



# KI für Verwaltungsräte

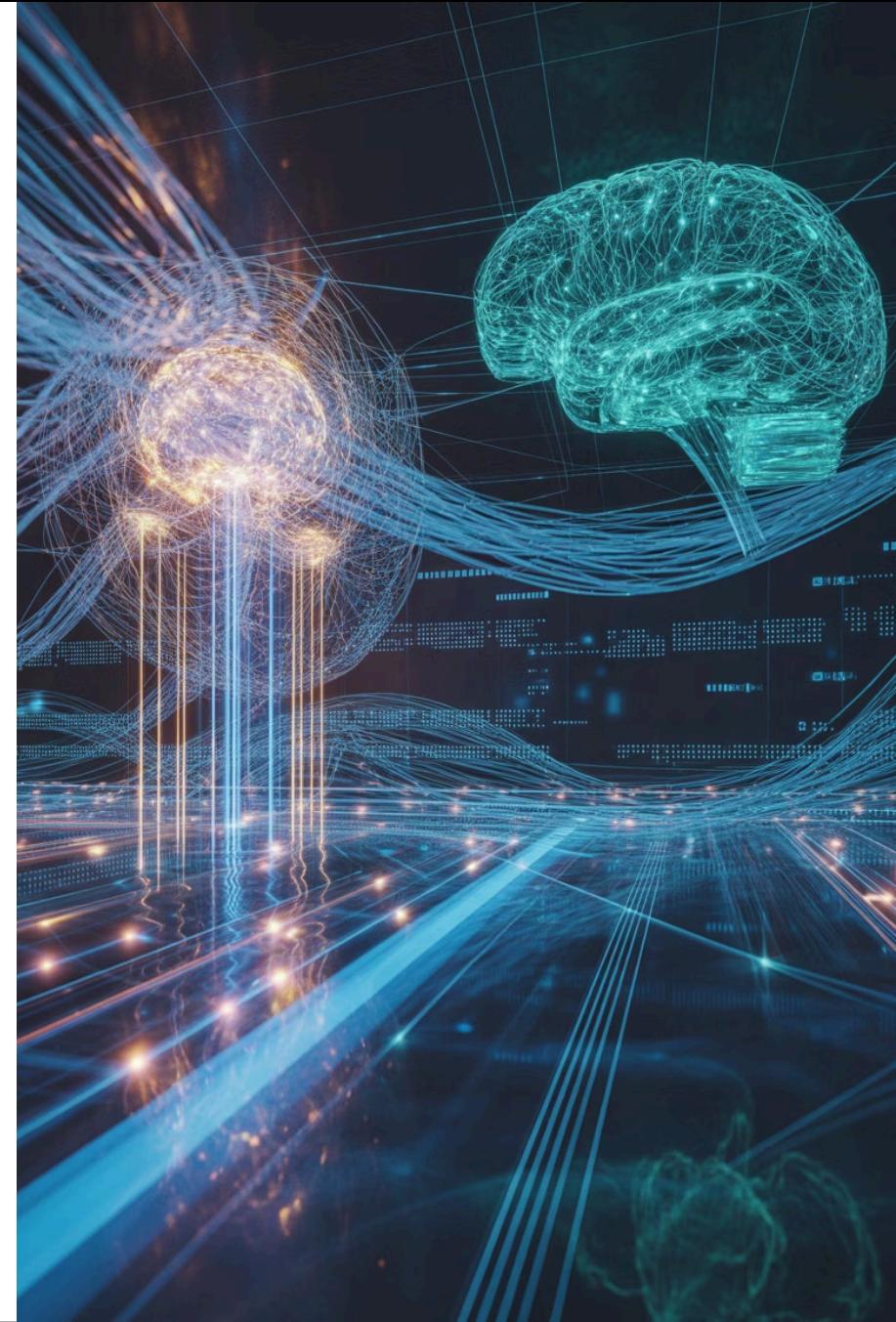
CAS VR - Comboeuf / Winter

HWZ

**Die Ära der  
Künstlichen Intelligenz ist da!**

**Unbegrenzte  
Möglichkeiten.  
Revolutionäre  
Transformation.**

KI ist nicht nur ein Werkzeug, sondern eine treibende Kraft, die jede Branche neu definiert und unser tägliches Leben in ungeahnte Richtungen lenken wird. Freuen Sie sich auf eine Zukunft voller Innovation und Fortschritt, in der das Unmögliche möglich wird!



# 85 AI TERMS EXPLAINED



# 3 menschliche Fähigkeiten, die Sie in einer KI-Welt unersetztlich machen

Quelle: Ed Elson, Prof G Markets

## 1. KURATIERUNG



Das Signal vom Rauschen filtern. Auswählen, was wirklich zählt aus riesigen Informationsmengen.



KI kann generieren, Menschen müssen Qualität und Relevanz erkennen.

**Kernaussage:** Seien Sie ein anspruchsvoller Filter, nicht nur ein Konsument.

## 2. NEUGIER

Nach dem 'Warum' und 'Was wäre wenn' fragen. Nach tieferem Verständnis jenseits von oberflächlichen Antworten suchen. KI beantwortet Fragen, Menschen stellen bessere.

**Kernaussage:** Fördern Sie Ihren Wunsch zu lernen und das Unbekannte zu erforschen.



## 3. KONNEKTIVITÄT

Echte Beziehungen aufbauen und Vertrauen fördern. Brücken zwischen unterschiedlichen Ideen und Menschen schlagen. KI verbindet Daten, Menschen verbinden sich auf emotionaler Ebene.

**Kernaussage:** Pflegen Sie bedeutungsvolle, bedeutungsvolle, menschliche Bindungen.

**FAZIT:** Nutzen Sie Ihren menschlichen Vorteil. Diese Fähigkeiten – Kuratieren, Hinterfragen und Verbinden – sind Ihre einzigartigen Stärken in einer KI-gesteuerten Zukunft.

# 3 Human Skills That Make You Irreplaceable in an AI World

Source: Ed Elson, Prof G Markets

## 1. CURATION



Filtering the signal from the noise. Selecting what truly matters from vast. AI can generate, humans must discern quality and relevance.

**Key takeaway:** Be a discerning filter, not just a consumer.

## 2. CURIOSITY



Asking 'why' and 'what if'. Seeking deeper understanding beyond surface-level answers. AI answers questions, humans ask better ones.

**Key takeaway:** Fuel your desire to learn and explore the unknown.

## 3. CONNECTIVITY



Building genuine relationships and fostering trust. Creating between diverse people. AI connects data, humans connect on an emotional level.

**Key takeaway:** Cultivate meaningful, human-to-human bonds.

**CONCLUSION:** Embrace your human edge. These skills—curating, questioning, and connecting—are your unique advantages in an AI-driven future.

# HOW PEOPLE USE GENERATIVE AI in 2025

This analysis by Marc Zao-Sanders for **Harvard Business Review** groups insights from thousands of forum posts in the last year.



CONTENT CREATION



SUPPORT



LEARNING & EDUCATION



TECHNICAL ASSISTANCE



RESEARCH & ANALYSIS



CREATIVITY & RECREATION

2024

2025

Change in rank  
from 2024

|                            |    |                            |     |
|----------------------------|----|----------------------------|-----|
| Generate Ideas             | 1  | Therapy & Companionship    | +1  |
| Therapy & Companionship    | 2  | Organize Life              | New |
| Specific Search            | 3  | Find Purpose               | New |
| Edit Text                  | 4  | Enhance Learning           | +4  |
| Explore Interests          | 5  | Generate Code              | +42 |
| Fun & Nonsense             | 6  | Generate Ideas             | -5  |
| Troubleshoot               | 7  | Fun & Nonsense             | -1  |
| Enhance Learning           | 8  | Improve Code               | +11 |
| Personalize Learning       | 9  | Creativity                 | +18 |
| General Advice             | 10 | Healthy Living             | +65 |
| Draft Emails               | 11 | Interview Preparation      | +24 |
| Explainers                 | 12 | Generate Images            | +53 |
| Write & Edit Résumé        | 13 | Specific Search            | -10 |
| Excel Formulas             | 14 | Explainers                 | -2  |
| Email Writing              | 15 | Cooking Guidance           | +10 |
| Evaluate Copy              | 16 | Troubleshoot               | -9  |
| Improve Decisions          | 17 | Personalize Learning       | -8  |
| Translation                | 18 | Boost Confidence           | New |
| Improve Code               | 19 | Email Writing              | -4  |
| Draft Document             | 20 | Explain Legalese           | +21 |
| Navigate Personal Disputes | 21 | Child Entertainment        | New |
| Summarize Content          | 22 | Corporate LLM              | New |
| Make a Complaint           | 23 | Student Essays             | New |
| Recommendations            | 24 | Travel Itinerary           | New |
| Cooking Guidance           | 25 | Childcare Help             | +8  |
| Generate Appraisals        | 26 | Medical Advice             | +2  |
| Creativity                 | 27 | Navigate Personal Disputes | -6  |
| Medical Advice             | 28 | Generate Legal Document    | +1  |
| Generate Legal Document    | 29 | Conversations              | New |
| Fix Code                   | 30 | Anti-trolling              | New |



VISUAL CAPITALIST



Where Data Tells the Story

Labels lightly edited from source for readability  
Source: Marc Zao-Sanders, Harvard Business Review



# Was ist ein Prompt?



Ein Prompt ist die **Eingabeaufforderung oder Anweisung**, die Sie einem Large Language Model (LLM) geben, um eine bestimmte Ausgabe zu generieren.

Es ist im Wesentlichen Ihre Art, mit der KI zu "sprechen" und ihr mitzuteilen, was Sie erreichen möchten. Ein effektiver Prompt ist präzise, klar und führt das Modell zum gewünschten Ergebnis.

- Denken Sie daran: Die Qualität des Prompts hat direkten Einfluss auf die Qualität der generierten Antwort.

# Prompt-Qualität entscheidet über Ergebnisqualität

## Schlechter Prompt

"Schreibe eine Zusammenfassung"

### Ergebnis:

Dieses Dokument diskutiert künstliche Intelligenz und ihre Anwendungen im Geschäftsbereich. KI kann Effizienz verbessern. Unternehmen sollten KI-Adoption erwägen.

**Nutzbarkeit:** Niedrig, generisch, unspezifisch

## Guter Prompt

"Erstelle eine 3-Absätze Executive Summary dieses Quartalsberichts für unsere VR-Sitzung. Fokus:

- 1) Umsatztrends vs. Vorquartal,
- 2) Drei operative Risiken und Mitigationen,
- 3) Strategische Implikationen für KI-Investitionen.

Max. 5 Sätze pro Absatz. Verwende spezifische Zahlen."

### Ergebnis:

Q3-Umsatz wuchs um 12% YoY auf CHF 145M, getrieben durch Enterprise-Segment (+18%), kompensiert durch Consumer-Segment (-5%). Margenrückgang von X auf Y ist Lieferkettenverzögerungen zuzuschreiben...

**Nutzbarkeit:** Hoch, spezifisch, VR-bereit

## 10x Qualitätsunterschied

Qualität des Inputs bestimmt Qualität des Outputs – nicht Magie, sondern Klarheit

# Prompting-Frameworks für strukturierte Ansätze



## Chain-of-Thought

**Anweisung:** "Denke Schritt für Schritt, bevor du antwortest"

**Warum:** Zwingt zu sequenziellem Denken, reduziert Halluzinationen

**Beispiel:** "Wir erwägen Fusion. Schritt für Schritt: Was sollten wir evaluieren?"



## Few-Shot Prompting

**Anweisung:** 2-3 Beispiele des gewünschten Formats vorab liefern

**Warum:** LLM lernt besser von Beispielen als von abstrakten Anweisungen

**Beispiel:** Beispiel 1: [Format] | Beispiel 2: [Format] | Nun bewerte diese Risiken:  
[...]



## Role-Playing / Persona

**Anweisung:** "Du bist ein erfahrener VR-Berater mit 20 Jahren M&A-Erfahrung"

**Warum:** Setzt Kontext, verbessert Antwortqualität

**Beispiel:** "Als CFO: Sollten wir Firma X akquirieren? Strukturierte Empfehlung."



## System vs. User Prompt

**System:** Hochrangige Anweisung, Ton, Leitplanken (einmalig gesetzt)

**User:** Spezifische Frage/Aufgabe (variiert pro Anfrage)

**Vorteil:** Konsistenz + Flexibilität

**Pro-Tipp:** Kombinieren Sie Frameworks für beste Ergebnisse – Few-Shot + Chain-of-Thought = kraftvoll

# Übung: Prompt Engineering

15 Minuten Übung, 10 Minuten Feedback-Sharing im Plenum

## Unstrukturierte Anfrage ("einfach drauflos fragen")

- Unser CIO sagt, wir müssen etwas mit Quantencomputing machen? Is that a thing?"
- Okay, und was heisst das für mich als board member?  
usw

## Strukturierter Prompt



Miro

CAS Verwaltungsrat (CAS VR25) – Miro



# Vier Prompt Konzepte für effektiven KI-Einsatz

## Prompt Engineering

Präzise Fragestellung für 10x bessere Ergebnisse

*Beispiel:* "Schreibe Zusammenfassung" → "Erstelle 3-Absätze Executive Summary für VR-Sitzung mit Fokus auf Finanzkennzahlen"

## Context Engineering

Hintergrundinformationen liefern für bessere Entscheidungen

*Vorteil:* LLM trifft fundierte Aussagen basierend auf Ihrem Kontext

## Reverse Prompting

Nach Iteration: "Welcher Prompt hätte direkt zu diesem Ergebnis geführt?"

*Nutzen:* Kontinuierliche Lernschleife, Zeitersparnis

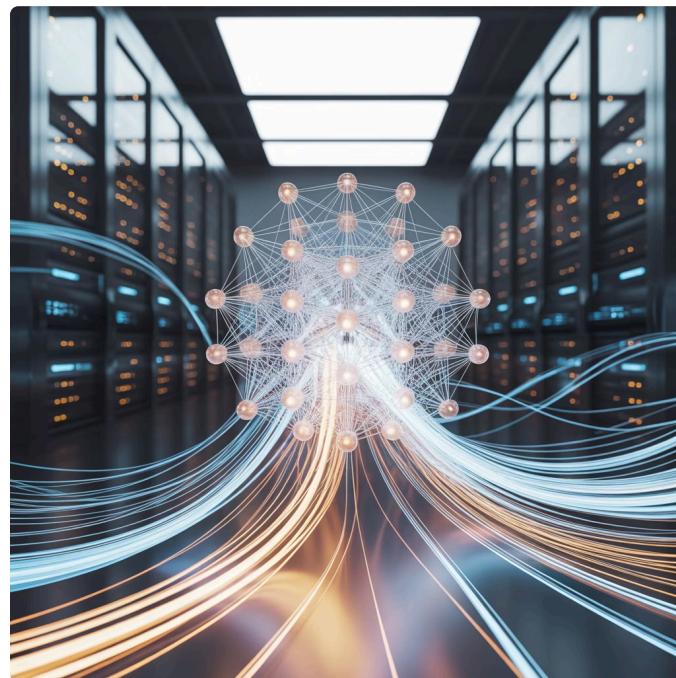
## Prompt-Optimierung

LLM helfen beim Verfeinern Ihrer Prompts

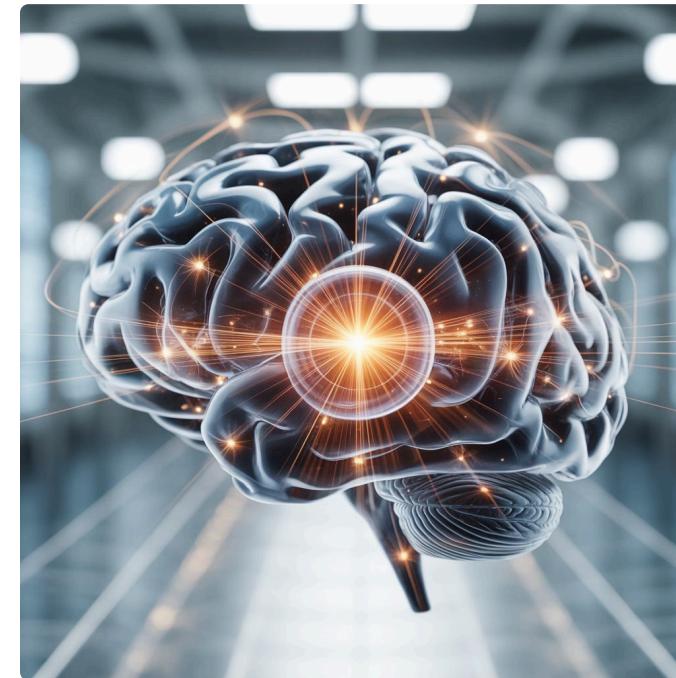
*Meta-Ebene:* "Hilf mir, diesen Prompt effektiver zu formulieren"

# Training vs. Inference – Die kritische Unterscheidung

Training Phase



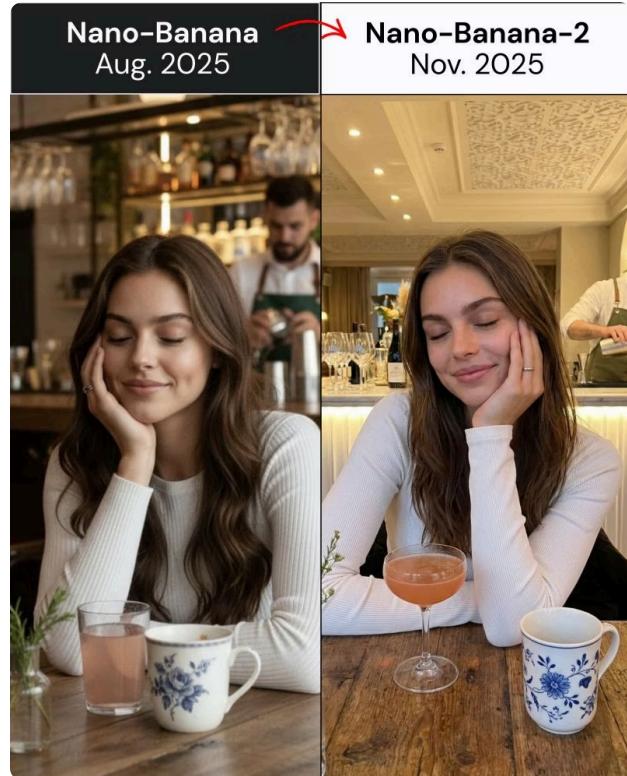
Inference Phase



- Hier lernt das Modell tatsächlich
- Erfordert massive Rechenressourcen und Zeit
- Ihr Unternehmen führt dies NICHT durch, wenn Sie ChatGPT oder ähnliche Tools verwenden

- Das Modell ist eingefroren; es findet kein Lernen statt
- Sie senden eine Anfrage und erhalten eine Antwort
- Häufiges Missverständnis: Viele denken, das Modell lernt in Echtzeit aus ihren Daten – das ist FALSCH

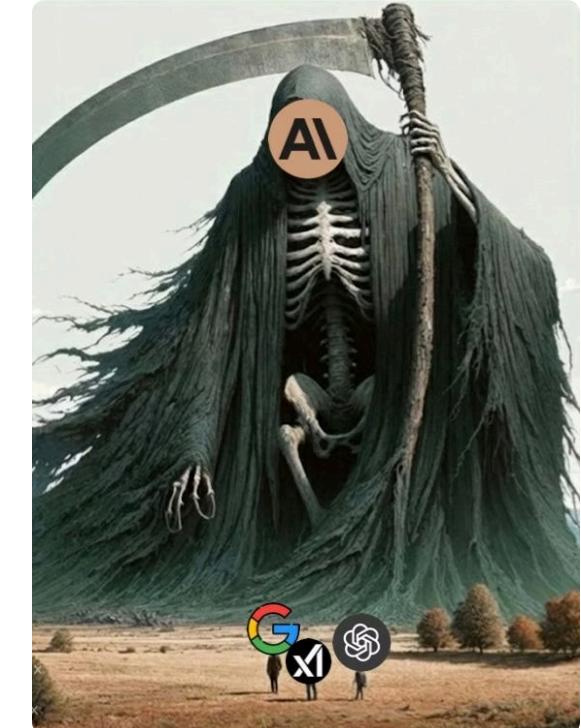
# The speed of AI: Don't chase the tools!



Anstatt sich nur auf einzelne Modelle zu fixieren, konzentrieren Sie sich auf eine strategischere Herangehensweise:

- Definieren Sie Ihre Prozesse, Datenflüsse und Integrationspunkte.
- Bauen Sie ein robustes architektonisches Gerüst auf.
- Integrieren und tauschen Sie neue KI-Tools nahtlos aus, sobald sie verfügbar sind.

Dieser anpassungsfähige Rahmen sichert Ihren strategischen Vorteil und langfristige Relevanz ohne ständige Neuentwicklung.



# McKinsey KI-Bericht 2025

Agenten-KI sieht in Demos magisch aus, aber die Skalierung über Teams und Systeme hinweg (insbesondere in Unternehmensumgebungen) ist der Punkt, an dem sie immer noch versagt.

Einfache, einstufige Automatisierung (Tool-Calling) ist derzeit zuverlässig umsetzbar.



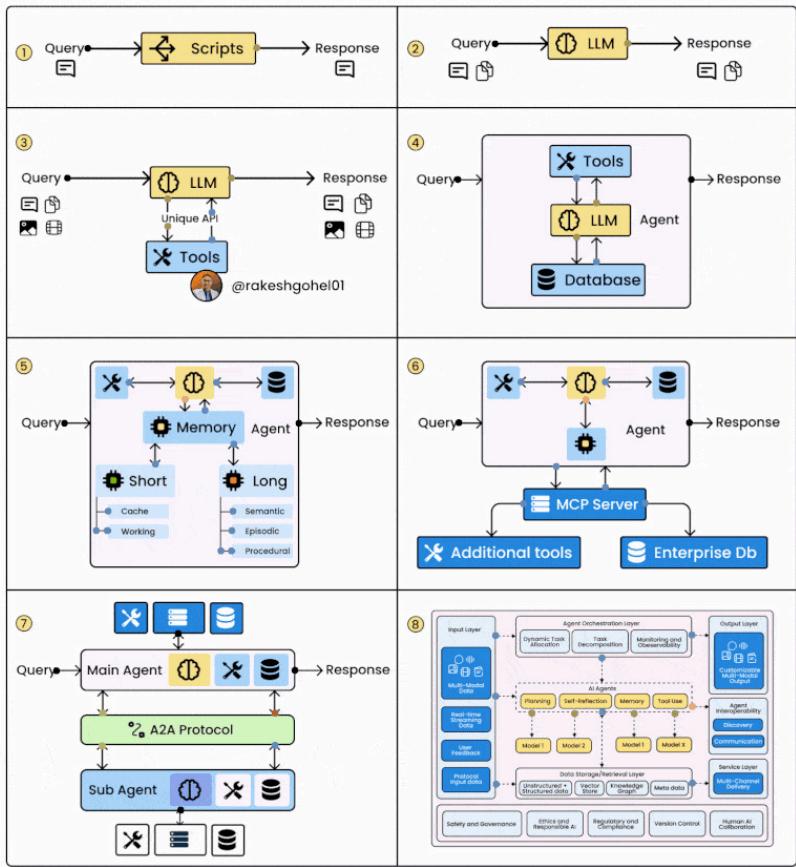
- 1** 90 % der Unternehmen „nutzen KI“, doch 67 % stecken immer noch im Pilotstadium fest. Das „Corporate AI Theater“ ist lebendig und allgegenwärtig.
- 2** 62 % der Organisationen experimentieren mit KI-Agenten, 23 % skalieren KI-Agenten. Die meisten davon sind in der Technologie- und Gesundheitsbranche zu finden.
- 3** Die Wirkungslücke ist enorm. 64 % geben an, dass KI die Innovation fördert, aber nur 39 % verzeichnen tatsächliche EBIT-Zuwächse.
- 4** Die Leistungsträger (Top 6 %) denken größer. Sie gestalten Arbeitsabläufe neu, setzen Wachstumsziele und investieren echte Budgets, nicht nur in Proof-of-Concepts (PoCs).
- 5** Führungskräfte, die KI persönlich vorantreiben, sind dreimal so wahrscheinlich, diese auch zu skalieren. Das ist nachvollziehbar.
- 6** Die Gewinner nutzen KI, um die Arbeitsweise zu transformieren, nicht nur, um sie zu beschleunigen.
- 7** Das durchschnittliche Unternehmen misst Effizienz. Die besten messen, wie schnell ihre Agenten agieren können.
- 8** Das Risikomanagement holt auf, wobei 51 % bereits erlebt haben, dass KI nach hinten losging, meist aufgrund von Ungenauigkeit.
- 9** Die Auswirkungen auf die Belegschaft sind unklar. 32 % erwarten Stellenstreichungen, 13 % erwarten Wachstum, alle anderen raten.

# KI Evolution



Rakesh Gohel @rakeshgohel01

## Evolution of AI Agents



### Stufe 1: Regelbasierte Skripte

- Einfache If-Then-Logik, die vordefinierte Antworten ausführt
  - Kein Lernen, nur fest codierte Entscheidungsbäume
- Determinierte Logik funktioniert, skaliert aber nicht mit Komplexität

### Stufe 2: Grundlegende LLM-Integration

- Direkte Abfrage an ein Sprachmodell für intelligente Antworten
  - Immer noch zustandslos, kein Kontext zwischen Interaktionen
- Intelligenz allein reicht nicht aus – wir brauchen Kontext und Tools

### Stufe 3: LLM mit Tool-Zugriff

- Sprachmodelle können nun externe APIs und Funktionen aufrufen
  - Beginn der dynamischen Kapazitätserweiterung
- LLMs müssen mit externen Systemen interagieren, um Antworten zu verbessern

### Stufe 4: RAG-gestützte Agenten

- Agenten rufen Echtzeitdaten aus Datenbanken ab, bevor sie antworten
  - Kombiniert Wissensabruft mit Generierung für Genauigkeit
- Echtzeit-Wissensabruft reduziert Halluzinationen

### Stufe 5: Speichergestützte Agenten

- Kurzzeit- (Cache, Arbeitsgedächtnis) und Langzeitgedächtnis (semantisch, episodisch, prozedural)
- Kontextsensitive Antworten basierend auf dem Gesprächsverlauf
- Kontextbeibehaltung über Interaktionen hinweg ist entscheidend für die Benutzererfahrung

### Stufe 6: MCP-verbundene Agenten

- Das Model Context Protocol (MCP) ermöglicht standardisierte Verbindungen zu Unternehmensdatenbanken und zusätzlichen Tools
  - Skalierbare Integration in bestehende Geschäftssysteme
- Standardisierte Protokolle ermöglichen eine skalierbare Unternehmensintegration über Tools hinweg

### Stufe 7: Multi-Agenten-Kollaboration

- Hauptagent koordiniert mit spezialisierten Unteragenten
  - Verteilte Aufgabenbearbeitung mit koordinierter Ausführung
- Spezialisierte Agenten, die zusammenarbeiten, übertreffen einzelne Generalisten-Agenten

### Stufe 8: Zukunft der Enterprise Agenten-Orchestrierung

- Vollständiges System mit Input-/Output-Layern, Aufgabenzerlegung, KI-Agenten-Orchestrierung
  - Mehrere Modelle, Datenspeicherung/-abruft und Service-Layer arbeiten zusammen
  - Produktionsreif mit Sicherheit, Governance und Versionskontrolle
- Produktionssysteme benötigen Governance, Überwachung und strukturierte Layer

# Das Technologie-Spektrum: Von einfach bis komplex



## RPA

**Deterministisch, regelbasiert**

Rechnungsverarbeitung, Dateneingabe, Legacy-Integration

*Limitierung:* Keine Ambiguität, explizite Regeln nötig

## LLM Workflows

**Strukturiert, KI-intelligent**

Lebenslauf → Extraktion → Weiterleitung

*Limitierung:* Keine Selbstkorrektur, folgt starrer Sequenz



## Chat-Modelle

**Interaktiv, menschengesteuert**

ChatGPT, Claude – Exploration & Ad-hoc-Analyse

*Limitierung:* Nicht skalierbar für Wiederholung

## Agentische Systeme

**Autonom, selbstkorrigierend**

Markt überwachen → Recherche → Aktionsplan → Genehmigung

*Frontier:* Mehrere Agenten, spezialisierte Expertise

# Agent-Frameworks: Sie müssen kein Python programmieren

| n8n                               | Zapier                                  | CrewAI                                    | Microsoft Copilot Stack            |
|-----------------------------------|---|---|------------------------------------|
| <b>Visueller Workflow-Builder</b> | <b>Cloud-nativ, anfängerfreundlich</b>  | <b>Python-basiert, elegant</b>            | <b>M365-integriert</b>             |
| Open-Source, 1000+ Konnektoren    | Schnelle Integrationen, "wenn X dann Y" | Multi-Agent-Orchestrierung, rollenbasiert | Native Integration, Single Sign-On |
| <i>Timeline:</i> Tage bis Wochen  | <i>Timeline:</i> Stunden bis Tage       | <i>Timeline:</i> Wochen                   | <i>Timeline:</i> Wochen            |

**Schlüsselentscheidung:** Das beste Framework ist jenes, das Ihr Team in 18 Monaten noch warten kann



# Mein eigener Workflow: Wie diese Vorlesung entstand

01

## Miro – Mentale Organisation (30 Min.)

Ideen sammeln, Themen strukturieren, Mind Map der Lernziele, logischen Ablauf erstellen

02

## Perplexity – Recherche & Iteration (60 Min.)

"Ich möchte 90-Min-KI-Vorlesung für angehende VR erstellen..." → Perplexity recherchiert, synthetisiert, schlägt Ansätze vor

03

## Gamma – Prompt-Generierung & Folienstruktur (90 Min.)

Verfeinerten Brief aus Perplexity nehmen → Detailliertes Gamma-Prompt erstellen → Initiales Deck generieren

04

## Iterative Verfeinerung – Einzelne Folien polieren (120 Min.)

Jede Folie überprüfen → Schwache Folien: gezieltes Prompting → Iterieren bis perfekt

**Zeitinvestition gesamt:**

**~5 Stunden** für 90-Min-Vortragsdeck

**Vorheriger Ansatz (ohne Workflow):**

**20-40 Stunden** manuelle Erstellung

**Der Multiplikator:** Gute Prompts + richtige Tools = 4x Geschwindigkeit



# Beispiel: Wie diese Vorlesung entstand

1

## Miro – Mentale Organisation (30 Min.)

Ideen sammeln, Themen strukturieren, Mind Map der Lernziele, logischen Ablauf erstellen

2

## Perplexity – Recherche & Iteration (60 Min.)

"Ich möchte 90-Min-KI-Vorlesung für angehende VR erstellen..."

→ Perplexity recherchiert, synthetisiert, schlägt Ansätze vor

3

## Gamma – Prompt-Generierung & Folienstruktur (90 Min.)

Verfeinerten Brief aus Perplexity nehmen

→ Detailliertes Gamma-Prompt erstellen

→ Initiales Deck generieren

4

## Iterative Verfeinerung – Einzelne Folien polieren (120 Min.)

Jede Folie überprüfen

→ Schwache Folien: gezieltes Prompting

→ Iterieren bis perfekt

**Zeitinvestition gesamt:**

**~5 Stunden** für 90-Min-Vortragsdeck

**Vorheriger Ansatz (ohne Workflow):**

**20-40 Stunden** manuelle Erstellung

**Der Multiplikator:** Gute Prompts + richtige Tools = 4x Geschwindigkeit

# Entscheidungsmatrix: Manuell vs. Automatisiert

## 1. Wiederkehrend oder Einmalig?

**Wiederkehrend**

(täglich/wöchentlich/monatlich) →

Automatisieren

**Einmalig** oder selten

→ **Oft manuell besser**

**Hybrid:** Einmalig, aber 20 Outputs →

Lohnt sich trotzdem

## 2. Aufwand:Nutzen-Verhältnis?

**Formel:** Setup-Zeit ÷ Zeitersparnis × Jahreshäufigkeit

**Beispiel 1:** 2h Setup, 30 Min. manuell,  
1×/Jahr = **Nicht automatisieren**

**Beispiel 2:** 8h Setup, 15 Min. manuell,  
250×/Jahr = **Automatisieren!** (Spart  
62.5h/Jahr)

## 3. Sind Hybrid-Modelle akzeptabel?

**Hybrid = unterschätzt:** 80%

Automatisierungsvorteil mit 20%

Setup-Kosten

**Beispiel:** 20 User-  
Zusammenfassungen kopieren  
(manuell, 10 Min.)

→ Workflow generiert 20  
personalisierte Reports (automatisiert)

Hybrid oft optimal

**Die grössten Fehler:** Etwas automatisieren, das nur alle 6 Monate passiert | Alles manuell belassen, was täglich passiert

# Informationsautomatisierung in der Praxis

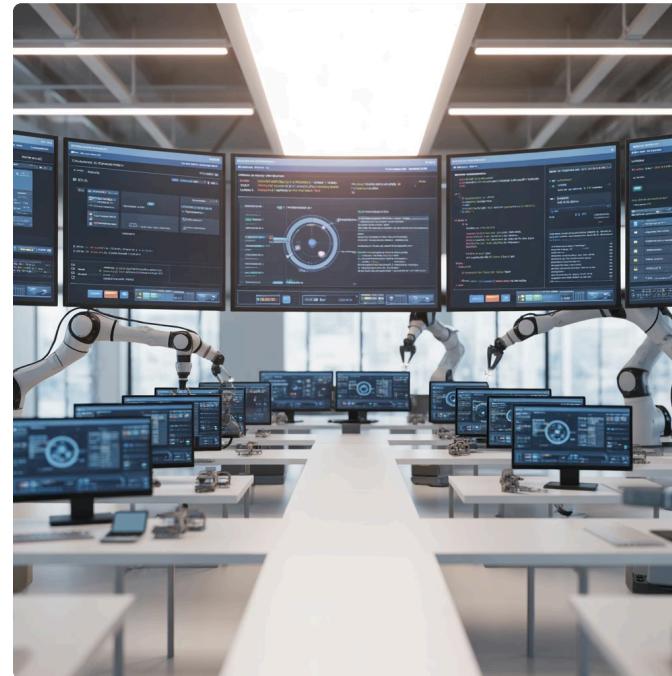
## Perplexity: Research-Automatisierung



- **Echtzeit-Recherche** über mehrere Quellen
- **Synthese** widersprüchlicher Informationen
- **Automatische Quellenangaben**
- **Planbare wiederkehrende Recherchen**

*Use Case:* Tägliche Markttrend-Analyse → automatisch geliefert

## ChatGPT: Task-Automatisierung



- **Informationen klassifizieren** (E-Mails, Tickets)
- **Strukturierte Antworten generieren**
- **Downstream-Aktionen auslösen**
- **Multi-Step-Reasoning** mit Tool-Aufrufen

*Use Case:* E-Mail → Klassifizierung → Weiterleitung → Antwort

**Hybrid-Ansatz:** Perplexity sammelt, ChatGPT verarbeitet und handelt – vollständige Informationspipeline

# Real-World Szenarien aus der Praxis

## Vollautomatisiert (Täglich)



### Daily User Report

- Was: Tägliche User-Metriken
- Häufigkeit: 365x/Jahr
- Setup: 8h, dann fire-and-forget
- Nutzen: Nie ein Tag verpasst

### Subscription Report

- Churn, Expansion, MRR
- Immer aktuell, automatisches Alerting

## Hybrid (Monatlich)



### User Behavior Analysis

#### Manuell (10-15 Min.):

- 20 User-Records kopieren
- In strukturiertes Format organisieren

#### Automatisiert (5 Min.):

- Workflow generiert 20 unique Summaries
- Personalisierte Insights pro User

Zeitersparnis: ~120-140h/Jahr

## Einmalig (Manuell)



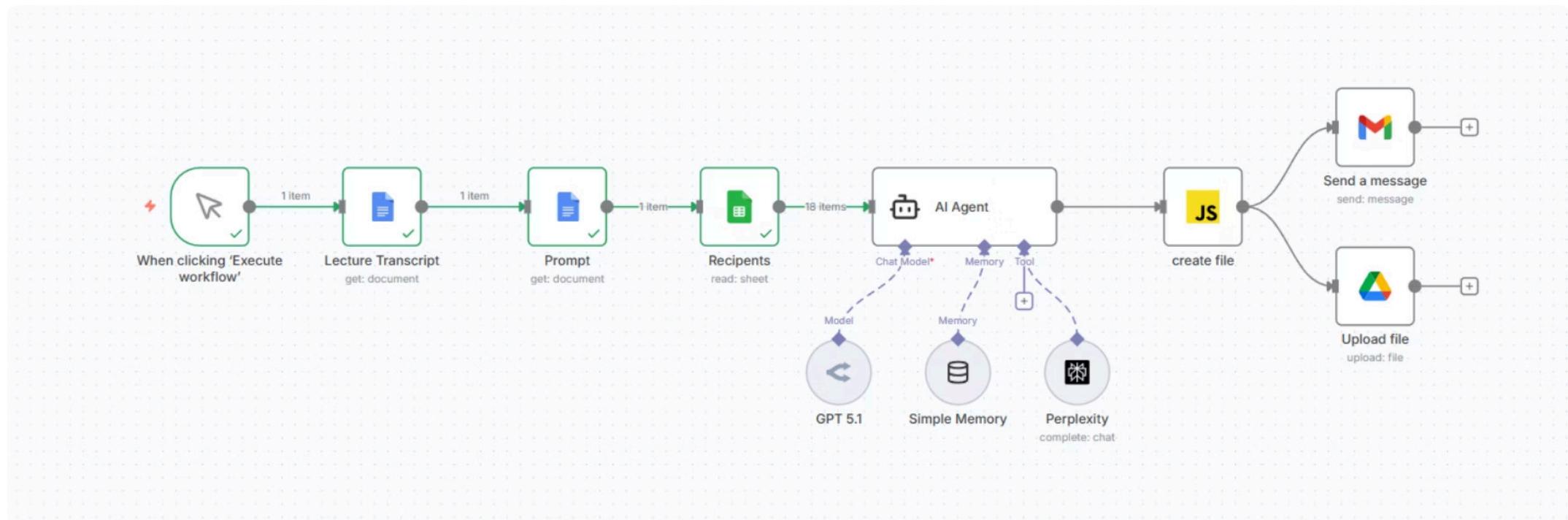
### Infografik-Summary

- Was: Custom Visual Summary
- Warum manuell: Einzigartiges Layout, kreative Entscheidungen
- Aufwand: 2-3h, aber hohe Wirkung
- Automation: 8-12h Templates erstellen

**Entscheidung:** Manuell, einmalig, hohe Qualität

**Strategisch denken:** Daily Reports = No-Brainer automatisieren | Hybrid = wo die Magie passiert | Kreatives = manuell belassen

# Beispiel: Personalisierte Zusammenfassung



# Der Forward-Deployed Engineer: Die kritische Rolle



## Was ist ein FDE?

Kein reiner Software-Engineer, kein reiner Product Manager, kein reiner Business Analyst

**Tatsächlich: Hybridrolle, die verbindet:**

- Produktdenken (versteht Business-Problem tief)
- Engineering-Expertise (kennt Architektur, Tools, Code)
- Business-Kompetenz (versteht ROI, Machbarkeit, Constraints)
- KI-Literacy (kennt Modelle, Prompting, Agentische Systeme)
- Kommunikationsstärke (übersetzt zwischen Business und Tech)



### Versteht Business-Problem

Frage: "Was lösen wir wirklich?" | Identifiziert Constraints



### Mapped auf richtige Technologie

Weiß wann KI vs. RPA vs. Manuell | Wählt Framework basierend auf Use Case



### Prototypisiert schnell

Tage, nicht Wochen | Validiert Annahmen | Zeigt Wert vor großer Investition



### Überbrückt Business ↔ Tech

Spricht beide Sprachen | Erklärt Trade-offs | Vertritt gute Architektur



**FORWARD  
DEPLOYED  
ENGINEER**

AI Engineering is dead ☠

The hottest tech job of 2026 is the Forward Deployed Engineer. And every AI lab is fighting to hire them.

If 2025 was the year of building AI prototypes, 2026 will be the year of deploying them and that's exactly why the Forward Deployed Engineer (FDE) is exploding.

Job postings have already grown 800%, and salaries are reaching levels we normally associate with hedge funds or top-tier quant firms.

Why? Simple:

95% of enterprise AI projects fail not because of models... but because nobody knows how to deploy them.

FDEs are the fix.

An FDE is part full-stack engineer, part product strategist, part embedded consultant. They don't just "build features", they go inside customer environments and make AI actually work in the real world: legacy systems, compliance, messy data, confused stakeholders, and all.

OpenAI, Mistral, xAI, Google, Anthropic, Palantir, and every serious AI Lab are hiring FDEs like crazy.

Salaries?

- ◆ \$150k-\$220k for mid-level
- ◆ \$200k-\$300k+ for senior
- ◆ \$300k-\$450k+ for principal
- ◆ Contractors in Europe? Up to £1,300/day.

This role will replace dozens of today's jobs: solution architects, consultants, customer engineers data scientists, even parts of product and data teams.

Because in 2026, the only thing that matters is getting AI into production (at scale).

And nobody does that better than a Forward Deployed Engineer.

# Zusammenfassung & Key Takeaways

## 1 KI ist Architektur, nicht Tools

Komponenten verstehen: LLM, Orchestrierung, Speicher, Integration | Prinzipien fokussieren, nicht Plattformen | Für Wandel bauen

## 2 Wissen, wann was einsetzen

RPA: Deterministisch, hohes Volumen | LLM Workflows: Strukturiert, KI-intelligent | Chat: Experimentation | Agents: Komplex, adaptiv, autonom

## 3 Prompt-Qualität = Output-Qualität

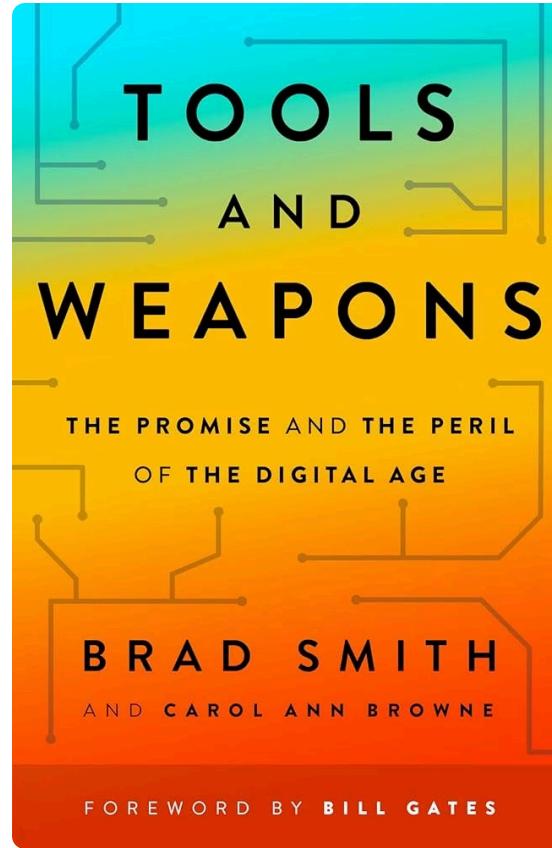
Gute Prompts = 10x bessere Ergebnisse | Frameworks helfen (Chain-of-Thought, Few-Shot, Role-Playing) | Iterieren und verfeinern

## 4 Strategisch automatisieren

Wiederkehrend + hohes Volumen → Vollautomatisierung | Einmalig oder kreativ → Manuell | Hybrid-Ansätze oft optimal

## 5 Kompetenz sicherstellen (intern oder extern)

Forward-Deployed Engineers sind 10x-Multiplikatoren | Sie überbrücken Business, Tech und KI-Literacy | Strategische Fähigkeit, nicht nur IT



—> Sundar Video

# Fragen & Diskussion

## Leitfragen zur Reflexion:

- Welche Automatisierungsmöglichkeiten existieren in Ihrer Organisation?
- Wer könnte Ihr FDE sein?
- Was ist Ihre größte KI-Sorge?
- Wo sehen Sie Quick Wins für KI-Adoption?



# Referenzen & Ressourcen

**Oliver Mölander – AI Architecture Thinking**

[LinkedIn Post](#)

**Rakesh Gohel – Building Blocks of AI Agents**

[LinkedIn Post](#)

**eordax – Architecture Thinking for Long-Term Resilience**

[LinkedIn Post](#)

**eordax – Forward-Deployed Engineer Concept**

[LinkedIn Post](#)

**Emily Smith – LLM Workflows vs RPA vs Agentic AI**

[LinkedIn Article](#)

**Visual Capitalist – Ranked: Things People Use AI for (2025)**

Adoption patterns and industry trends



# Heutige Lösungen sind morgiges Legacy

## ✗ Tool-zentrisch (Fragil)

Zapier + GPT-4 + Azure fest verkoppelt

**Problem:** Neues Tool erscheint → Vollständiges Redesign

12 Monate später: "GPT-5 ist besser, aber unsere Prompts sind für GPT-4 optimiert. Zapier zu teuer, wir wollen n8n."

**Resultat:** Kompletter Neuaufbau

## ✓ Architektur-zentrisch (Resilient)

### LLM-Schicht

Austauschbar: OpenAI, Anthropic, Azure, Google

### Orchestrierung

Swappable: n8n, Zapier, LangChain

### Speicher/Kontext

Datenbank, Vector Store, Session

12 Monate später: "Lass uns GPT-4 gegen Claude in der LLM-Schicht austauschen. Migration zu n8n in Orchestrierung."

**Resultat:** Modulare Upgrades, kein Neuaufbau

# Architektonisches Denken = Zukunftssicher

**Für Verwaltungsräte:** Keine spezifischen Tools in Strategie mandatieren, sondern architektonische Prinzipien: "Wir nutzen modulare, austauschbare Komponenten"

# Deterministic Automation vs. Probabilistic AI

Understanding the fundamental difference between algorithmic process automation and Large Language Models is critical for strategic technology decisions. While both play important roles in digital transformation, they serve distinctly different purposes and deliver fundamentally different types of outcomes.

- **Strategic Imperative:** For deterministic, auditable results in mission-critical processes, algorithmic automation is essential. LLMs deliver variable outputs and are unsuitable as the backbone of regulated business processes.

## Algorithmic Automation



### Deterministic Logic

Rule-based execution with guaranteed repeatability. Same input always produces identical output.



### Process Reliability

Ensures consistency, auditability, and compliance with regulatory requirements.



### Workflow Control

Structured process flows with defined decision points and clear governance.

**Use Case Example:** Invoice verification and payment processing — each invoice follows identical validation rules, ensuring consistent outcomes and full audit trails.

## LLMs & Generative AI



### Probabilistic Models

Statistical pattern recognition generating contextually relevant but variable responses.



### Language Understanding

Natural language processing enables intuitive interaction and content generation.



### Creative Flexibility

Generates diverse, context-appropriate outputs for complex communication tasks.

**Use Case Example:** Customer inquiry response drafting — AI generates helpful, contextual replies that vary based on nuanced understanding, requiring human review.

## Strategic Implications for Board Governance

### Compliance & Risk Management

Algorithmic automation provides the deterministic foundation required for regulatory compliance, financial controls, and risk mitigation. Auditors can trace every decision through transparent logic.

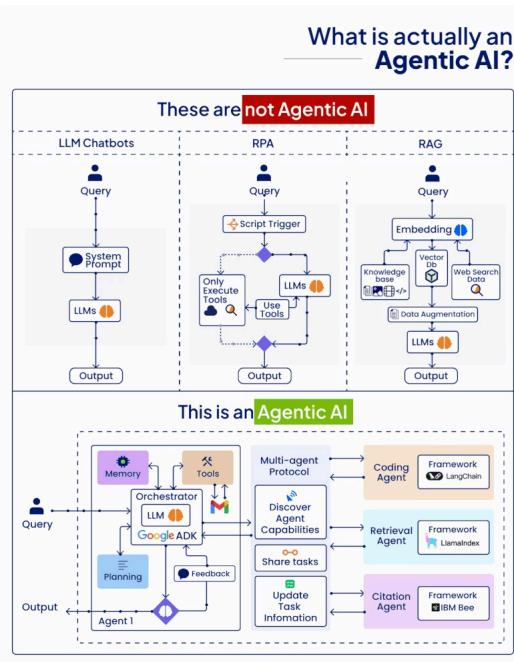
### Innovation & Customer Experience

LLMs excel as intelligent interfaces, enhancing user experience through natural interaction, content creation, and insights generation — but always with human oversight for critical decisions.

### Hybrid Architecture Strategy

Leading organizations combine both: deterministic automation for process backbone and compliance, with LLMs providing intelligent, user-friendly interaction layers and decision support tools.

The key takeaway for board members: recognize that these technologies are complementary, not interchangeable. Strategic technology investments must align the right tool with the right business need — deterministic automation for reliability and compliance, generative AI for flexibility and enhanced user engagement.



Not EVERYTHING powered by LLMs qualifies as an AI Agent !

Here's how to spot what's Agentic, and what isn't ...

There's major confusion between agent-like architectures and true Agents.

This creates challenges for beginners, as many users label almost every workflow as "Agentic" nowadays.

👉 To illustrate the confusion:

Last month I closed my course survey, and one of the most common questions was,

"What is actually an AI Agent?" or "How can I identify an AI Agent?"

I'll be teaching everything about real Agentic AI in my cohort,

But today, I'll simplify this doubt so beginners can start on the same level as everyone else.

👉 Let's clear it up:

1. **LLM Chatbots**
  - Think of them as advanced calculators for text. You ask a question → they answer.
  - Short-term memory of past queries, no planning, no dynamic tool use.
  - Great for customer support, but adaptability is limited to what they're trained on.
2. **RPA (Robotic Process Automation)**
  - Scripted bots that run predefined workflows, sometimes calling tools or LLMs, but always in a fixed sequence.
  - Works well for repetitive tasks like Email Automation, but fails when something unexpected appears.
3. **RAG (Retrieval-Augmented Generation)**
  - A smart reference system: it pulls information from a database or the web, then feeds it to an LLM.
  - Useful for fact-checking/knowledge retrieval, but simple RAG doesn't know how to plan multiple steps or adapt workflows.
  - It gives answers, not strategies about what to do ahead.

Cred: Rakesh

👉 How to spot what is Agentic AI

- It remembers context using short-term(for current tasks) and long-term(for scheduled tasks).
- It plans: breaking down a goal into smaller tasks, without even reasoning models, by using clever prompt engineering.
- It uses Tools and even schedules tools as per your prompt.
- It improves with feedback using approaches like ReACT, Reflexion(self-reflection), to improve the responses.
- It can work in multi-agent teams, where each agent specializes and collaborates.

👉 An awesome example is how Manus AI Works,

- When you give Manus a task, it plans the steps, selects tools, collaborates with other agents, executes them, and then reviews its own work.

One thing that makes agents unique is that they can also ask for your feedback during the workflow, something most linear LLM pipelines don't do.