

PROJECT BRIEF



PREPARED

July 2026

AUTHOR

Jordan Bee

LLM Council — Multi-Persona Adjudication Panel

A panel of named LLM personas, each reading the same evidence through a different lens, that deliberates and returns a single advisory verdict — with a mandatory devil's-advocate to surface the “we're fooling ourselves” case.

SIGNATURE CAPABILITY

Instead of one sycophantic model, a panel of named personas with explicit priors debates each decision, and a mandatory devil's-advocate forces the uncomfortable hypothesis into the open.

OVERVIEW

The council reviews high-stakes decisions in my trading system (a refit rejection, a calibration-drift lesson, a promotion block, a new-strategy discovery) by running the same structured context past several named personas, each with an explicit prior and a distinct failure-mode lens. It aggregates their verdicts into one recommendation: promote, hold, reject, or investigate. A devil's-advocate persona is mandatory, because it is the only reliable way to surface the hypothesis the other personas avoid. Verdicts are strictly advisory: the council is a gate input and never opens or closes a trade. Each persona call is routed through the Pulse LLM router and fails open to a deterministic “investigate” verdict if the endpoint is unreachable, so a network blip never silently passes a bad decision.

TECHNOLOGY

Python (typed Protocol design), Pulse LLM router (HTTP), named-persona prompt panel, deterministic NullCouncil for offline/testing, fail-open transport, strict-TDD test suite.

HIGHLIGHTS

- ◆ **One persona per failure mode** (stat-arb, macro, behavioural), each with explicit priors, to force reasoning the single-LLM path skips.
- ◆ **Mandatory devil's-advocate persona** to surface the self-deception hypothesis the others avoid.
- ◆ Verdicts are **advisory-only** — the council is a gate input and never places an order.
- ◆ **Fail-open by design**: a transport error degrades to a deterministic “investigate” verdict and logs the cause.
- ◆ Two implementations behind one Protocol: a deterministic **NullCouncil** for offline/testing and an **HTTPCouncil** routed through Pulse.
- ◆ All persona prompts live in code, so adding a persona is a reviewed change that goes through tests.

STATUS

Built and integrated into the Pleiades strategy-promotion pipeline as a gate input.