Tautai

Judgment Zones

How a unit finds its position with AI, and what follows from it

Version 0.8 · 17 November 2026

Krishan Mathis · tautai · krishan@tautai.net

What this paper is for

This paper explains a method, not a product. It describes how Judgment Zones establishes where an organisational unit stands with AI, what assumptions sit behind that, what is measured, and what the method deliberately does not deliver. It is not a sales document: the method's limits appear in it, in the same place as its claims.

The occasion is an observation that repeats in almost every conversation about AI. The organisation now has tools, pilots, a strategy and a policy. What it often does not have is an answer to where it actually stands — and whether it knows that or is only guessing. Without that answer, every decision about AI becomes a decision about tools, and nobody can say afterwards whether it worked.

The paper reads in one pass. It starts with the problem the method addresses, then sets out the picture of two axes and four fields, the process and the measurement, and ends with the question of what this means for a unit that has no target picture yet.

In five sentences. AI makes output cheap and judgment expensive. Maturity ladders measure the wrong thing and deliver one grade for the whole company. Judgment Zones instead reads two axes per unit — how deeply AI is embedded, and how well the unit can check and direct its results — and draws a location in one of four fields from them. From that location follow six steps, of which locating is the second, and a plan the unit owns itself. Every structural claim carries the observation that would prove it wrong. If you have more time, read on; if you have one page, read section 2.

1. The problem: cheap output, expensive judgment

AI makes output cheap. The draft, the quote, the forecast, the classification, the answer — everything that used to cost time now arrives in seconds. What stays expensive is something else: the capacity to recognise when the result is wrong, to override it, and to point it somewhere useful. That capacity decides whether AI strengthens a unit or hollows it out. And it is the capacity that practically no established maturity model measures as a dimension of its own.

Most of these models share a shape inherited from CMM: one axis, five levels, one company, one level — and the quiet assumption that the top level is everyone's goal. That shape produces five failures, and Judgment Zones is built against them.

1. **One level for the whole company.** Units sit in different places at any given moment. An average hides exactly what leadership needs to see.
2. **The level as the goal.** A ladder cannot say: "you are right where you are." Some units should stay in Exploration; some processes should never be obliterated.
3. **Volume instead of judgment.** What is measured is how much AI is in use. What is not measured is whether anyone can still tell when the result is wrong.
4. **Access counted as capability.** Rollout is counted as progress. Access without judgment is exposure, not capability.
5. **Untestable.** No observation is named that would show the model to be wrong.

Why not another ladder? Because ladders need three assumptions that do not hold for AI adoption: a known target, sequential progress, and "more is better". AI adoption has no known end state, does not proceed in phases, and is decided by decision capacity, not by tool coverage. Where conditions are stable relative to the response cycle — payroll, compliance — planning is right. AI adoption in a unit is not one of those cases.

2. The idea: a location, not a grade

Judgment Zones draws two axes.

Horizontal: AI entrenchment. How deeply is AI woven into a unit's routines, its talent systems and its steering? Four stages, from Exploration through Codification and Propagation to Obliteration — the stage names are working terms of the model, and a German version is not yet decided. The stage is read from present and absent signals in four areas: tooling, enablement, steering, structure. A unit can show signals of two adjacent stages; its position is then a range, not a point.

Vertical: decision capacity. Can the unit filter, check, override and direct the results of AI? Four dimensions:

- **Decision latitude** — may people override, and do they have the time to look?
- **Evaluation skill** — can they tell a plausible wrong answer from a right one?
- **Judgment-broker roles** — is there a named person with override authority and review duty, and is time allocated for it?

- **Experience pipeline** — where do the next people who can do this come from when AI absorbs the novice work?

The four bars together form the axis. If one is missing, the axis is not half full; it is incomplete.

That produces four fields:

	Low entrenchment	High entrenchment
High decision capacity	**Blocked** — people who could check and direct AI, in a unit that has not let them. Skill without access.	**Augmented** — AI takes the routine, people keep the exceptions, the overrides and the direction. Not an end state.
Low decision capacity	**Dormant** — little AI, little capacity. "Wait, and build judgment first" can be the right answer.	**Hollow** — AI is entrenched and nobody can tell when it is wrong. Access without judgment. Looks like success on the tool dashboard.

Three rules belong to the picture:

- **A location, not a grade.** Units of the same company sit in different fields at the same time, and should. There is no target in the top right.
- **A boundary is a location.** A unit on a line is drawn on the line. Rounding it into the nearer field is the ladder habit coming back through the graphic.
- **Blocked and Dormant need opposite advice.** From outside, both look like "little AI". In Blocked the constraint is access and latitude; in Dormant it is decision capacity, which can be built without AI.

The one question for a leadership team: **which of your units is in which field — and do you know, or are you guessing?**

3. The process: locating is the second step

Judgment Zones is not an assessment that ends with a result. It is the second of six steps, and the other five are what a unit does with the location.

Step	Question	What happens
1 Orient	Why navigate rather than plan?	The usual way of introducing AI is named: roadmaps for a technology that changes quarterly, copied case studies, more tools read as more capability.
2 Locate	Where does each unit stand?	Judgment Zones: entrenchment × decision capacity, per unit and per core workflow, through one of four ways in. A location, not a grade.
3 Make sense	What does the location mean for this unit — and can it act?	The unit reads its own profile in a perspective session: what is happening, what it means, what comes next. Judgment is read as one element of the ability to act.
4 Set course	What is fixed, what is cheap to move, which way is downhill?	The elements of the profile are placed on a map (energy × time). Direction as "more of / less of". A one-page plan: at most three probes with stop conditions, and a date.
5 Move	Are the moves working?	The unit runs its probes. Monthly: continue, stop, scale — against the conditions written down beforehand.
6 Re-measure	Did it work, on our data?	Six months later: the same items, the same measures, the delta. "Nothing moved" and "we were right to wait" are admissible answers.

Two things follow. The steps after the location are addressed to the unit — a department, a division, a team with its own customers — not only to leadership. And what follows a location is qualification first: people who can read the profile and act on it. Consulting is a need that may emerge from the plan, later, and only in some units.

4. What is measured

Seven dimension groups on one shared item set. Every item has a reading in the Tautai model and one in the Viable System Model; the quick tier reports the first, the depth tier both.

Group	Question	How
A Adaptive potentials	Does AI amplify or erode the capacity to Perceive, Respond, Monitor, Learn, Anticipate?	Management items, bipolar scale; absolute variant for units without AI.
B Viability functions	Has AI strengthened, bypassed or hollowed out the system's functions?	Interview readings with evidence, by a VSM-literate consultant. Read, not measured.

Group	Question	How
C Information culture (gate)	Can bad news about AI output travel upward?	Six Westrum/DORA items and six AI-specific twins, anonymous member survey. A condition, not a score.
D Judgment and agency	Can people tell when output is wrong, may they override it, who holds the role, where do the next experts come from?	Latitude, evaluation skill, broker roles, pipeline. The four bars of the vertical axis.
E Delegation coherence	Does delegation depth exceed the judgment available, per decision class?	Card per decision class: depth per automation stage, exception rate, last override, judgment available.
F Entrenchment location	Where does the unit stand, and each core workflow?	Sixteen present/absent signals (four in the short form). A position, never scored.
G Governance	Do the practices that make delegation safe exist from the first pilot?	Eight present/absent practices.

Deliberately not measured: data readiness, technology stack, number of use cases, budget share. They are recorded as context so the research can control for them; they describe the tooling, not the position. No composite, no weighting, no company level.

Every element of a report carries one of three evidence marks: **measured** (a value on a stated scale from a stated instrument, with n), **read** (a categorical judgment by a named reader from named evidence), or **asserted** (a statement without instrument or evidence). An element moves up only when evidence or an instrument is attached; it never slips down silently.

Individual answers are reported only in aggregate, and only from five respondents per unit. No item asks about individual performance, and no result can be used in a performance conversation.

5. Four ways in

One shared item set across all four, so that a unit measured before can be measured again afterwards. The effort figures are estimates.

Way in	For	Effort and client time	What the client holds afterwards
Check	A leadership team that wants to see its units on the picture	20 minutes (30 for five units), no preparation	The map of three to five units, each position marked evidence or guess, and the unit to measure first
Profile	Units with AI in use, without a consultant	28 management items (10–14 min, in discussion), 18 anonymous member items (5–7 min, n ≥ 5), three workflow cards; report within one working day	Per unit: potentials radar, four decision-capacity bars with contradictions, entrenchment location, culture gate, governance strip, up to three probes with stop conditions
Potential	Units that do not yet use AI	3 consultant days per unit (up to 4.5) plus 1.5 days of engagement overhead	Pre-AI baseline, abstention reading (deliberate / deferred / blocked / unexamined), delegation ceiling per decision class, governance floor, entry probes. “Do not start in the next twelve months” is an admissible result
Diagnostic	Units with AI in a core workflow	3–5 consultant days per unit plus 1.5 days of engagement overhead	Viability reading with evidence, pathology list, delegation-coherence map, governance record for an ISO/IEC 42001 implementation, baseline of the outcome measures, probes, re-measurement date

The routing between them: start with the Check. A unit on the right of the map (Hollow, Augmented) continues with the Profile and, where needed, the Diagnostic. A unit on the left (Blocked, Dormant) continues with the Potential.

6. What the method does not promise

No way in delivers: a score, a benchmark against other companies, a certificate, a roadmap to a level, a use-case catalogue, an ROI figure — or the assurance that the outcome measures will move. The re-measurement exists precisely for that, and it can also show that nothing happened.

One result is explicitly admissible: **that a unit should not start in the next twelve months** — with the reason, and with the observation that would change our mind. A method that cannot produce that result is a sales instrument.

7. How you know whether it worked

Seven outcome measures per unit, at baseline and every six months, each from a named record or reported as “not observable”: decision latency (routine and non-routine), ratio of probes to scale-ups, governance violations

and near-misses, output per full-time equivalent in a judgment role, experience pipeline, override rate and override quality. Latency is never reported alone: a faster decision bought by skipping the check is the Hollow field in the making.

And the model can lose. Every structural claim carries with it the observation that would refute it:

- **F1:** Hollow units degrade — rising non-routine latency, rising violations, a weakening pipeline, relative to Augmented units. Refuted if there is no difference across two rounds.
- **F8:** Blocked units that are given access and latitude reach Augmented faster and with fewer incidents than Dormant units with the same tooling. Refuted if there is no difference.
- **F10:** Decision capacity measured before adoption predicts the field reached afterwards. Refuted if starting capacity has nothing to do with the field reached.

That is the reason the method fits an organisation that has already tried a great deal: it brings its own doubts with it.

8. What this means for a unit without a target picture

A target picture cannot sensibly be written before the location. If you do not know where a unit stands, you cannot say where it should go — and a target picture that is written anyway usually describes what was planned in the first place.

The method therefore reverses the order. What comes first is not the document but a session: a leadership team places three to five units on the picture of two axes and marks each position as evidence or guess. It takes twenty minutes and needs no preparation. Two things come out of it: a map that shows how much is actually known, and a named unit to measure first.

That is also the point at which the missing target picture becomes visible — not through a talk about what is missing, but through the picture that emerges in those twenty minutes. A gap nobody sees is not worked on; a gap that is on a map has an address.

What follows is a measurement the unit can read for itself, a plan with at most three probes, and a date for re-measurement. The target-picture draft emerges from that reading, not before it.

9. Lineage

The method has extensive foundations. It takes its structure from five patterns and names what it leaves out: the entrenchment stages and the claim that judgment is the scarce resource from the Tautai model, the argument that “access without agency is exposure”, the Viable System Model of Stafford Beer with the pathology families of Pérez Ríos, the picture of flows and movable elements instead of milestones from Dave Snowden, and the distinction between capability, room to act and will from the agency literature. The critique of maturity models rests on the systematic reviews by Sadiq et al. (2021) and Smajli, Feldman & Cox (2024).