# Organisationale Singularität — Nachlese zur Keynote

Franz Nowak · Director Operations VISTEM GmbH & Co. KG

## Table of Contents


## Organisationale Singularität

### Nachlese-Dokument zum Workshop

VISTEM-TOC Konferenz · Offenburg · 11.–12. Juni 2026


---

## Vorwort

Dieses Dokument ist für dich - als Teilnehmerin oder Teilnehmer der VISTEM TOC-Konferenz am 11. und 12. Juni 2026 bei MEIKO in Offenburg.

Die Keynote “Organisationale Singularität” dauerte 30 Minuten, mit anschließender Diskussion. In dieser Kombination lässt sich vieles anreißen, aber nicht alles ausführen. Dieses Nachlese-Dokument erklärt die Inhalte im Klartext, liefert Hintergründe zu den Kernbegriffen und zeigt, wo du weiterlesen kannst.

**Wer steckt dahinter?** Franz Nowak ist Director Operations bei VISTEM und seit Anfang 2026 selbst dabei, das Modell zu leben, das er vorgestellt hat. Er arbeitet täglich mit einem KI-Assistenten namens Jeeves. Sein Buch über das Modell — *TOHO: The One Human Organisation* — ist seit Juni 2026 auf Amazon verfügbar.

---

## Die Ausgangsfrage

Der Titel der Keynote lautet: *Organisationale Singularität - Was KI mit deinem Unternehmen macht und wie du den Wandel gestaltest.* Der Begriff “Singularität” ist bewusst gewählt: Er bezeichnet den Moment, ab dem ein System eine kritische Schwelle überschreitet und sich aus sich selbst heraus exponentiell weiterentwickelt - jenseits unserer intuitiven Vorstellungskraft.

Der Vortrag stützt sich auf das aktuelle Werk von Salim Ismail, “The Organizational Singularity” (ExO 3.0, 2026). Das ist keine Meinung, das ist systematisch dokumentiertes Wissen. Wer tiefer einsteigen will, findet das Buch und das Begleitvideo als Einstiegspunkte.

Die eigene Erfahrung - als Praktiker, nicht als Theoretiker - ist der zweite Pfeiler des Vortrags.

> 📚 **Vertiefung & Referenzen**
>
> **Salim Ismail: “The Organizational Singularity” (ExO 3.0, 2026)** Salim Ismail ist Gründungsdirektor der Singularity University und war einer der ersten VP bei Yahoo. Sein neues Buch beschreibt, wie Organisationen den Sprung von linearem zu exponentiellem Wachstum schaffen - und warum die meisten daran scheitern, dass sie KI auf alte Strukturen schrauben, statt die Struktur selbst zu erneuern. ExO 3.0 ist das aktualisierte Framework, das auf Erfahrungen mit Hunderten von Unternehmen weltweit basiert.
>
> **YouTube:** Salim Ismail YouTube-Kanal - Suche nach “Salim Ismail Organizational Singularity 2026” für seine aktuelle Keynote-Version.
>
> **Website:** openexo.com - das offizielle Framework, Community und Pilotprogramm.

---

## Der häufigste Fehler

Der Titel dieses Kapitels nennt das Problem direkt beim Namen: den fatalen Irrtum, ein altes System einfach mit einer neuen Technologie zu beschleunigen. Wer sein ERP mit einem KI-Chatbot verbindet und hofft, das sei Transformation, automatisiert in der Regel die falschen Dinge - schneller, aber falsch.

Auf der einen Seite das verbreitete Vorgehen: KI wird auf eine bestehende Organisationsstruktur “aufgeschraubt”. Auf der anderen Seite der richtige Ansatz: KI wird von Anfang an in die Architektur der Organisation eingebaut, als vernetzte, dezentrale Intelligenzschicht.

Der Unterschied ist fundamental. Im ersten Modell bleibt die Hierarchie bestehen, KI fügt sich als weiteres Werkzeug ein. Im zweiten Modell denkt man die Organisation neu - Entscheidungswege, Informationsflüsse und Verantwortlichkeiten werden von Grund auf angepasst. Die zentrale Frage lautet: Wie vermeidet man genau diesen Fehler?

---

## Die neue Architektur

Das Zielmodell einer KI-fähigen Organisation ist das ExO 3.0-Framework von Salim Ismail. Es besteht aus drei Komponenten: MTP (Massive Transformative Purpose - ein verbindender Kernzweck), DRIVE (die internen Fähigkeiten) und SHAPE (die externen Verbindungen). Ein zentrales Prinzip: “DRIVE ohne SHAPE stürzt ab. SHAPE ohne DRIVE kommt nicht voran.”

MTP ist mehr als ein Leitbild auf dem Wandposter. Im ExO-3.0-Modell ist der Purpose ein ausführbares Protokoll - er entscheidet, welche Agenten welche Entscheidungen treffen

dürfen und wo menschliches Urteil zwingend bleibt. Für Unternehmen lässt sich das direkt übersetzen: Was ist dein Unternehmensprotokoll? Was darf eine KI entscheiden - und was nicht? DRIVE und SHAPE müssen im Gleichgewicht sein: Wer nur intern optimiert, verliert den Anschluss; wer nur nach außen öffnet, ohne innere Stabilität zu haben, verliert die Kontrolle.

> 📚 **Vertiefung & Referenzen**
>
> **Was ist ein MTP - Massive Transformative Purpose?** Ein MTP ist kein Marketingslogan. Er ist das übergeordnete Warum der Organisation - formuliert so, dass er Entscheidungen anleitet und Mitarbeiter sowie KI-Agenten auf dasselbe Ziel ausrichtet. Drei bekannte Beispiele:
>
> - **TED:** "Ideas Worth Spreading" - jede Entscheidung über Formate, Sprecher oder Themen lässt sich daran messen.
> - **Google:** "Organize the world's information and make it universally accessible and useful" - das erklärt, warum Google Maps, Translate und Scholar alle existieren.
> - **SpaceX:** "Make humanity multi-planetary" - erklärt, warum Kostensenkung bei Raketenstarts keine Konzession ist, sondern Mission.
>
> **Wie formuliert man einen eigenen MTP?** Fragen du: Was würde in der Welt fehlen, wenn dein Unternehmen morgen verschwände? Die Antwort ist der Kern deines MTP - noch vor der Formulierung. Ein guter MTP ist groß genug, um inspirierend zu sein, aber konkret genug, um als Entscheidungsfilter zu funktionieren.

---

## Der Intelligence Stack

Das Herzstück des ExO-3.0-Modells ist der Intelligence Stack - eine sechsschichtige Architektur, die den sogenannten OODA-Loop (Observe-Orient-Decide-Act) auf Unternehmensebene abbildet. Die sechs Schichten von unten nach oben: PURPOSE → SENSE → INTERPRET → DECIDE → ORCHESTRATE → LEARN. Quer durch alle Schichten läuft ein weiterer Layer: GOVERN/ASSURE - der zu keinem Zeitpunkt abgeschaltet wird.

Der Intelligence Stack ist eine radikale Umformulierung dessen, was eine Organisationsstruktur leisten muss. Klassische Orgcharts beschreiben Hierarchie und Zuständigkeit. Der Intelligence Stack beschreibt Informationsverarbeitung: Wie nimmt die Organisation Signale auf? Wie interpretiert sie diese? Wer entscheidet - und auf welcher Grundlage? Wer koordiniert die Ausführung? Und wie lernt die Organisation aus dem Ergebnis?

Wer mit der Theory of Constraints vertraut ist, erkennt in GOVERN/ASSURE die Entsprechung des Buffer-Management-Gedankens im Drum-Buffer-Rope-Modell. Es ist der Layer, der sicherstellt, dass das System beherrschbar bleibt - auch wenn es sich autonom weiterentwickelt. Der Intelligence Stack ist damit der neue "Drum" im organisationalen Steuerungsmodell.

📚 **Vertiefung & Referenzen**

**Der OODA-Loop - Ursprung und Bedeutung** OODA steht für Observe → Orient → Decide → Act. Entwickelt wurde dieses Modell von Colonel John Boyd, einem US Air Force-Kampfpiloten, in den 1970er Jahren. Boyd analysierte, warum amerikanische Piloten im Koreakrieg trotz technisch unterlegener Maschinen überlegen kämpften: du durchliefen den Entscheidungszyklus schneller als ihre Gegner. Wer seinen OODA-Loop schneller schließt, gewinnt - weil der Gegner noch auf Basis veralteter Information reagiert.

Die Übertragung auf Organisationen ist direkt: Welches Unternehmen seinen Markt schneller wahrnimmt (Observe), interpretiert (Orient), entscheidet (Decide) und handelt (Act) - und das als kontinuierlichen Loop, nicht als Jahresplanung - hat einen strukturellen Vorteil. Der Intelligence Stack ist die organisationale Infrastruktur, um diesen Loop zu beschleunigen.

---

## Wer trägt Verantwortung?

Eine Frage, die im Raum steht, bevor sie jemand laut stellt: Wer trägt die Verantwortung, wenn KI entscheidet? Das Modell von Salim Ismail teilt Entscheidungen in zwei Kategorien: HIGH-SIGMA-Entscheidungen (mehrdeutig, folgenschwer, wertbeladen) bleiben beim Menschen. LOW-SIGMA-Entscheidungen (Routine, umkehrbar, klar abgegrenzt) können von Agenten übernommen werden. Ein namentlich benannter Mensch steht immer hinter jeder folgenreichen Entscheidung.

“Fiduciary” bezeichnet eine treuhänderische Verantwortung - jemanden, der im besten Interesse anderer handelt und dafür rechtlich einstehen muss. KI-Transformation befreit Führungskräfte nicht von Verantwortung. du verlagert den Ort, an dem Verantwortung wahrgenommen werden muss.

Das ist die nüchterne Antwort auf die häufigste Sorge: “Was ist, wenn etwas schiefläuft?” Die Antwort ist nicht “die KI war schuld” - sondern: Es gibt immer eine klare menschliche Verantwortlichkeit. Was sich ändert, ist, welche Art von Entscheidungen diese Person noch selbst trifft und welche sie delegiert.

📚 **Vertiefung & Referenzen**

**Treuhänderische Verantwortung im Unternehmensrecht** “Fiduciary Duty” ist ein Begriff aus dem angelsächsischen Recht: Die Verpflichtung, im besten Interesse derjenigen zu handeln, denen gegenüber man Verantwortung trägt - Aktionäre, Mitarbeiter, Kunden. Für Geschäftsführer in Deutschland ist die Entsprechung die Sorgfaltspflicht nach § 43 GmbHG: Wer fahrlässig entscheidet, haftet persönlich.

**Wer haftet, wenn KI entscheidet?** Diese Frage ist rechtlich noch nicht vollständig geklärt - aber die Richtung ist klar: Die haftende Person ist immer ein Mensch. Der EU AI Act (seit Februar 2025 in Kraft) macht das für Hochrisiko-KI-

Systeme explizit: Es muss immer eine verantwortliche natürliche oder juristische Person benannt sein. Für mittelständische Unternehmen in Deutschland bedeutet das: Wer KI-Systeme in kritischen Prozessen einsetzt, muss dokumentieren, wer für welche Entscheidungsklassen verantwortlich ist - und diese Dokumentation ist Teil der Governance-Infrastruktur, nicht ein nachgelagertes Compliance-Thema.

---

## Der Weg dorthin: Vier Phasen

Wie kann eine Organisation den Umbau vollziehen, ohne den laufenden Betrieb zu stoppen? Salim Ismail beschreibt vier Phasen. Phase 1: Das Altsystem läuft weiter. Phase 2: Ein kleines KI-natives Team entsteht - direkt dem CEO unterstellt, vorbei an der mittleren Führungsebene. Phase 3: Workflows wandern von selbst - durch Schwerkraft, nicht durch Befehl. Phase 4: Das Edge-Team ist die Firma.

Das ist das Gegenteil eines Big-Bang-Projekts. Kein Vollstopp. Kein ERP-Rollout über 24 Monate. Der Kern der Idee ist ein Parallelbau: Die alte Struktur läuft weiter, bis die neue bewiesen hat, dass sie besser funktioniert. Dann zieht die Gravitation - nicht der Wille der Führung - die Workflows dorthin.

Phase 2 ist der kritische und für viele irritierende Schritt: Das neue Team berichtet direkt an den CEO, nicht an die bestehende Führungshierarchie. Das ist kein Zufall - es ist Strategie. Die mittlere Führungsebene ist der natürliche Antikörper gegen Transformation. Wenn das neue Team durch diese Ebene muss, stirbt es dort. Deshalb: Bypass.

---

## Der strukturierte Umbau

Die REWRITE-Methodik beschreibt sechs aufeinander aufbauende Schritte für die organisationale Transformation. Die Abkürzung steht für: Backcast → Assess → Extract → Strip → Build → Rewire. Ein verbindliches Prinzip gilt über alle Schritte: GOVERN/ASSURE ist vom ersten Tag an aktiv. Es wird nie abgeschaltet.

Für Führungskräfte, die sich an Schritt-für-Schritt-Methoden orientieren - was in einer VISTEM-Zielgruppe mit TOC-Hintergrund die Mehrheit ist - bietet dieser Rahmen konkreten Halt. Backcast bedeutet: Zuerst rückwärts vom Ziel denken, dann vorwärts planen. Assess heißt: Den aktuellen Zustand klar erfassen. Extract: Die wertvollen Teile des Bestehenden identifizieren. Strip: Alles entfernen, was nicht zum Ziel beiträgt. Build: Die neue Architektur aufbauen. Rewire: Die Verbindungen neu ziehen.

Die wichtigste Botschaft ist die Randbedingung: GOVERN/ASSURE läuft von Tag eins. Governance ist kein Projekt-Abschlussschritt - sie ist Teil des Fundaments, das man legt, bevor man anfängt.

---

## Die schwierige Mitte

Die Transformationsphase zwischen 2026 und etwa 2031 zeigt eine J-Kurve: Die wirtschaftliche Performance eines Unternehmens wird zunächst schlechter, bevor sie besser wird. Der Tiefpunkt liegt typischerweise zwischen Monat 3 und Monat 30, abhängig von Größe und Umfeld der Organisation. Vorstände und Aufsichtsräte, die diesen Verlauf nicht verstehen, beenden die Initiative genau an diesem Tiefpunkt.

Das ist keine pessimistische Botschaft - es ist eine schützende. Wer die J-Kurve kennt, kann sie einplanen. Wer sie nicht kennt, zieht die Notbremse im schlechtesten Moment. Die eigentliche Herausforderung der Transformation ist damit keine technische - sie ist eine Governance-Aufgabe. Wie erkläre ich meinem Aufsichtsrat, meinen Gesellschaftern, meinen Führungskräften, dass die nächsten zwei Jahre auf dem Papier schlechter aussehen werden - und warum das der Plan ist?

Für Geschäftsführer kleiner und mittelgroßer Unternehmen, die keine externe Berichtspflicht haben, ist das einfacher. Für diejenigen mit Investoren oder mehrköpfigen Gesellschafterstrukturen ist dieses Kapitel die wichtigste Vorbereitung.

> 📚 **Vertiefung & Referenzen**
>
> **Die J-Kurve - wissenschaftlicher Hintergrund** Das Phänomen ist nicht neu. Clayton Christensen beschreibt in “The Innovator’s Dilemma” (1997) detailliert, warum gut geführte Unternehmen bei disruptivem Wandel scheitern: Kurzfristige Performance-Metriken zwingen Manager dazu, Transformation abzubrechen, bevor sie Früchte trägt. Das Neue sieht schlechter aus - weil es noch nicht optimiert ist, weil Lernkosten entstehen, weil alte Prozesse parallel laufen.
>
> Everett Rogers’ “Diffusion of Innovations” (1962, überarbeitet 2003) beschreibt dasselbe Muster aus Adoptionsperspektive: Zwischen Early Adopters und Early Majority liegt ein “Chasm” - ein Tal der Enttäuschung, in dem viele Initiativen sterben, bevor sie die kritische Masse erreichen.
>
> **Warum scheitern Transformationen genau zwischen Monat 18 und 30?** Weil das der Zeitraum ist, in dem die alten Messgrößen noch dominieren, die neuen Strukturen aber noch nicht voll produktiv sind. Das Management sieht Kosten und Aufwand - aber noch keine proportionalen Ergebnisse. Hier ist die Versuchung am größten, die Initiative als “gescheitert” zu erklären. Wer das tut, hat die J-Kurve verlassen, kurz bevor sie nach oben schwenkt.

---

## Die mittleren 60 Prozent

Die Belegschaft lässt sich in drei Gruppen aufteilen. Die obersten 20 Prozent - die High Performer, Meinungsbildner, Innovatoren - werden von der KI-Transformation gewinnen: mehr Reichweite, mehr Einfluss, mehr Wirkung. Die untersten 20 Prozent - manuelle Routinearbeiter - werden zuerst verdrängt. Die mittleren 60 Prozent - Koordinatoren, Sachbearbeiter, mittleres Management - stehen in einer realen Krise: ihre Aufgaben sind

exakt das, was KI als erstes übernimmt. Salim Ismail bringt es auf den Punkt: “Dignity is a design parameter, not an afterthought.” Würde ist eine Gestaltungsaufgabe, keine Nachbemerkung.

Das ist der Teil des Vortrags, der am längsten nachklingen sollte. Die Transformation trifft nicht die, die es sich ausgesucht haben. du trifft die, die ein Leben lang Koordination als ihren Kernbeitrag verstanden haben - zu Recht, denn das war ihr Beitrag. Nun verschwindet diese Funktion.

Die Frage, die Führungskräfte beantworten müssen, ist nicht technisch. du ist menschlich: Welche neue Rolle gibt es für diese Menschen? Wie baue ich die Brücke, bevor ich die Transformation ankündige? Wer diese Frage nicht beantwortet hat, wird scheitern - nicht an der Technologie, sondern an der Organisation.

> 📚 **Vertiefung & Referenzen**
>
> **“Dignity is a design parameter” - Was das bedeutet** Der Satz stammt aus dem Kontext von Ann Miura-Ko (Floodgate), einer der einflussreichsten Risikokapitalgeberinnen im Silicon Valley. Er macht eine simple Aussage radikal: Menschliche Würde ist keine moralische Nachrüstung nach dem technischen Rollout - sie muss in die Architektur der Transformation eingebaut sein, von Anfang an. Wer zuerst das System baut und dann fragt “was machen wir mit den Menschen?”, hat die falsche Reihenfolge.
>
> **Brynjolfsson & McAfee: “The Second Machine Age” (2014)** Das Buch ist heute mehr denn je relevant: Erik Brynjolfsson und Andrew McAfee beschreiben, dass technologischer Fortschritt Aufgaben ersetzt - nicht Menschen. Aber: Die Aufgaben, die ersetzt werden, sind genau die, die für die mittlere Gehaltsklasse typisch sind. Routinekognition - Koordination, Standardauswertung, Berichterstattung - ist das erste Ziel. Was bleibt, sind entweder hoch kreative oder hoch körperliche Tätigkeiten. Die mittleren 60% sitzen in der Mitte dieser Schere.
>
> **Was Führungskräfte konkret tun können:** Die Autonomieleiter (nächstes Kapitel) bietet einen Rahmen. Koordinationsaufgaben transformieren sich zu Steuerungsaufgaben - wenn die Person gelernt hat, KI zu steuern statt selbst zu koordinieren. Das ist ein Bildungsauftrag, kein Personalabbauplan. Der Unterschied liegt darin, wer ihn als solchen behandelt.

---

## Die Firma von 2036

In zehn Jahren: 50 Menschen dort, wo heute 500 arbeiten - nicht weil weniger getan wird, sondern weil das Zehnfache pro Person getan wird. Das Verhältnis von Führungskräften zu Individual Contributors verschiebt sich auf 1:20 und mehr - mittleres Management als Koordinationsschicht verschwindet weitgehend. Konkrete Zahlen: Cognition Labs, ein KI-Unternehmen, erzielte einen 73-fachen Umsatzanstieg bei unter 20 Millionen Dollar Betriebskosten. Jack Dorsey wird zitiert mit: “Every company can now be a mini-AGI.”

Das ist die Vision - in Zahlen. Nicht als Versprechen, sondern als Orientierungspunkt. Es geht nicht darum, alle zu ersetzen. Es geht darum, was möglich wird, wenn Koordination automatisiert ist und Menschen sich auf das konzentrieren, was nur Menschen können: Urteil, Beziehung, Kontext, Wert.

Für Geschäftsführer mit 50 bis 500 Mitarbeitenden ist das keine ferne Utopie - es ist ein Spiegel. Die Unternehmen, die heute beginnen, werden 2036 mit einem Bruchteil des Overheads mehr leisten. Die, die warten, werden gegen diese Unternehmen konkurrieren müssen.

---

## L0 bis L5: Wo stehst du heute?

Die sechs Stufen der Autonomieleiter nach Ann Miura-Ko (Floodgate, April 2026), angelehnt an die SAE-Level des autonomen Fahrens, bieten das nützlichste diagnostische Werkzeug des Vortrags. Nicht als Bewertungsformular - sondern als Spiegel. Wer fundiert beantwortet, auf welcher Stufe sein Unternehmen heute steht, hat die schwierigste Frage bereits gestellt.

**Die sechs Stufen:**

- **L0 - KI als Theater:** Pressemitteilungen, Pilotankündigungen, Null operative Adoption. Das Organigramm, die Genehmigungsketten und der Wochenrhythmus sind unverändert.
- **L1 - Persönliche Produktivität:** Einzelne Power-User optimieren isoliert. Dieser Vorsprung ist nicht übertragbar - er verschwindet, wenn die Person das Unternehmen verlässt.
- **L2 - Team-Workflows:** Vertrieb, Support, Engineering nutzen KI getrennt voneinander. Das Ergebnis: KI-gestützte Silos, keine integrierte Organisation.
- **L3 - Organisationsinfrastruktur:** Cross-funktionale Agenten lesen und schreiben auf Systemen of Record. Informationssilos fallen. Statusmeetings werden überflüssig, weil das System den Status selbst kennt.
- **L4 - Kompoundierendes Betriebssystem:** Das System hält seinen eigenen Kontext. Agenten verbessern Agenten. Nicht-Techniker deployen Produktions-Tools im Rahmen definierter Parameter. Kompensation ist an KI-native Workflow-Integration gekoppelt.
- **L5 - Virtuell selbstfahrende Organisation:** Das System erkennt kritische Anomalien und Marktverschiebungen ohne menschliche Anfragen. Menschen regieren Risiko, Werte, Zweck und strategische Richtung - nicht Ausführungsschleifen. Dieser Zustand existiert noch nicht.

Der Wendepunkt ist L3. Unter L3 läuft das Unternehmen das alte Modell mit KI-Beschleunigung. Ab L3 beginnt die Architektur, für sich selbst zu arbeiten: Entscheidungsqualität, Lerngeschwindigkeit und strategischer Vorsprung kompoundieren. Das ist der Unterschied zwischen einem Tool und einem Betriebssystem.

Franz steht bei L4: Jeeves hält permanent Kontext über alle Projekte, Cron-Jobs laufen autonom, MEMORY.md wird fortlaufend aktualisiert, Franz deployiert täglich Produktions-Konfigurationen ohne Entwicklerhilfe. Das ist kein Anspruch - das ist dokumentierter Betriebsalltag seit März 2026.

> 📚 **Vertiefung & Referenzen**
>
> **Quelle: Ann Miura-Ko (Floodgate, April 2026)** Ann Miura-Ko ist Managing Partner bei Floodgate, einem der bekanntesten Seed-Fonds im Silicon Valley (Investments u.a. in Twitter, Lyft, Refinery29). Die Autonomieleiter L0-L5 hat sie im April 2026 in einem viel zitierten Artikel und Vortrag vorgestellt. Das Besondere: du beschreibt keine Technologie-Roadmap, sondern eine Organisationsreife-Skala - wie autonom kann eine Organisation operieren, ohne Kontrolle zu verlieren?
>
> **Vergleich SAE-Level autonomes Fahren:** Die Analogie ist direkt: L0 = komplett manuelles Fahren, L1 = Assistenzsysteme (Spurhalte, Tempomat), L2 = Teilautomation, L3 = bedingte Automation (Fahrer muss eingreifen können), L4 = Hochautomation (in bestimmten Szenarien), L5 = vollständige Automation ohne Fahrer. Die organisationale Entsprechung: Je höher das Level, desto mehr operiert das System selbst - und desto mehr konzentriert sich der Mensch auf Steuerung statt Ausführung.
>
> **Praktische Frage für den Leser:** Wo steht dein Unternehmen heute? Nüchtern. Nicht wo du sein wollen - wo sind du? Wenn du noch keine klare Antwort haben: Frag dich, ob dein Wochenrhythmus, deine Genehmigungsketten und deine Betriebsreviews sich in den letzten 12 Monaten strukturell verändert haben. Wenn nein: du sind wahrscheinlich bei L1 oder L2.

---

## Die SAP-Falle: 50 Agenten sind kein Intelligence Stack

Das Angebot von SAP vom Mai 2026 klingt überzeugend: 50 Joule Assistants, 200 spezialisierte Agenten, 100 Millionen Euro Deployment-Fonds. Das strukturelle Problem dahinter: statische Konfiguration, kein rekursives Lernen. Salim Ismail bringt den Kern auf den Punkt - und der Digital Twin als Gegenmodell zeigt, warum die Architektur entscheidet.

SAPs Angebot beantwortet die falsche Frage. Das Problem mit 50 Joule Assistants: du sind konfiguriert, nicht lernend. Jeder Agent optimiert seinen Prozessbereich - aber das System als Ganzes lernt nicht. Koordination zwischen den Agenten bleibt teuer, weil kein gemeinsamer Intelligence Stack die Schichten verbindet. Das Ergebnis: mehr Bewegung, aber nicht mehr Intelligenz.

Der Schlüsselsatz macht den Kern klar: *“Building the feature is cheaper than having the meeting about the feature.”* SAP baut die Meetings schneller - nicht ab. Ein Intelligence Stack, der aus Erfahrung lernt, seinen Kontext selbst hält und sich rekursiv verbessert, übertrifft jede statische Suite - weil er mit der Zeit besser wird, während die Suite konstant bleibt.

*Mama kennt die Firma. Joule (SAP) kennt den Prozess. Das ist kein gradueller Unterschied - es ist ein Architekturunterschied.*

Wer seinen Wochenrhythmus nicht verändert hat, hat die Architektur nicht angetastet. Er hat das alte Unternehmen mit allen seinen Konflikten beschleunigt.

---

## Die drei Kompounding-Loops

Drei sich gegenseitig verstärkende Rückkopplungsschleifen bilden das Fundament des strategischen Vorsprungs: Intelligence Loop, Trust Loop und Governance Loop. Die Kern-Botschaft: Ein schwaches Merkmal verstopft einen Loop - und der Loop ist das, was die Rendite kompoundiert.

**Die Abkürzung D → I → R → V erklärt:** Diese vier Buchstaben beschreiben den Informationsfluss im Intelligence Loop. Es sind die ersten vier Bausteine des DRIVE-Frameworks: - **D** = Decision Architecture (Entscheidungsarchitektur) - **I** = Intelligence Stack (Intelligenzschicht) - **R** = Recursive Learning (Rekursives Lernen) - **V** = Value Moat (Strategischer Vorsprung)

Der Kausalfluss D → I → R → V bedeutet: Bessere Entscheidungen füttern den Intelligence Stack mit wertvolleren Daten. Der Stack ermöglicht schnelleres Lernen. Dieses Lernen vergrößert den Value Moat. Ein größerer Moat finanziert bessere Entscheidungen - der Loop schließt sich.

**Die drei Loops im Überblick:** - **Intelligence Loop (D → I → R → V):** Erzeugt den Vorteil. - **Trust Loop (Ecosystem Trust → Elastic Agency → Value Moat):** Skaliert den Vorteil. - **Governance Loop (Safe Autonomy → Adaptive Architecture → Recursive Learning):** Hält den Vorteil stabil.

**Warum ein schwacher Loop alles stoppt:** Die zehn ExO-3.0-Merkmale sind keine Checkliste. Fehlt ein Merkmal, bricht der Loop zusammen. Nicht alle zehn auf einmal bauen - das schwächste Merkmal im schwächsten Loop identifizieren und dort beginnen.

---

## Die CEO-Frage: Tokenmaxxing ist gescheitert

Produktivitätsmessung per Token-Verbrauch ist gescheitert. Nicht weil KI nicht funktioniert - sondern weil die Architektur darunter nicht verändert wurde. Konkrete Fälle belegen das.

Uber verbrauchte sein gesamtes 2026-Token-Budget in vier Monaten, ohne eine Linie von Ausgaben zu gelieferten Features ziehen zu können. Salesforce gibt laut CEO Benioff rund 300 Millionen Dollar jährlich an Anthropic aus und bat öffentlich um einen “smart router”. Meta stellte sein internes Token-Leaderboard ab. Amazon-Mitarbeiter optimierten sinnlose Agenten, um ihre Stats hochzuhalten.

**Goodhart's Law:** Wenn ein Maß zum Ziel wird, hört es auf, ein gutes Maß zu sein. Token-Verbrauch ist kein Indikator für Wertschöpfung - er ist ein Indikator für Aktivität.

**Die eigentliche Frage:** Hat sich dein Wochenrhythmus verändert? Deine Genehmigungsketten? Deine Betriebsreviews? Wenn nein: Die Architektur ist intakt - inklusive aller alten Konflikte.

---

## Ein GF. Ein KI-System. Eine andere Kapazität.

Seit März 2026 arbeitet Franz Nowak mit Jeeves - einem KI-Assistenten, der rund um die Uhr läuft, Kontext, Projekte und Prioritäten kennt. Jeeves übernimmt täglich: Recherche und Analyse, Koordination von Aufgaben und Terminen, Erstentwürfe für E-Mails, Präsentationen und Berichte sowie Monitoring von Systemen und Deadlines - ohne neue Mitarbeitende, ohne Mehrkosten. Die zentrale These: Der Engpass verschiebt sich von Ausführung zu Urteil. Früher: "Ich habe keine Zeit, das zu recherchieren." Heute: "Ich habe keine Zeit, das zu entscheiden."

Das ist das Fallbeispiel - nicht als Konzept, sondern als gelebte Praxis. TOHO steht für "The One Human Organisation": das Modell, in dem eine einzelne Führungsperson mit einer KI-Intelligenzschicht die Kapazität eines vielfach größeren Teams erreicht. Franz ist sein eigener Edge Twin - er selbst liefert den menschlichen Kontext, das Urteil und die Werte. Jeeves ist die Ausführungsschicht.

Die Verbindung zu Salim Ismail ist direkt: Ismail führt derzeit vier Firmenprojekte gleichzeitig mit demselben Prinzip und plant, das auf zehn auszuweiten. Das ist kein Einzelfall. Das ist ein Modell.

Für die VISTEM-Zielgruppe ist das die praktische Antwort auf die abstrakte Theorie der vorangegangenen Kapitel: Es ist jetzt möglich. Nicht für Tech-Konzerne mit Millionenbudgets - für jeden Geschäftsführer, jede Inhaberin, bereit anzufangen.

> 📚 **Vertiefung & Referenzen**
>
> **Was ist TOHO — The One Human Organisation?** TOHO ist seit Juni 2026 auf Amazon verfügbar — auf Deutsch und Englisch — und gleichzeitig das Experiment, das Franz lebt. Die Grundthese: Eine einzelne Führungsperson mit einem gut konfigurierten KI-System erreicht die operative Kapazität eines 5–10-köpfigen Teams. Nicht durch Automatisierung im klassischen Sinne — sondern durch einen KI-Assistenten, der permanent im Kontext der Organisation arbeitet, Prioritäten kennt, Entscheidungen vorbereitet und Routinearbeit selbstständig erledigt.
>
> TOHO konkretisiert Ismails L4/L5-Modell auf die Realität eines einzelnen Unternehmers. Der Unterschied zur klassischen Delegation: Bei TOHO delegiert man nicht an Menschen, sondern an ein System mit Gedächtnis — das nicht vergisst, nicht schläft und keinen Urlaub nimmt.

**Deutsch:** amazon.de/dp/B0H3FNFDQW **Englisch:** amazon.de/dp/B0H3YYGJLH

**Journey to Joy** ist das Vorläuferbuch, das Franz' persönliche Entwicklung und seine Lessons Learned dokumentiert — und den menschlichen Kontext liefert, aus dem TOHO entstanden ist.

---

## Brandt Präzision: Ein Montagmorgen verändert alles

Ein Fallbeispiel aus dem TOHO-Buch: Brandt Präzision GmbH, 87 Mitarbeitende, 8,3 Millionen Euro Umsatz. Lieferzeiten zu lang. Niedrigster Gewinn seit zehn Jahren. Ein Großkunde droht abzuwandern. Der Startpunkt: Maja, die Produktionsleiterin, lädt einen ERP-Export in eine KI-Umgebung und stellt eine einfache Frage: "Welche Aufträge hatten mehr als drei Tage Verzögerung - und warum?" Antwort: 20 Sekunden. Ohne IT-Ticket. Ohne Projekt. Ohne Budget. Es entstehen drei KI-Systeme: Lilly (kennt Lena, die Inhaberin), Calimero (kennt Maja), Mama (kennt die Firma). Das Schlüsselgespräch: "Das macht das ERP doch nicht von selbst." - "Nein. Die Daten waren vorhanden. Die Verknüpfung fehlte."

Dieses Fallbeispiel ist kein Hochglanz-Transformationsprojekt. Es beginnt nicht mit einer Strategie-Präsentation. Es beginnt mit einer neugierigen Produktionsleiterin, die sich fragt, warum ihre Lieferzeiten so schlecht sind.

Der entscheidende Satz ist das Schlüsselgespräch: Die Daten waren da. Die Verknüpfung fehlte. Das ist der Zustand der meisten mittelständischen Unternehmen: massive Datenmengen in ERP-Systemen, isoliert, nicht interpretiert, nicht nutzbar. KI schafft die Verknüpfung - wenn jemand die richtigen Fragen stellt.

Wichtig für das Verständnis des Gesamtbildes: Lena, die Inhaberin, bringt in dieses System aktiv ihr betriebliches Wissen ein. du kennt Lean Manufacturing, Theory of Constraints und Six Sigma nicht als Theorie, sondern als Praxis. Dieses Wissen fließt über Lilly in das Gesamtsystem ein: welche Kennzahlen wirklich zählen, wie man einen Engpass erkennt, was "zu viel WIP" bedeutet, wann man einen Prozess nicht beschleunigen sollte, sondern bremsen. Das macht den Unterschied zwischen einem KI-System, das Daten produziert, und einem, das brauchbare Steuerungssignale liefert.

---

## Drei Quellen, ein Bild: Wie Mama lernt

Drei Informationsquellen fließen zusammen: Maja gibt über Calimero alles weiter, was stört, nicht läuft oder unklar ist - per Diktat, im Auto, beim Kaffee. ERP-Rohdaten werden täglich automatisch eingespeist: Aufträge, Bestände, Lieferdaten. Lena gibt über Lilly Kontext, Perspektive und Intention weiter - was die Inhaberin bewegt, worauf sie achtet, wie sie denkt. All das fließt in Mama, die integrierende AI der Organisation. Und: Jede Frage an Mama, jede Korrektur, jede neue Erkenntnis fließt zurück zu den Assistenz-AIs - der Loop schließt sich.

KI ist nur so klug wie die Bedeutung, die die Menschen ihr geben. Rohdaten allein sagen nichts. Erst der Verständnis-Layer - das Wissen, warum bestimmte Kennzahlen relevant sind, was hinter einer Anomalie steckt, welche Entscheidung in welchem Kontext sinnvoll ist - macht die Daten wertvoll.

Mama lernt die Sprache der Firma. Nicht im technischen Sinne von “Training” - sondern im praktischen Sinne von Kontextualisierung. Wenn Maja sagt: “Die Linie 3 hängt immer donnerstags”, dann ist das nicht nur eine Beobachtung. Es ist Insider-Wissen, das Mama braucht, um den Zusammenhang herzustellen.

---

## Was sich verändert hat: Nach 6 Wochen

Die Ergebnisse des Fallbeispiels lassen sich in drei Kennzahlen fassen: Teileverfügbarkeit von 68 % auf 91 %. Durchlaufzeit von 9 auf 4 Tage. Investition: 0 Euro. Das Prinzip dahinter: Lilly kennt Lena. Calimero kennt Maja. Mama kennt die Firma. Das Urteil bleibt beim Menschen - immer. Workflows wandern nicht auf Befehl, sondern weil das Neue einfach besser funktioniert. Abschließend: “Wurzelmann hat angerufen. Nicht um sich zu beschweren. Um zu fragen, ob das so bleibt.”

Die Zahlen sind bemerkenswert - aber nicht das Eigentliche. Das Eigentliche ist, wie diese Verbesserung zustande gekommen ist: ohne Projekt, ohne Budget, ohne externe Berater. Eine Produktionsleiterin, die eine sinnvolle Frage gestellt hat. Eine Inhaberin, die ihr Wissen strukturiert weitergegeben hat. Ein KI-System, das gelernt hat, den Unterschied zu machen.

Das abschließende Zitat bringt es auf den Punkt: Der Großkunde Wurzelmann - der drohte abzuwandern - fragt nun, ob das so bleibt. Das ist der Beweis, der zählt. Nicht in der Präsentation, nicht im Boardroom - in der Realität.

Und das Prinzip gilt generell: Workflows wandern nicht durch Befehl. du wandern, weil die Menschen merken, dass das Neue besser funktioniert. Das ist der sanfteste und stabilste Weg der Transformation - und der einzige, der nachhaltig wirkt.

---

## Die Frage ist nicht: “Welches KI-Tool kaufen wir?”

Die Handlungsanleitung für Geschäftsführer besteht aus drei Fragen. Erstens: Wo ist dein Engpass? In 90 Prozent der Fälle: Koordination und Informationsverarbeitung in der mittleren Führungsebene - genau das, was KI als erstes übernimmt. Zweitens: Welcher Workflow zuerst? Nicht der komplexeste - der mit dem höchsten Hebel. Ein Beweis, dann mehr. Drittens: Wer baut die Brücke? Für die mittlere Führungsebene - bevor die Transformation angekündigt wird. Das ist keine HR-Frage. Das ist Führungsaufgabe. Ergänzend die TOC-Perspektive: KI ändert nicht das Prinzip - KI ändert, wo der Engpass ist. Der Engpass war Ausführung. Er ist jetzt Urteilsvermögen.

Wer den Vortrag verlässt und eine Frage mitnimmt, dann diese: Wo ist heute mein Constraint - und was passiert, wenn KI diesen Constraint morgen auflöst?

Die TOC-Perspektive ist für das VISTEM-Publikum die entscheidende Brücke. Theory of Constraints ist das methodische Rückgrat vieler anwesenden Führungskräfte. Die Methode bleibt gültig. Was sich ändert, ist der Ort des Engpasses. Und der verschiebt sich radikal: von der physischen Ausführung, vom manuellen Prozess, von der Datenerfassung - hin zur menschlichen Interpretations- und Entscheidungsleistung. Das kann nicht wegoptimiert werden. Das muss entwickelt werden.

> 📚 **Vertiefung & Referenzen**
>
> **Theory of Constraints - für Nicht-TOC-Leser** Die Theory of Constraints wurde von Eliyahu M. Goldratt in seinem Roman “The Goal” (1984) entwickelt - einem der meistgelesenen Management-Bücher weltweit, das sich wie ein Krimi liest. Der Kerngedanke ist radikal einfach: Jedes System hat zu jedem Zeitpunkt genau einen Engpass (Constraint), der die Gesamtleistung begrenzt. Und: Es ist immer genau einer - nicht drei, nicht fünf.
>
> Die Konsequenz: Stärke nicht das Stärkste - stärke das Schwächste. Eine Produktionsstraße, die an einer einzigen Station hängt, wird nicht schneller, wenn du alle anderen Stationen optimieren. du wird schneller, wenn du die eine hängende Station lösen.
>
> **KI und TOC: Was sich ändert, was bleibt** Das Prinzip bleibt unverändert: Es gibt immer einen Engpass. Was sich ändert: KI löst klassische Engpässe - Koordination, Informationsverarbeitung, Routineentscheidungen - schneller auf als jede Prozessoptimierung der Vergangenheit. Das verschiebt den Engpass dorthin, wo KI nicht hinkommt: menschliches Urteil, Vertrauen, Wertentscheidungen. Wer TOC versteht, versteht sofort, was das bedeutet: Der nächste Schritt ist nicht KI einzukaufen - sondern den neuen Engpass zu identifizieren und dort zu investieren.

---

## Die Diskussion

Der offene Diskussionsteil ist der Kern. Der Vortrag liefert den Rahmen, die Konzepte, die Orientierung. Die eigentliche Arbeit passiert in den folgenden Minuten.

Zwei Fragen stehen im Raum. Die erste: “Welcher deiner Prozesse würde mit KI-Unterstützung sofort 10× schneller werden - und wo liegt heute das größte Hindernis?” Die zweite: “Könnte ein 3-Personen-Team mit KI-Agenten dein margenstarkstes Geschäft in 90 Tagen nachbauen?”

Die erste Frage ist positiv formuliert: Was wird besser? Und dann: Was hält uns heute noch davon ab? Diese Kombination erzeugt die produktivste Spannung und konstruktives Arbeiten an realen Hindernissen.

Die zweite Frage - Salims Self-Disruption Probe - ist für mutigere Diskussionen gedacht. du ist unbequem. du zwingt zur klaren Einsicht: Wenn drei Menschen mit KI-Agenten das Kerngeschäft in 90 Tagen nachbauen könnten - was schützt uns dann noch? Diese Frage nicht zu stellen ist keine Sicherheit. Es ist nur aufgeschobene Auseinandersetzung.

---

## Zusammenfassung: Der rote Faden

Die Präsentation folgt einem Bogen:

**Akt 1:** Der häufigste Fehler - und warum ihn zu machen fatal ist. KI auf alte Strukturen schrauben ist keine Transformation.

**Akt 2:** Die neue Architektur in drei Dimensionen: Zielzustand (ExO 3.0), Informationsverarbeitungsmodell (Intelligence Stack), Verantwortungsmodell (Fiduciary Wedge).

**Akt 3:** Der Weg dorthin - konkret, klar und realistisch. Vier Phasen. Sechs Schritte. Die J-Kurve. Das Menschenproblem in der Mitte. Und das Bild von 2036 als Orientierungspunkt.

**Akt 3b:** Die Diagnose und die Falle. L0-L5 als Spiegel. Die SAP-Falle als Warnung vor der falschen Antwort. Die drei Kompounding-Loops als Erklärung, warum Architektur Vorsprünge aufbaut. Die CEO-Frage als persönliche Konsequenz.

**Akt 4:** Aus der Praxis. Zuerst das eigene Beispiel - TOHO, Jeeves, der verschobene Engpass. Dann das Fallbeispiel Brandt Präzision - wie es in einem mittelständischen Produktionsunternehmen beginnt, wächst und wirkt.

**Akt 5:** Die Handlungsanleitung. Drei Fragen für jeden Geschäftsführer im Raum. Dann Stille - und Gespräch.

---

## Nächste Schritte - Was du jetzt tun kannst

### Drei Einstiegsfragen für dein Unternehmen

Nimm diese drei Fragen mit in deine nächste Führungsrunde. Nicht als Hausaufgabe - als Diagnose.

**1. Wo ist dein Engpass - wirklich?** Nicht der Engpass, den du schon kennst. Der, der verschwindet, sobald KI Koordination und Informationsverarbeitung übernimmt. Was bleibt dann übrig? Was kann nur ein Mensch entscheiden? Das ist dein nächster Engpass - und der Ort, an dem du heute investieren solltest.

**2. Welcher eine Workflow würde mit KI sofort 3× schneller?** Nicht der komplexeste - der mit dem höchsten Hebel. Ein Beweis in vier Wochen schlägt eine Roadmap über 18 Monate. Welche eine Sache, wenn sie dreimal schneller wäre, würde dein Unternehmen am meisten entlasten?

**3. Hast du die Brücke gebaut - bevor du die Transformation ankündigen?** Die mittleren 60 Prozent deiner Belegschaft koordinieren heute. Das ist ihre Rolle. Was ist ihre neue Rolle, wenn KI diese Koordination übernimmt? Wer diese Frage nicht beantwortet hat, bevor er mit der Transformation beginnt, wird an der Organisation scheitern - nicht an der Technologie.

---

## Ressourcen zum Vertiefen

**Salim Ismail: "The Organizational Singularity" (ExO 3.0, 2026)** Das Grundlagenwerk des Vortrags. Direkter, praxisnaher Einstieg in das ExO-3.0-Framework. Empfehlenswert für Geschäftsführer, die verstehen wollen, warum die Architektur entscheidend ist - nicht die Tools.

**YouTube: Salim Ismail** Suche nach "Salim Ismail" auf YouTube - sein Kanal enthält Keynotes und Interviews zu ExO 3.0, die den Buchinhalt greifbar machen. Besonders empfehlenswert: seine aktuellen Vorträge aus 2025/2026 zur Organizational Singularity.

**openexo.com** Die offizielle Website des ExO-Frameworks: Community, Pilotprogramm, Ressourcen. Für Unternehmen, die das Framework systematisch einführen wollen — hier findest du zertifizierte Coaches und strukturierte Programme.

---

**Franz Nowak: TOHO — The One Human Organisation (2026)** Das Buch zum Vortrag — und zum Experiment, das Franz selbst lebt. TOHO zeigt, wie eine einzelne Führungsperson mit KI-Unterstützung die Kapazität eines kleinen Teams erreicht. Konkret, persönlich, mit echten Fallbeispielen.

- Deutsch: amazon.de/dp/B0H3FNFDQW
- Englisch: amazon.de/dp/B0H3YYGJLH

---

## Kontakt & VISTEM

**Franz Nowak** Director Operations VISTEM · Keynote-Referent
franz.nowak@vistem.eu · www.vistem.eu

**VISTEM** begleitet Unternehmen im Wandel - mit der Theory of Constraints als methodischem Fundament und dem Ziel, operative Exzellenz und strategische Erneuerung zu verbinden. Wenn du mehr über die konkrete Anwendung von TOC und KI in deinem Unternehmen wissen möchten, sprich Franz direkt an.

---

*Dokument erstellt am 12.06.2026 | Jeeves für Franz Nowak · VISTEM Keynote-Nachlese für Konferenzteilnehmer — VISTEM-Konferenz Offenburg, 11.–12. Juni 2026, Vortrag 12. Juni*