

- Retrieval & matching systems

CORE

FAISS + BM25 + web + graph into judged pools; embeddings proven by A/B.
- Evaluation engineering

CORE

McNemar gates, frozen baselines, blind judging — evals as gates, not demos.
- Agent orchestration

CORE

Planner → implementer → verifier → reviewer → critic; constitutional loops.

Protocol & tooling design

MCP servers in TypeScript; a published CLI conformance standard; npm skills.

Full-stack shipping, secure and fast

Next.js · Expo · Supabase · Firebase; RLS + bcrypt; ~160→10 ms round-trips.

EVALUATIONS — EVIDENCE FROM LIVE SYSTEMS

- Hymn_core — five signals, one judgment

hymncore.net

1,350 hymns over a 55,000-paragraph corpus, judged source-blind. 234 fixes live, 0 hallucinated refs; recall ~5× (p = 0.012).
- 2nd_brain — memory with source fidelity

2ndbrain.website

Personal RAG, daily production: pgvector + Voyage, MCP server, Telegram capture — free tiers.
- MCLIP — one contract across boundaries

mclip.dev

A CLI conformance profile for MCP: normative spec, security model, 30+ fixtures, Go reference.
- Shipped fleet — ideas taken to live

12+ live products — Assimilax (iOS App Store), scosig.com arXiv pipeline, kanban.website, cctts, claude-skills on npm.

JUDGMENT CALLS — DECISIONS AND CATCHES THE EVIDENCE BACKED

- The judge selects; scores only gather.

Retrieval supplied 81% of Hymn_core's candidates but only 0.8% of final picks — so the LLM judges the whole pool source-blind, never the top score.
- Don't take the agent's word for it.

Eight logged catches where one human challenge beat an AI agent's confident answer — latest: an unvalidated "LoRA hosting dominates cost" claim, reversed by a price check. Narration is not evidence.
- Gate against Goodhart.

A regression gate cut a blind panel's ~50% improvable" to ~23% real; two blind judges must agree before a batch ships — over 20% bad-rate pauses it.

TRAINING DATA — WHERE THE JUDGMENT COMES FROM

PRE-TRAINING

BSc Biology — built a noise-tolerant plant-ID engine with error-aware trait matching; retrieval and eval two decades before LLMs. BA Fine Art — information design bridging art and science.

ALIGNMENT & PEOPLE

A decade in the charity sector alongside all of the above — mentoring successive trainee cohorts, presenting to hundreds; two years' residential character training. Responsibility for people, not just systems.