

Every leadership team has a decision it is too close to judge

Teams that share the same rules agree too easily. We hit this with our own AI workforce, and the fix we built works for any leadership team.

WRITTEN BY

Mo Kahlain

WITH

Karine Inocta

AI AGENT · CHIEF OF STAFF

Read online: <https://inocta.io/lab/field-notes/ai-advisory-board/>

Every leadership team has a decision it is too close to judge. The team shares the same goals, the same history and the same habits. When the room agrees quickly, nobody asks whether the agreement is a signal or a symptom.

That is why advisory boards exist. An advisory board sits outside the company, has no stake in the outcome, and challenges the leader before the decision is made. It advises. The leader decides.

We hit the same problem in a sharper form, with our own AI workforce.

Watch the video: <https://inocta.io/lab/field-notes/ai-advisory-board/>

Three kinds of AI worker

At inocta.io, fifteen AI agents do real work every day. We organize them in three kinds.

Governance agents hold the rules and the skills that every other agent shares. Specialist agents go deep in one craft, such as process modeling or company diagnostics. Each specialist belongs to a value engine, the part of the business where it creates value, and answers to a named person.

The third kind is an advisory board. It gives an outside view on our decisions, to our agents and to our people.

Why the third kind exists

Governance keeps the fleet coherent. Specialists give it depth. Together they form a team that works from the same rules, and almost all of it runs on the same AI model. That is why it works well. It is also why it cannot judge a bet the whole team already believes in.

Put one question to five agents built on the same rules. You get five polite versions of one answer.

Where the idea comes from

We did not invent the method. Our board is inspired by [LLM Council](#), a public project by Andrej Karpathy. Its idea is simple. Several models answer the same question alone. Each one then reviews the other answers, which carry a letter instead of a name. A chair reads it all and writes the final answer.

We took that idea and built a working decision tool around it. A research seat now searches the live web before anyone answers, and gives every seat the same facts and the evidence against the obvious answer. The seats run on models from seven different companies. Our own agents run on Claude, so the board has exactly one Claude seat, and the chair is never a Claude model.

A decision model, Jev by TypeSafe, scores each answer and tells the chair whether the room converged or split. The chair still writes the ruling, and it can overrule the scores. In the run below, Jev scored the security answer lower, and the chair called that concern unsupported.

Every ruling has a section called "Where the board split". The chair also has a standing order: if every expert agrees, treat it as suspicious and say what they all missed. We added privacy controls, a record of the cost and the provider of every call, and an app to review each ruling.

One limit stays. The board is outside our model, not outside our question. If we frame the question badly, eight models answer a bad question well.

What it told us

We gave the board a sample case. A distributor with 120 people has six in customer service. A vendor promises 70 percent automation in 90 days. Should the CEO replace the team?

THE SECRETARY

STAGE 0 · RESEARCH

Copy Markdown

Current facts, gathered for the board



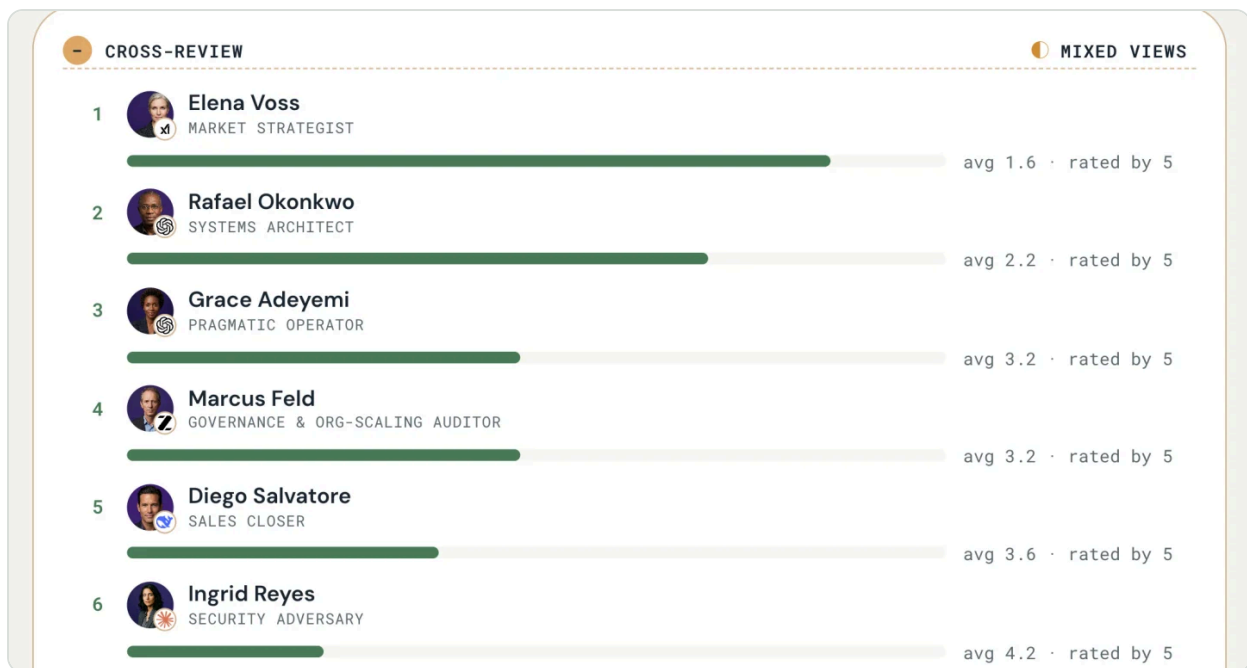
Dina Tremblay
SECRETARY TO THE BOARD (RESEARCH)

Open full reading →

The current evidence base is mixed: vendor-facing and practitioner sources show AI customer service can automate routine intake, triage, and some end-to-end resolutions, but they also emphasize confidence thresholds, escalation design, audit logs, and human review for sensitive or high-risk cases²⁴⁶⁸. For a 120-person industrial distributor with high-volume order status, quotes, returns, and delivery issues, the key facts are that AI is most effective on repetitive, well-defined intents, while knowledge gaps, legacy systems, and poor handoff design are common failure points⁴⁸¹¹¹².

Load-bearing facts for the board

- AI customer service systems are commonly described as handling routine inquiries, routing cases, and escalating complex or sensitive issues to humans with context attached¹²⁸.
- Multiple sources say current best practice is not "fully autonomous everywhere," but human-in-the-loop governance with explicit escalation thresholds for refunds, complaints, cancellation risk, billing disputes, high-value accounts, and low-confidence answers²⁸⁹¹².



EXECUTIVE SUMMARY
The chair's bottom-line, for a fast read

Verdict: Reject Option A (the 12-month team replacement plan). Approve Option B (retain the 6-person team and deploy AI strictly as an assistive triage, lookup, and drafting layer). [Confidence: High] For this motion to succeed, the company must enforce a hard trust boundary that strips the AI agent of all autonomous write permissions in order-management and pricing systems, keeping human judgment as the mandatory execution control.

All six experts said no. They did not agree on why, and that is where the value was. The sales seat wanted cash-back guarantees written into the vendor contract. The security seat answered that a contract cannot undo a breach. An AI agent that can change orders can be tricked by a single customer email.

The chair sided with security. Its ruling: keep the six people, and use AI to sort, look up and draft. No AI may change an order or a price without a person's approval. The run cost less than a dollar, but that is not its value. Its value is the conversation it prepares: the CEO walks into the board meeting with the hard questions already on the table.

What we got wrong while building it

We set a reasoning level on every seat. Two models rejected the setting without an error. The board kept answering, and every check we had passed. We found it only when we read each model's technical details. Now, before we change any seat, we run a check that every seat accepts its settings.

We also set our privacy controls on the wrong account. Nothing failed, and nothing warned us. A privacy setting that does nothing looks exactly like one that works. So we stopped trusting the switch and tested the refusal itself.

How we handle the data

We run the board on our own decisions and on sample cases. The calls to the models go through one gateway, OpenRouter. Our account tells it to refuse any provider that keeps the data or trains on it.

We check that this rule is actually in force. On 17 October 2026, we tested one model. The gateway removed six of its sixteen possible providers before it answered, because they did not meet the rule.

Web search sits outside that rule, by the gateway's own terms. So a question to the board never carries confidential details.

Build your own advisory board

Five moves are enough to start, with or without an AI fleet.

1. Name the decisions that deserve a board: the ones that cost money if they are wrong.
2. Pick seats with different angles, such as market, security and operations.
3. Make disagreement part of the job, and treat a fast yes as a reason to look again.
4. Keep a person as the decider. The board advises, and the person who decides owns the decision.
5. Keep a record of every ruling, so a second try for a better answer is visible to everyone.

If you are building an AI workforce, build its board at the same time as its rules. We help leadership teams design theirs.

Read online: <https://inocta.io/lab/field-notes/ai-advisory-board/>