Tautai

Judgment Zones

Wie eine Einheit ihren Stand beim Einsatz von KI bestimmt — und was daraus folgt

Version 0.8 · 17. November 2026

Krishan Mathis · tautai · krishan@tautai.net

Wofür dieses Papier da ist

Dieses Papier erklärt eine Methode, kein Produkt. Es beschreibt, wie Judgment Zones den Stand einer Organisationseinheit beim Einsatz von KI bestimmt, welche Annahmen dahinterstehen, was gemessen wird und was das Verfahren bewusst nicht liefert. Es ist kein Verkaufsdokument: Die Grenzen der Methode stehen darin, an derselben Stelle wie ihre Behauptungen.

Der Anlass ist eine Beobachtung, die sich in fast jedem Gespräch über KI wiederholt. Die Organisation hat inzwischen Werkzeuge, Pilotprojekte, eine Strategie und eine Richtlinie. Was sie oft nicht hat, ist eine Antwort auf die Frage, wo sie eigentlich steht — und ob sie das weiß oder nur schätzt. Ohne diese Antwort wird jede Entscheidung über KI zu einer Entscheidung über Werkzeuge, und niemand kann hinterher sagen, ob sie gewirkt hat.

Das Papier lässt sich in einem Zug lesen. Es beginnt mit dem Problem, das die Methode adressiert, erklärt dann das Bild aus zwei Achsen und vier Feldern, den Prozess und die Messung, und endet mit der Frage, was das für eine Einheit bedeutet, die noch kein Zielbild hat.

In fünf Sätzen. KI macht Output billig und Urteil teuer. Reifegradleitern messen das Falsche und liefern eine Note fürs Ganze. Judgment Zones liest stattdessen je Einheit zwei Achsen — wie tief KI eingewachsen ist und wie gut die Einheit ihre Ergebnisse prüfen und steuern kann — und zeichnet daraus einen Ort in einem von vier Feldern. Aus dem Ort folgen sechs Schritte, von denen das Verorten der zweite ist, und ein Plan, den die Einheit selbst verantwortet. Jede strukturelle Behauptung trägt die Beobachtung mit sich, die sie widerlegen würde. Wer mehr wissen will, liest weiter; wer eine Seite hat, liest Abschnitt 2.

1. Das Problem: billiger Output, teures Urteil

KI macht Output billig. Der Entwurf, das Angebot, die Prognose, die Klassifikation, die Antwort — alles, was früher Zeit gekostet hat, kommt heute in Sekunden. Teuer bleibt etwas anderes: die Fähigkeit zu erkennen, wann das Ergebnis falsch ist, es zu überschreiben und es in eine Richtung zu lenken. Diese Fähigkeit entscheidet darüber, ob der Einsatz von KI eine Einheit stärkt oder sie aushöhlt. Und sie ist die Fähigkeit, die praktisch kein verbreitetes Reifegradmodell als eigene Dimension misst.

Die meisten dieser Modelle teilen eine Form, die aus dem CMM stammt: eine Achse, fünf Stufen, ein Unternehmen, eine Stufe — und die stille Annahme, dass die oberste Stufe das Ziel aller sei. Diese Form produziert fünf Fehler, gegen die Judgment Zones gebaut ist.

1. **Eine Stufe für das ganze Unternehmen.** Einheiten stehen zu jedem Zeitpunkt an verschiedenen Stellen. Ein Durchschnitt verdeckt genau das, was eine Führung sehen müsste.
2. **Die Stufe als Ziel.** Eine Leiter kann nicht sagen: „Du bist dort richtig, wo du stehst.“ Manche Einheiten sollten in der Exploration bleiben; manche Prozesse sollten nie ausgelöscht werden.
3. **Volumen statt Urteil.** Gemessen wird, wie viel KI im Einsatz ist. Nicht gemessen wird, ob jemand noch erkennen kann, wann das Ergebnis falsch ist.
4. **Zugang als Fähigkeit.** Der Rollout wird als Fortschritt gezählt. Zugang ohne Urteil ist aber Exposition, keine Fähigkeit.
5. **Nicht widerlegbar.** Es wird keine Beobachtung genannt, die das Modell falsch machen würde.

Warum keine weitere Leiter? Weil Leitern drei Annahmen brauchen, die auf die Einführung von KI nicht zutreffen: ein bekanntes Ziel, sequenziellen Fortschritt und „mehr ist besser“. KI-Einführung hat kein bekanntes Endziel, verläuft nicht in Phasen, und entschieden wird sie über Entscheidungsfähigkeit, nicht über Werkzeugabdeckung. Wo Bedingungen stabil sind relativ zum Reaktionszyklus — Lohnabrechnung, Compliance —, ist Planung richtig. Die Einführung von KI in einer Einheit ist nicht einer dieser Fälle.

2. Die Grundidee: Ort statt Note

Judgment Zones zeichnet zwei Achsen.

Waagerecht: die KI-Durchdringung. Wie tief ist KI in die Routinen, die Talentprozesse und die Steuerung einer Einheit eingewachsen? Vier Stufen: Exploration, Codification, Propagation, Obliteration. Gelesen wird die Stufe an vorhandenen und fehlenden Signalen in vier Bereichen: Werkzeuge, Befähigung, Steuerung, Struktur. Eine Einheit kann Signale zweier benachbarter Stufen zeigen; dann ist ihre Lage eine Spanne, kein Punkt.

Senkrecht: die Entscheidungsfähigkeit. Kann die Einheit die Ergebnisse von KI filtern, prüfen, überschreiben und lenken? Vier Dimensionen:

- **Entscheidungsspielraum** — dürfen Menschen überschreiben, und haben sie die Zeit, nachzusehen?
- **Prüfkompetenz** — können sie eine plausible falsche Antwort von einer richtigen unterscheiden?

- **Urteils-Broker-Rollen** — gibt es eine benannte Person mit Überschreibungsbefugnis und Prüfpflicht, und ist dafür Zeit eingetragen?
- **Erfahrungspipeline** — woher kommen die nächsten Leute, die das können, wenn KI die Anfängerarbeit übernimmt?

Die vier Balken zusammen bilden die Achse. Fällt einer aus, ist die Achse nicht halb voll, sondern unvollständig.

Daraus entstehen vier Felder:

	Geringe Durchdringung	Hohe Durchdringung
Hohe Entscheidungsfähigkeit	**Blocked** — Menschen, die prüfen und steuern könnten, in einer Einheit, die sie nicht gelassen hat. Können ohne Zugang.	**Augmented** — KI übernimmt das Routinehafte, Menschen behalten die Ausnahmen, die Überschreibungen und die Richtung. Kein Endzustand.
Geringe Entscheidungsfähigkeit	**Dormant** — wenig KI, wenig Fähigkeit. „Warten und zuerst Urteil aufbauen" kann die richtige Antwort sein.	**Hollow** — KI ist durchgedrungen, und niemand kann erkennen, wann sie falsch liegt. Zugang ohne Urteil. Sieht auf dem Werkzeug-Dashboard wie Erfolg aus.

Drei Regeln gehören zum Bild:

- **Ort, nicht Note.** Einheiten desselben Unternehmens stehen gleichzeitig in verschiedenen Feldern, und das ist richtig so. Es gibt kein Ziel rechts oben.
- **Eine Grenze ist ein Ort.** Eine Einheit, die auf einer Linie liegt, wird auf der Linie gezeichnet. Sie in das nähere Feld zu runden, ist die Leitergewohnheit, die durch die Grafik zurückkommt.
- **Blocked und Dormant brauchen entgegengesetzten Rat.** Von außen sehen beide nach „wenig KI" aus. In Blocked ist die Schranke der Zugang und der Spielraum; in Dormant ist es die Entscheidungsfähigkeit, die man auch ohne KI aufbauen kann.

Die eine Frage für eine Führung lautet: **Welche Ihrer Einheiten ist in welchem Feld — und wissen Sie es, oder schätzen Sie es?**

3. Der Prozess: Verorten ist der zweite Schritt

Judgment Zones ist kein Gutachten, das mit einem Ergebnis endet. Es ist der zweite von sechs Schritten, und die anderen fünf sind das, was eine Einheit mit dem Ort macht.

Schritt	Frage	Was passiert
1 Orientieren	Warum navigieren statt planen?	Die übliche Art, KI einzuführen, wird benannt: Roadmaps für eine Technologie, die sich vierteljährlich ändert, kopierte Fallstudien, mehr Werkzeuge als mehr Fähigkeit.
2 Verorten	Wo steht jede Einheit?	Judgment Zones: Durchdringung × Entscheidungsfähigkeit, je Einheit und je Kernablauf, über einen von vier Zugängen. Ein Ort, keine Note.
3 Verstehen	Was bedeutet der Ort für diese Einheit — und kann sie handeln?	Die Einheit liest ihr eigenes Profil in einer Perspektivsitzung: was los ist, was es bedeutet, was als Nächstes kommt. Urteil wird als ein Element von Handlungsfähigkeit gelesen.
4 Kurs setzen	Was ist fixiert, was billig zu bewegen, wo geht es bergab?	Die Elemente des Profils werden auf eine Karte gelegt (Energie × Zeit). Richtung als „mehr von / weniger von". Ein Ein-Seiten-Plan: höchstens drei Versuche mit Abbruchbedingungen, ein Datum.
5 Bewegen	Wirken die Bewegungen?	Die Einheit führt ihre Versuche. Monatlich: weiter, stoppen, skalieren — gegen die Bedingungen, die vorher aufgeschrieben wurden.
6 Nachmessen	Hat es gewirkt, an unseren Daten?	Sechs Monate später: dieselben Items, dieselben Maße, das Delta. „Nichts hat sich bewegt" und „wir hätten zu Recht gewartet" sind zulässige Antworten.

Zwei Dinge folgen daraus. Die Schritte nach dem Verorten sind an die Einheit adressiert — eine Abteilung, ein Bereich, ein Team mit eigenen Kunden —, nicht nur an die Führung. Und was auf einen Ort folgt, ist zuerst Qualifizierung: Leute, die das Profil lesen und danach handeln können. Beratung ist ein Bedarf, der aus dem Plan entstehen kann, später, und nur in manchen Einheiten.

4. Was gemessen wird

Sieben Dimensionsgruppen auf einem gemeinsamen Item-Satz. Jedes Item hat eine Lesart im Tautai-Modell und eine im Viable System Model; die Kurzstufe berichtet die erste, die Tiefenstufe beide.

Gruppe	Frage	Wie
A Adaptive Potenziale	Verstärkt oder erodiert KI die Fähigkeit zu Wahrnehmen, Reagieren, Beobachten, Lernen, Vorwegnehmen?	Management-Items, bipolare Skala; absolute Variante für Einheiten ohne KI.
B Viabilitätsfunktionen	Hat KI die Funktionen des Systems gestärkt, umgangen oder ausgehöhlt?	Interview-Lesarten mit Beleg, durch einen VSM-kundigen Berater. Gelesen, nicht gemessen.
C Informationskultur (Gate)	Kann schlechte Nachricht über KI-Output nach oben wandern?	Sechs Westrum-/DORA-Items und sechs KI-spezifische Zwillinge, anonyme Mitgliederbefragung. Ein Zustand, kein Score.
D Urteil und Handlungsfähigkeit	Können Leute erkennen, wann Output falsch ist, dürfen sie überschreiben, wer hält die Rolle, woher kommen die nächsten Experten?	Spielraum, Prüfkompetenz, Broker-Rollen, Pipeline. Die vier Balken der senkrechten Achse.
E Delegationskohärenz	Übersteigt die Delegationstiefe je Entscheidungsklasse das vorhandene Urteil?	Karte je Entscheidungsklasse: Tiefe je Automatisierungsstufe, Ausnahmenrate, letzte Überschreibung, vorhandenes Urteil.
F Durchdringungsort	Wo steht die Einheit und jeder Kernablauf?	Sechzehn Anwesend-/Abwesend-Signale (vier in der Kurzform). Position, nie bewertet.
G Governance	Gibt es die Praktiken, die Delegation sicher machen, ab dem ersten Piloten?	Acht Anwesend-/Abwesend-Praktiken.

Bewusst nicht gemessen: Datenreife, Technologie-Stack, Anzahl der Use Cases, Budgetanteil. Sie werden als Kontext erfasst, damit die Forschung sie kontrollieren kann; sie beschreiben das Werkzeug, nicht die Position. Kein Gesamtwert, keine Gewichtung, keine Unternehmensstufe.

Jedes Element eines Berichts trägt eine von drei Belegmarken: **gemessen** (ein Wert auf einer genannten Skala aus einem genannten Instrument, mit n), **gelesen** (eine kategoriale Einschätzung eines genannten Lesers aus genannten Belegen) oder **behauptet** (eine Aussage ohne Instrument und ohne Beleg). Ein Element steigt nur auf, wenn ein Beleg oder ein Instrument dazukommt; es fällt nie stillschweigend herunter.

Individuelle Antworten werden nur aggregiert berichtet, und nur ab fünf Antworten je Einheit. Kein Item fragt nach individueller Leistung, und kein Ergebnis kann in einem Leistungsgespräch verwendet werden.

5. Vier Zugänge

Ein gemeinsamer Item-Satz über alle Zugänge, damit eine vorher gemessene Einheit nachher wieder gemessen werden kann. Die Aufwände sind Schätzungen.

Zugang	Für	Aufwand und Zeit des Kunden	Was der Kunde danach in der Hand hat
Check	Ein Führungskreis, der seine Einheiten auf dem Bild sehen will	20 Minuten (30 für fünf Einheiten), keine Vorbereitung	Die Karte von drei bis fünf Einheiten, je Position die Marke belegt oder geschätzt, die zuerst zu messende Einheit
Profil	Einheiten mit KI im Einsatz, ohne Berater	28 Management-Items (10–14 Min., im Gespräch), 18 anonyme Mitglieder-Items (5–7 Min., n ≥ 5), drei Ablaufkarten; Bericht binnen eines Arbeitstags	Je Einheit: Potenzial-Radar, vier Balken der Entscheidungsfähigkeit mit Widersprüchen, Durchdringungsort, Kultur-Gate, Governance-Streifen, bis zu drei Versuche mit Abbruchbedingungen
Potenzial	Einheiten, die noch keine KI nutzen	3 Beratertage je Einheit (bis 4,5) plus 1,5 Tage Übergabeaufwand	Baseline vor KI, Abstinenz-Lesart (bewusst / aufgeschoben / blockiert / ungeprüft), Delegationsdecke je Entscheidungsklasse, Governance-Boden, Einstiegsversuche. „In den nächsten zwölf Monaten nicht anfangen“ ist ein zulässiges Ergebnis
Diagnose	Einheiten mit KI in einem Kernablauf	3–5 Beratertage je Einheit plus 1,5 Tage Übergabeaufwand	Viabilitätslesart mit Beleg, Pathologieliste, Delegationskohärenz-Karte, Governance-Aufzeichnung für eine ISO/IEC-42001-Umsetzung, Baseline der Wirkungsmaße, Versuche, Nachmessungsdatum

Die Weiterleitung zwischen den Zugängen: mit dem Check anfangen. Eine Einheit rechts auf der Karte (Hollow, Augmented) geht weiter mit Profil und, wo nötig, Diagnose. Eine Einheit links (Blocked, Dormant) geht weiter mit Potenzial.

6. Was die Methode nicht verspricht

Kein Zugang liefert: einen Score, ein Benchmark gegen andere Unternehmen, ein Zertifikat, eine Roadmap auf eine Stufe, einen Use-Case-Katalog, eine ROI-Rechnung — oder die Zusage, dass sich die Wirkungsmaße bewegen. Die Nachmessung ist genau dafür da, und sie kann auch zeigen, dass nichts passiert ist.

Ein Ergebnis ist ausdrücklich zulässig: **dass eine Einheit in den nächsten zwölf Monaten nicht anfangen sollte** — mit der Begründung und mit der Beobachtung, die uns umstimmen würde. Ein Verfahren, das dieses Ergebnis nicht produzieren kann, ist ein Verkaufsinstrument.

7. Woran man merkt, ob es funktioniert

Sieben Wirkungsmaße je Einheit, zur Baseline und alle sechs Monate, jedes aus einer benannten Aufzeichnung oder als „nicht beobachtbar" gemeldet: Entscheidungslatenz (routine und nicht-routine), Verhältnis von Versuchen zu Skalierungen, Governance-Verstöße und Beinahe-Vorfälle, Output je Vollzeitstelle mit Urteilsaufgabe, Erfahrungspipeline, Überschreibungsrate und Überschreibungsqualität. Latenz wird nie allein berichtet: Eine schnellere Entscheidung, die durch Überspringen der Prüfung erkauft ist, ist das Feld Hollow in Entstehung.

Und das Modell kann verlieren. Jede strukturelle Behauptung trägt die Beobachtung mit sich, die sie widerlegen würde:

- **F1:** Einheiten im Feld Hollow verschlechtern sich — steigende nicht-routine Latenz, steigende Verstöße, schwächere Pipeline, relativ zu Einheiten im Feld Augmented. Widerlegt, wenn sich über zwei Runden kein Unterschied zeigt.
- **F8:** Blockierte Einheiten, die Zugang und Spielraum bekommen, erreichen Erweitert schneller und mit weniger Vorfällen als ruhende Einheiten mit denselben Werkzeugen. Widerlegt, wenn kein Unterschied besteht.
- **F10:** Die vor der Einführung gemessene Entscheidungsfähigkeit sagt das danach erreichte Feld voraus. Widerlegt, wenn die Startfähigkeit mit dem erreichten Feld nichts zu tun hat.

Das ist der Grund, warum die Methode in eine Organisation passt, die schon viel ausprobiert hat: Sie bringt ihre eigenen Zweifel mit.

8. Was das für eine Einheit ohne Zielbild bedeutet

Ein Zielbild lässt sich nicht sinnvoll vor der Verortung schreiben. Wer nicht weiß, wo eine Einheit steht, kann nicht sagen, wohin sie soll — und ein Zielbild, das trotzdem geschrieben wird, beschreibt meist das, was ohnehin geplant war.

Die Methode dreht deshalb die Reihenfolge um. Am Anfang steht nicht das Dokument, sondern eine Sitzung: Ein Führungskreis setzt drei bis fünf Einheiten auf das Bild aus zwei Achsen und markiert jede Position als belegt oder geschätzt. Das dauert zwanzig Minuten und braucht keine Vorbereitung. Danach gibt es zwei Ergebnisse: eine Karte, die zeigt, wie viel man tatsächlich weiß, und eine benannte Einheit, die zuerst gemessen wird.

Das ist auch der Punkt, an dem das fehlende Zielbild sichtbar wird — nicht durch einen Vortrag über das Fehlende, sondern durch das Bild, das in diesen zwanzig Minuten entsteht. Eine Lücke, die niemand sieht, wird nicht bearbeitet; eine, die auf einer Karte steht, hat eine Adresse.

Was danach folgt, ist eine Messung, die eine Einheit selbst lesen kann, ein Plan mit höchstens drei Versuchen, und ein Datum, an dem nachgemessen wird. Der Zielbild-Entwurf entsteht aus dieser Lesart, nicht vor ihr.

9. Herkunft

Die Methode hat extensive Grundlagen. Sie nimmt die Struktur von fünf Mustern und benennt, was sie weglässt: die Durchdringungsstufen und die These vom Urteil als knapper Ressource aus dem Tautai-Modell, das Argument „Zugang ohne Handlungsfähigkeit ist Exposition", das Viable System Model von Stafford Beer mit den Pathologiefamilien von Pérez Ríos, das Bild von Strömungen und beweglichen Elementen statt Meilensteinen von Dave Snowden, und die Unterscheidung von Fähigkeit, Spielraum und Wille aus der Agency-Literatur. Die Kritik an Reifegradmodellen stützt sich auf die systematischen Übersichten von Sadiq et al. (2021) und Smajli, Feldman & Cox (2024).