

We Are Back!

AI/Copilot STAMMITSCH

November 25

AlfaPeople

www.alfapeople.com



Agenda

- About Us
- Demo
- Austausch



Sven Endres

Head of Delivery DACH



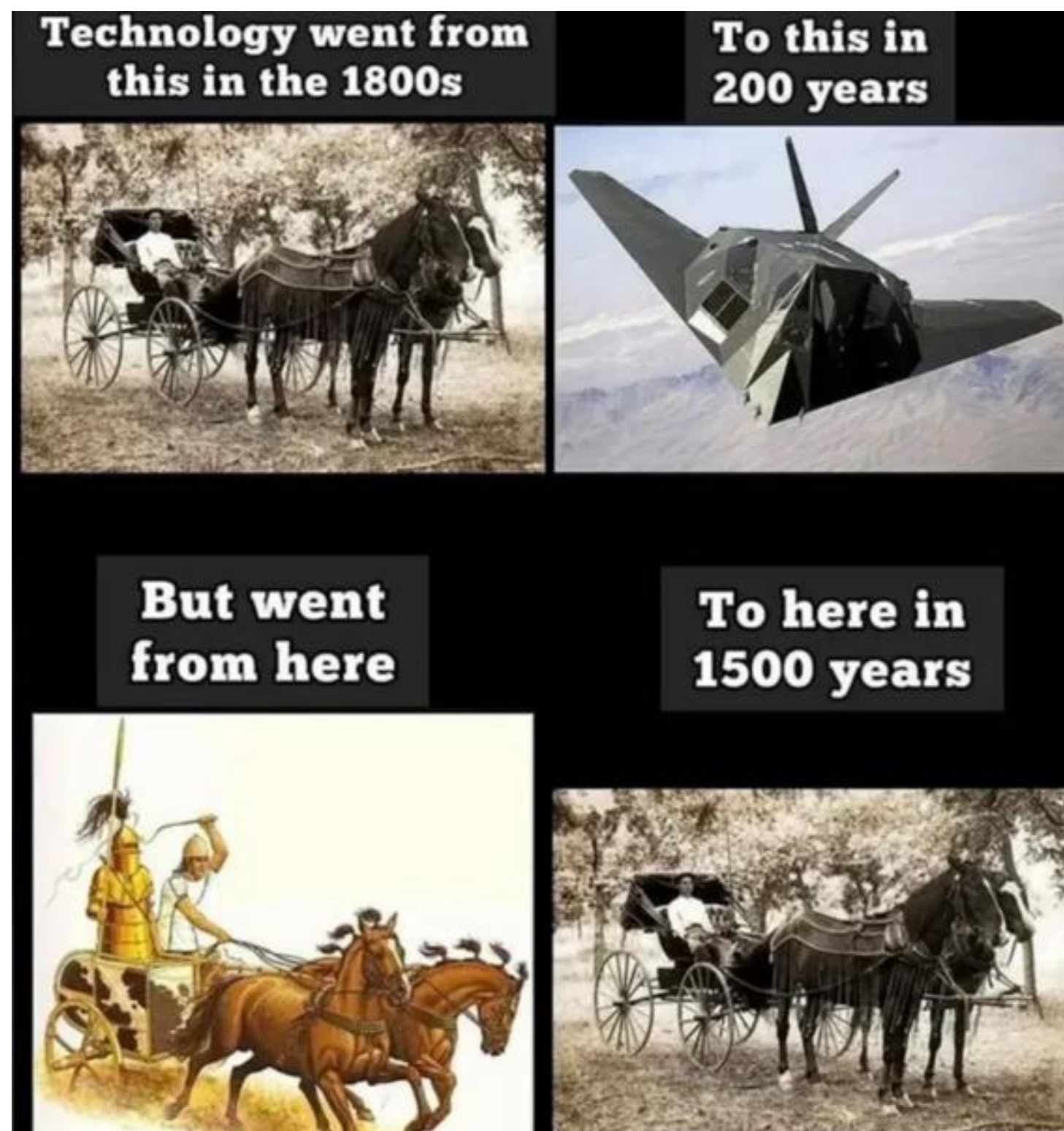
KI Grundlagen

Unleash your potential with **AlfaPeople**

KI und Geschwindigkeit

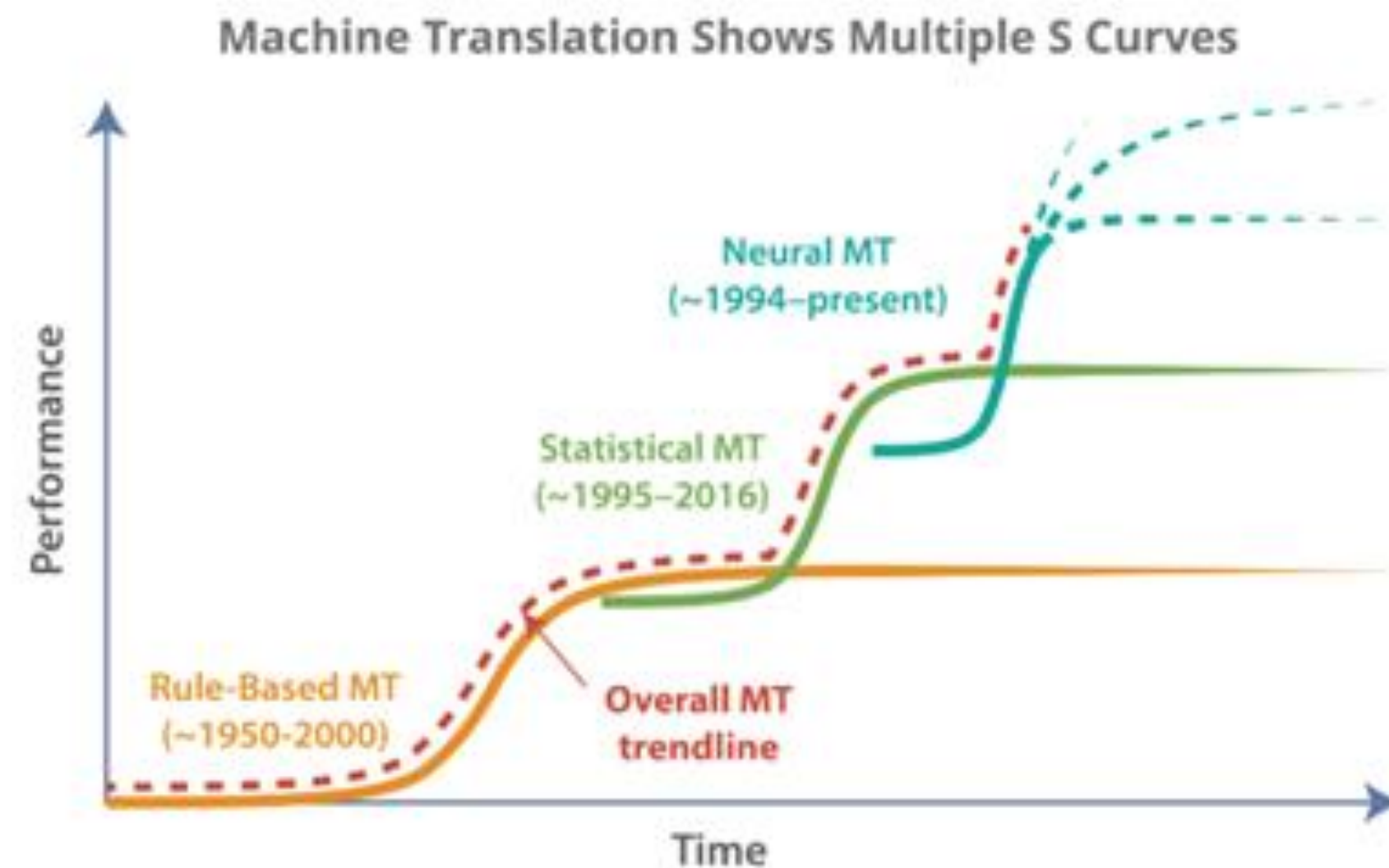


Die Geschichte der Technologie



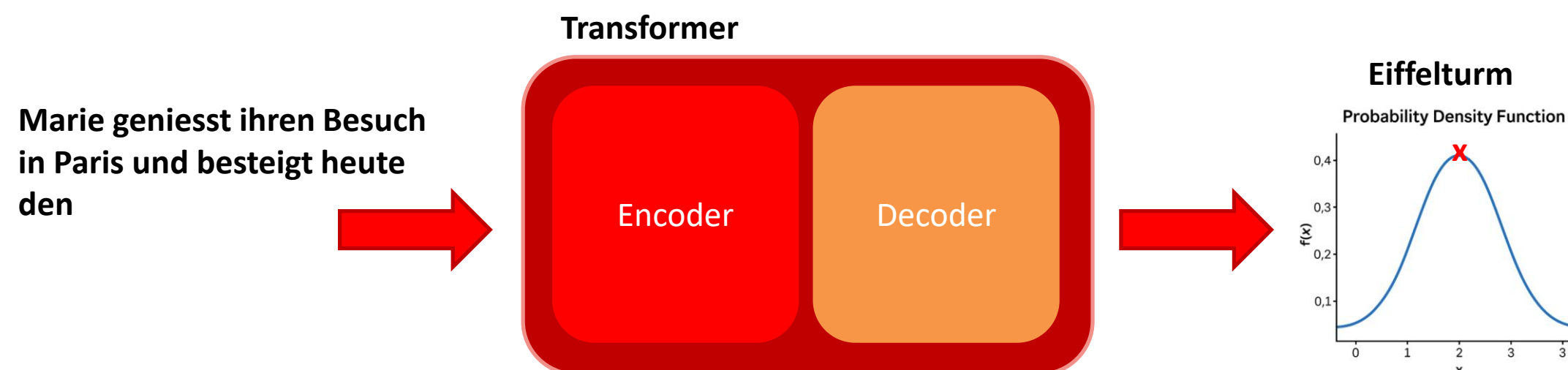
Die Geschichte der Technologie

Der Technologiefortschritt insgesamt wächst exponentiell. Einzeltechnologien wachsen erst langsam, dann schnell und flachen zuletzt wieder ab. Dies wird Wachstum in Form von “Interlocking S-Curves” genannt.



Wie ein LLM arbeitet

Ein Transformer sagt die Wahrscheinlichkeit jedes Worts im Wörterbuch vorher.



Dazu werden alle Wörter Tokenisiert.

Marie genießt ihren Besuch in Paris und besteigt heute den

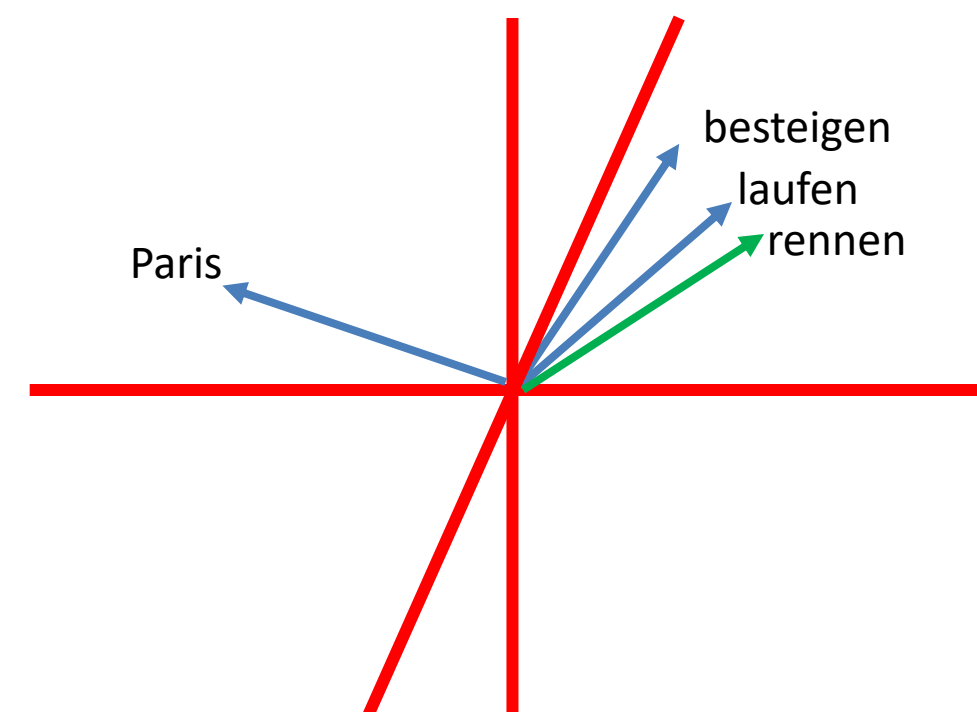
Wie ein LLM arbeitet

Alle Token werden dann in Vektoren umgewandelt.

Marie genießt ihren Besuch in Paris und besteigt heute den												
↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓
4	2	4	1	3	8	2	0	5	2	3	0	1
4	3	4	2	4	7	3	9	6	3	4	9	1
5	4	5	3	5	4	1	8	4	6	5	8	2
6	7	6	4	6	5	3	6	3	5	9	7	2
7	1	7	5	3	8	5	5	2	4	8	6	3
8	6	8	6	2	7	4	4	1	8	7	5	3
8	8	8	7	4	6	4	3	8	7	5	4	4

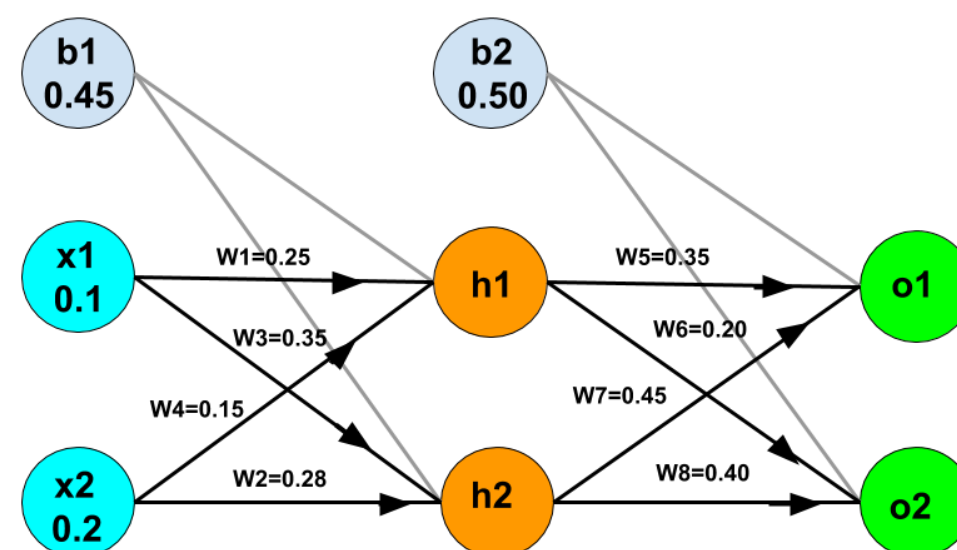
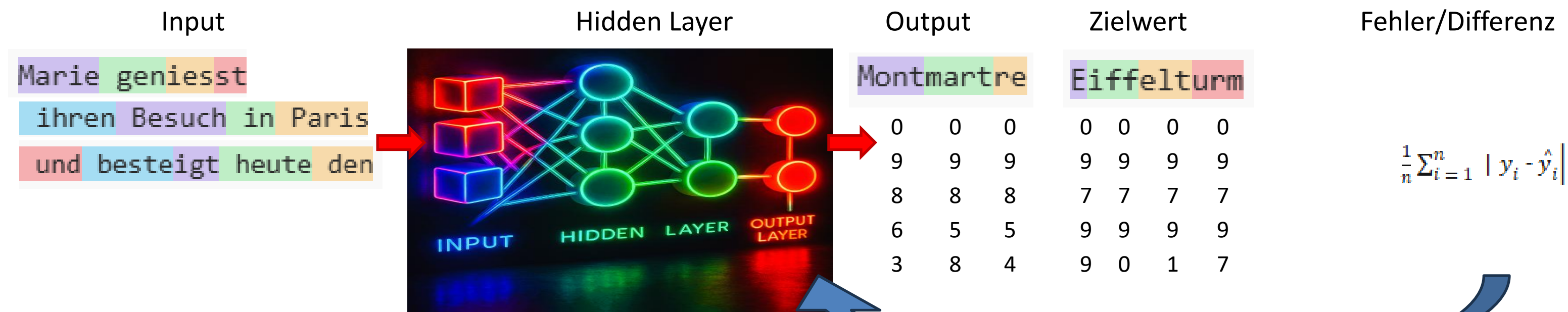
Vektoren werden in einer multidimensionalen Vektordatenbank gespeichert. Ähnliche Vektoren liegen dabei nahe beieinander.

Dies nennt man Vektor-Embedding.



Wie ein LLM arbeitet

Basis jedes LLMs ist ein neuronales Netzwerk. Dieses wird durch die Anpassung der Gewichte (weights) und Bias Werte optimiert um die Fehlfunktion zu minimieren.

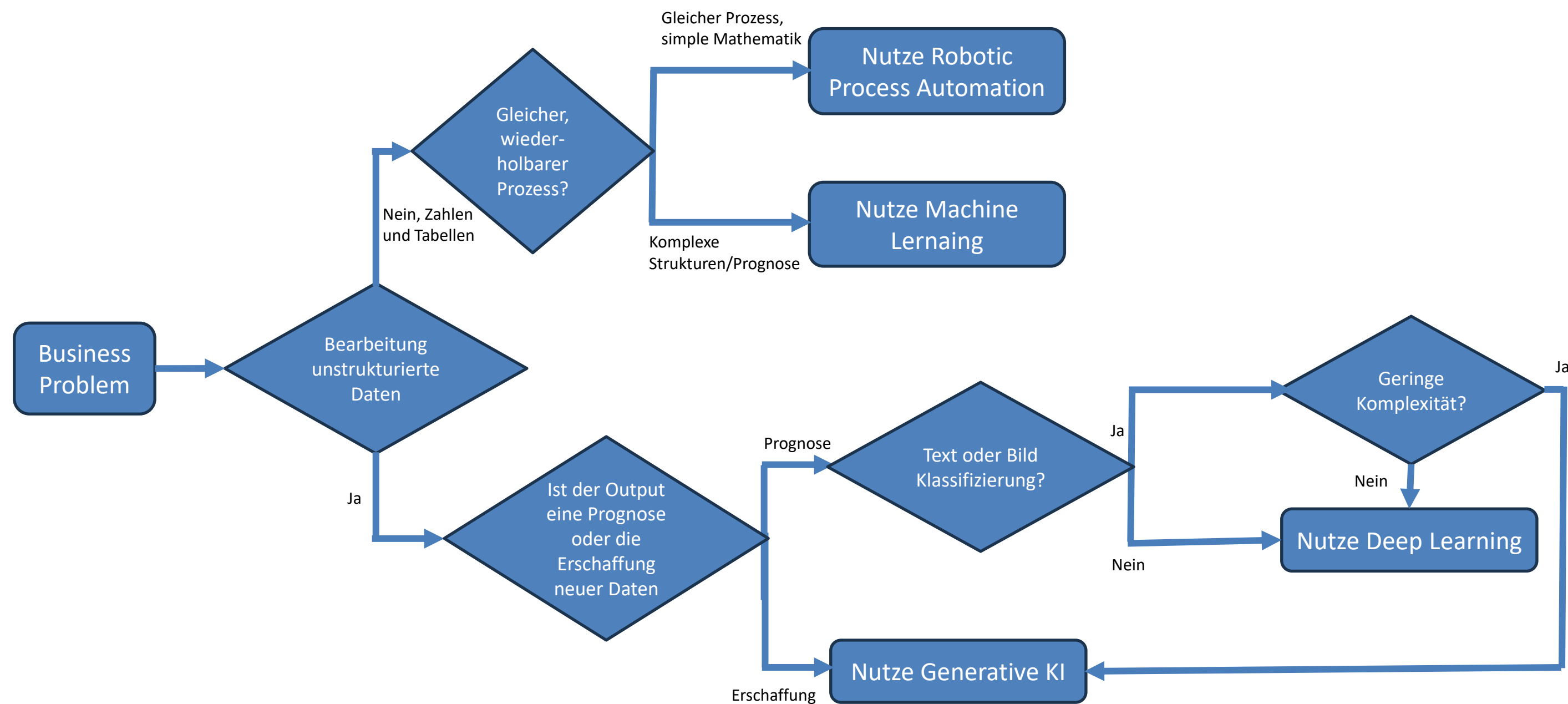


$$h1 = f(x1 * W1 + x2 * W4) + b1$$

$$h2 = f(x1 * W3 + x2 * W2) + b1$$

Was verwende ich wann?

Basis jedes LLMs ist ein neuronales Netzwerk. Dieses wird durch die Anpassung der Gewichte (weights) und Bias Werte optimiert um die Fehlfunktion zu minimieren.





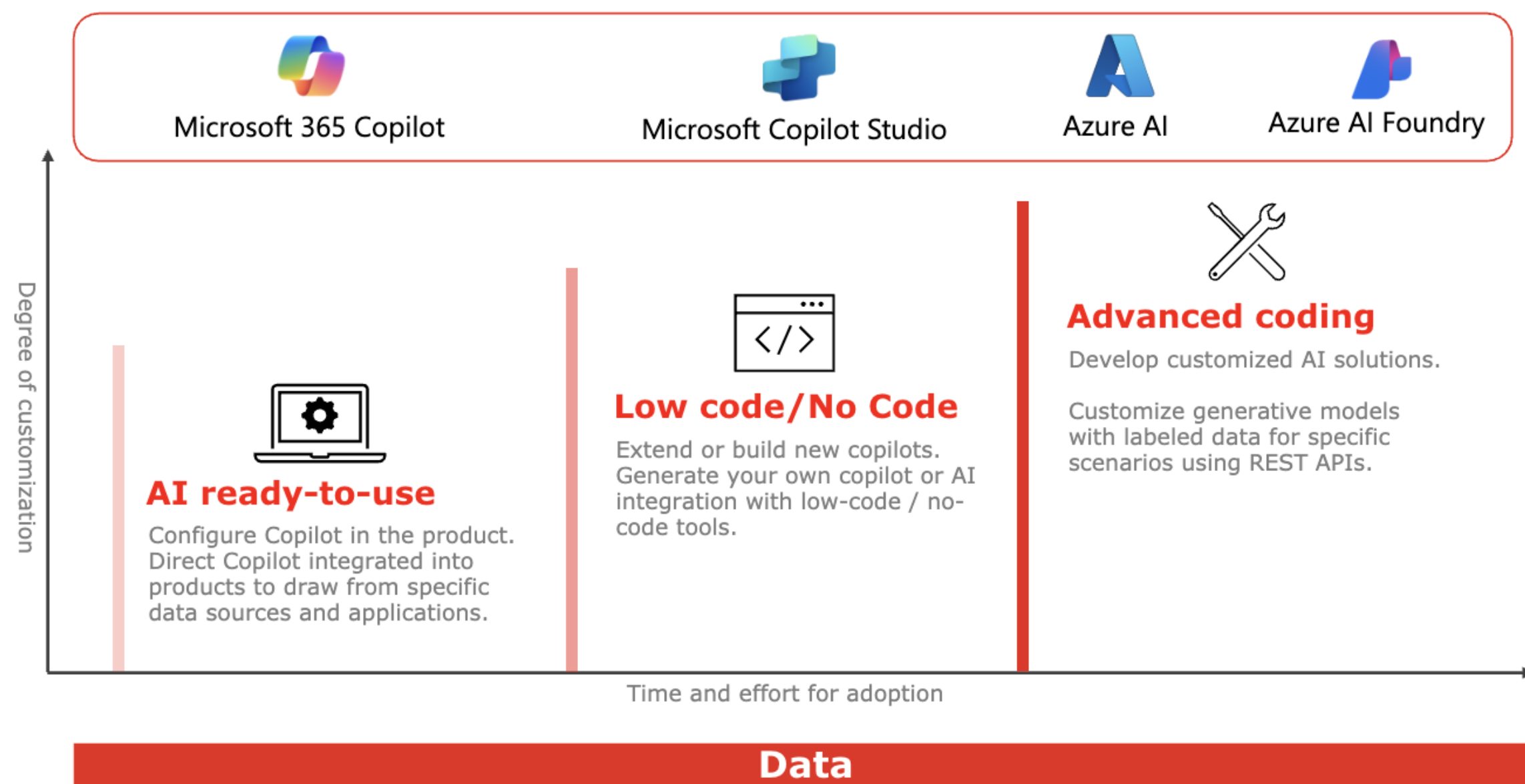
85%

der GenAI Lösungen stellen sich in der Produktivphase als nicht Funktional heraus (fullstack)

95%

der GenAI Pilotprojekte in Firmen schaffen es nicht einen guten fail ROI zu erwirtschaften (MIT)

Was tun wir bei AlfaPeople?



Wir müssen aufhören, KI als ein einzelnes Werkzeug zu betrachten, und anfangen, sie als ein Portfolio von Fähigkeiten zu sehen – einige generativ, einige analytisch und einige agentic.

Microsoft Copilot Studio Grenzen

- Vektor-Einbettungen können nicht mit Finetuneing bearbeitet oder konfiguriert werden.
- Kein Zugriff auf die zugrunde liegende Indexstruktur.
- Semantic Chunking-Strategien können nicht angewendet werden.
- Die Parameter für die Vektorsuche sind nicht konfigurierbar.

Sie können einen SharePoint- oder OneDrive-Ordner anbinden und Antworten erhalten, haben jedoch keinerlei Kontrolle darüber, wie das Wissen embedded, retrieved oder bewertet wird.

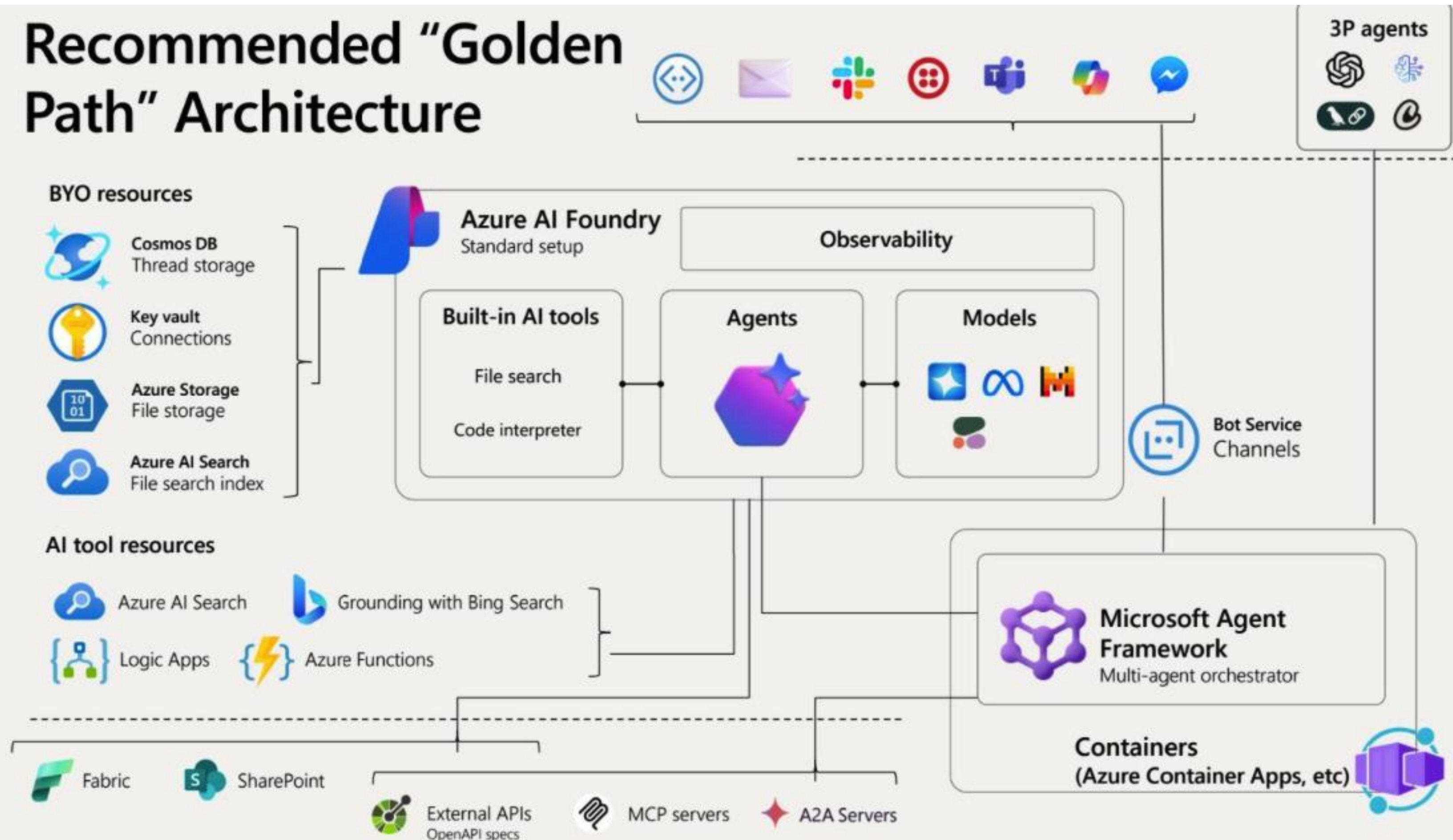
- Sie können keine mehrstufigen Reasoning-Chains erstellen

Copilot Studio eignet sich hervorragend für FAQs und Einfachlogik, kann jedoch keine Schlussfolgerungen aus sich entwickelnden Kontexten ziehen oder mehrere KI-Schritte koordinieren.

Voll autonome vs Konversations KI Agenten

Aspect	Fully autonomous AI Agents	Conversational AI Agents
Autonomiegrad	Hoch: Sie arbeiten unabhängig, können planen, ausführen und optimieren ohne menschliche Prompts.	Moderat: Reaktiv. Sie benötigen Interaktion und vordefinierte Dialogabläufe.
Komplexitätsbewältigung	Sie verwalten in der Regel komplexe, mehrstufige Aufgaben wie Content-Platzierung und Bestandsmanagement durch Koordination von Unteragenten.	Sie bearbeiten strukturierte, routinemäßige Interaktionen wie FAQs und grundlegenden Support mit begrenzter adaptiver Logik.
Integration	Tiefe Integration mit Unternehmenssystemen einschließlich Data Lakes, APIs und DaaS-Plattformen zur Unterstützung autonomer Workflows.	Hauptsächlich Integration mit Kommunikationsschnittstellen und ausgewählten Backend-Systemen wie CRM- und E-Commerce-Plattformen.

Recommended "Golden Path" Architecture



Von der Automatisierung zur Intelligenz

- Copilot Studio demokratisiert den Zugang zu KI, aber echte Transformation beginnt erst, wenn wir die Black Box öffnen.
- Mit Azure AI Foundry und Data Agents können wir Architekturen entwerfen, in denen Intelligenz erklärbar, kontrollierbar und entwicklungsfähig ist.

Wir brauchen keine weiteren Bots, sondern intelligentere Systeme, die Kontext, Absicht und Zweck verstehen.



Austausch





Vielen Dank für Ihre Teilnahme!

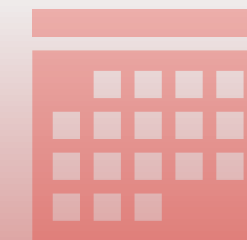
Wir freuen uns darauf, Sie auch bei einem unserer nächsten AI/Copilot Stammtische im 2026 begrüßen zu dürfen.

Wir werden die neuen Termine frühzeitig kommunizieren.



<https://alfapeople.com/ch/webinars/>

Stay tuned!



➡ 2026 geht's weiter ⬅