



Vertrauenswürdige Technologie und der Mensch

Spannungsfeld und Lösungsansätze

Impulsvortrag, 03.03.2026

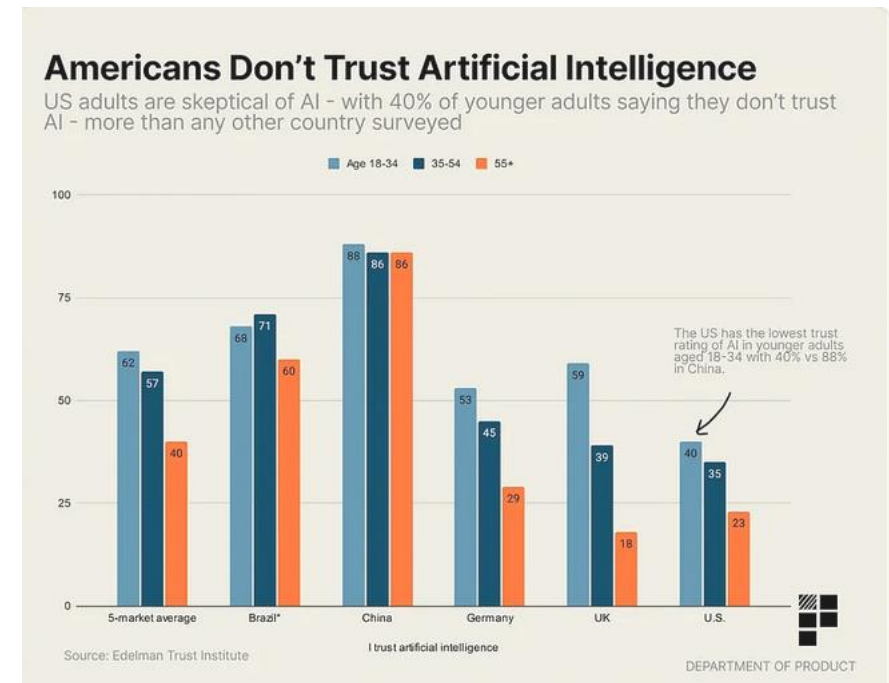
Dr. Gertraud Leimüller, MPA (Harvard)
CEO winnovation & leiwand.ai



Vertrauen in Technologie wächst nicht mit zunehmender Vertrautheit

Anti-KI Demonstration
am 28.02.2026
in London

Nutzungs-Vertrauens-Paradoxon:
Obwohl die Nutzung von KI-Systemen steigt, sinkt das Vertrauen in KI (insbesondere in entwickelten Ländern).



Warum gibt es dieses Misstrauen? Einige wichtige Faktoren



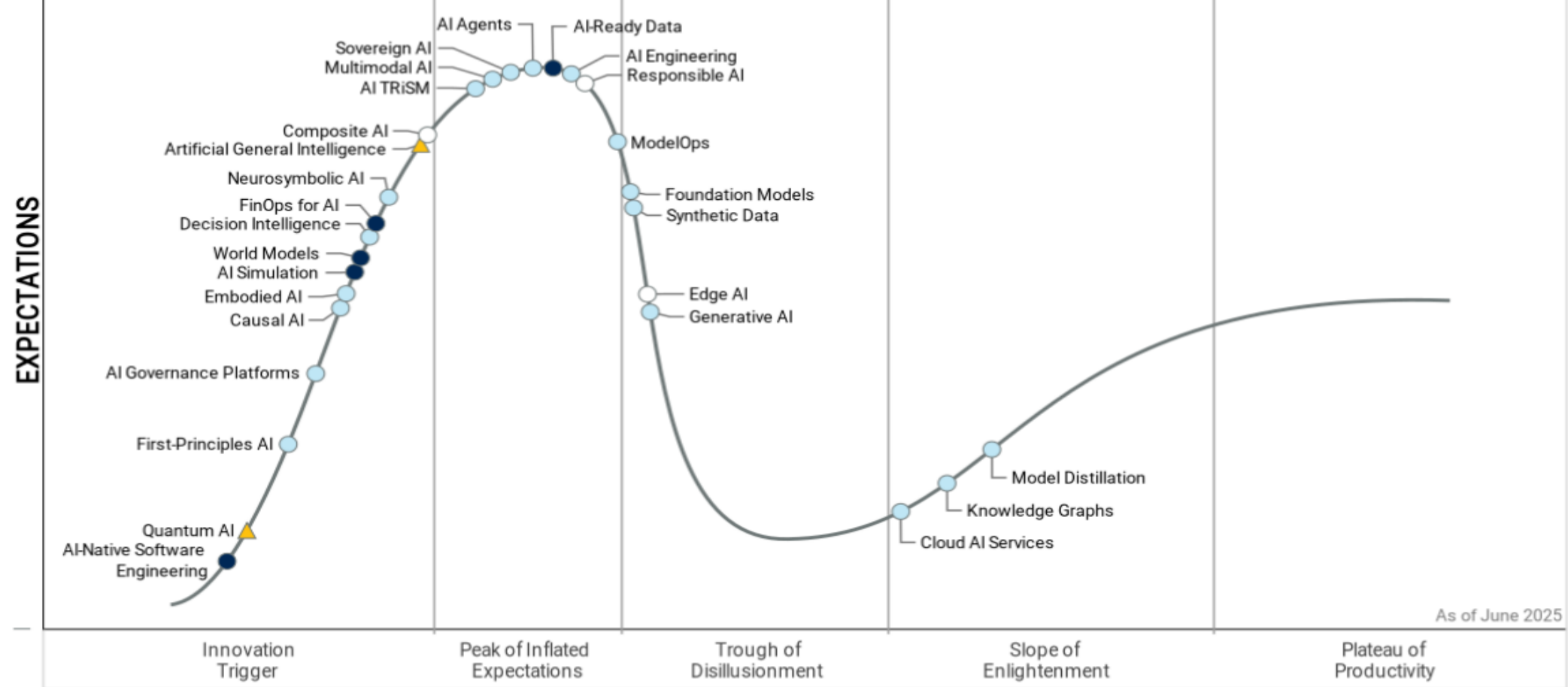
1. Überzogene Erwartungen

Narrative für KI

- Eine Superintelligenz, die zu unglaublichem Wachstum führt, die Wissenschaft beschleunigt und den Menschen unsterblich machen kann; wichtiger Begriff: AGI
- Eine Superintelligenz, die zu großen Umbrüchen und Jobverlusten führt und sich eines Tages vielleicht verselbständigt und die Menschheit ausrotten wird
- Eine überschätzte Technologie, die sich nicht durchsetzen und wieder verschwinden wird

1. Überzogene Erwartungen

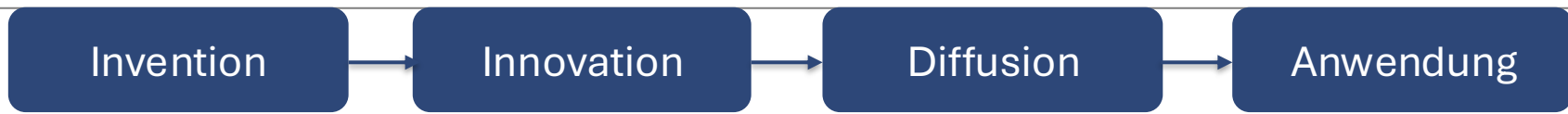
Hype Cycle für Künstliche Intelligenz, 2025 (Gartner)



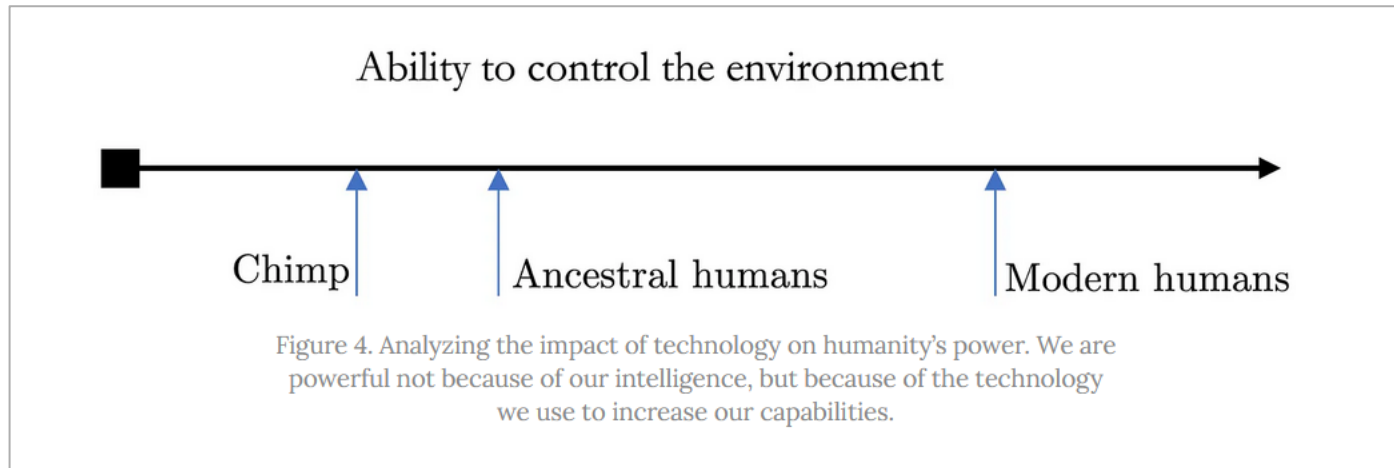
As of June 2025

<https://www.pasqa.com/resources/new-gartner-hype-cycle-for-ai-report-2025/>

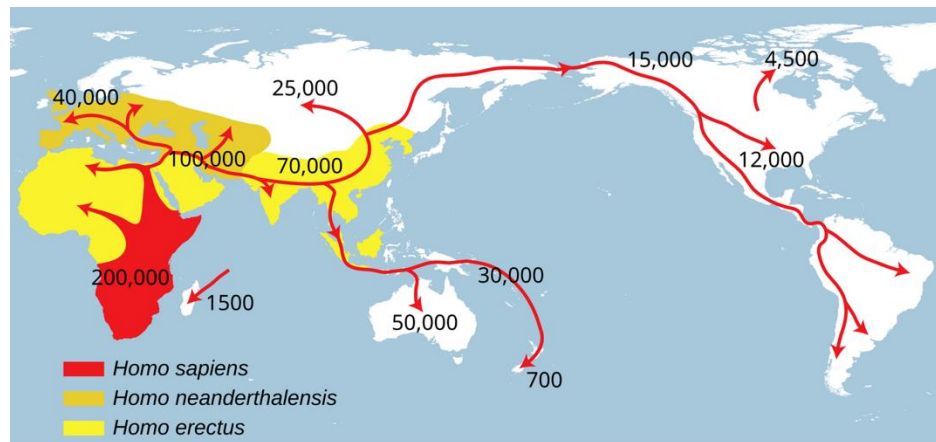
Plateau will be reached: ○ <2 yrs. ● 2-5 yrs. ● 5-10 yrs. ▲ >10 yrs. ✗ Obsolete before plateau



2. Empfundene Ohnmacht



Quelle: [Arvind Narayanan and Sayash Kapoor, https://www.normaltech.ai](https://www.normaltech.ai)



**Technologie ist ein
Machtinstrument**
Zugang entscheidet über
Entwicklungs-
möglichkeiten

Geschlossene Natur vieler
Entwicklungen mit
geringer Offenheit

3. Wirkung von neuen Technologien zeigt sich erst mit zunehmender Anwendung

Beispiele:

- Systematischer AI Bias benachteiligt viele Menschen: z.B. Frauen, Ältere, Migrant:innen
- Negative Effekte von „kognitive Offloading“ mit AI: kritisches Denken nimmt ab
- Verzerrungen am Arbeitsmarkt



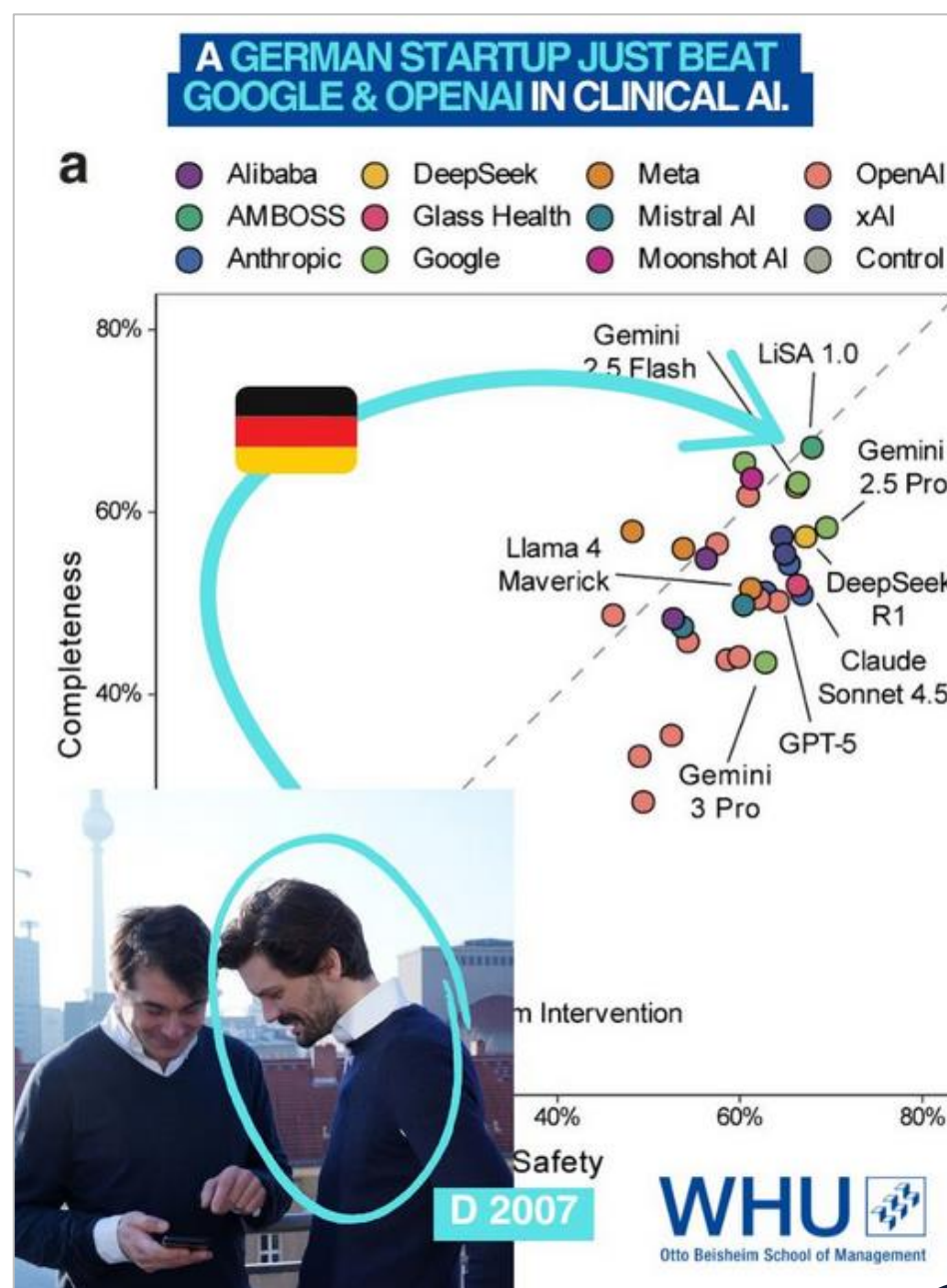
Lösungsansätze

Effektive Gestaltungsmacht anstreben

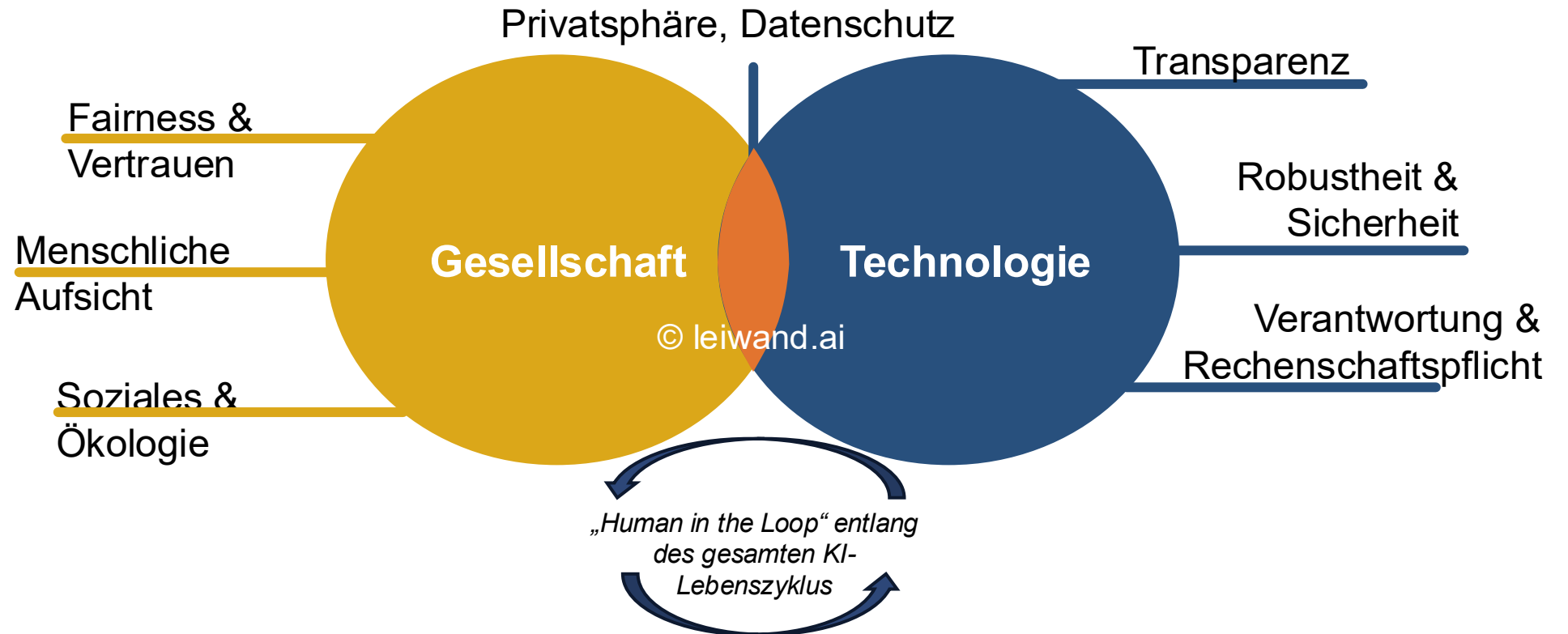
Digitale Souveränität in Europa

- **Europäische Cloud:** Souveräne Cloud Alternativen (OVH, Scaleway, Hetzner, IONOS)
- **EU AI GRID:** private Unternehmen, welche unabhängige, europäische KI-Infrastrukturen zur Verfügung stellen
- **Gaia-X:** Europäische Cloud Infrastructure Zusammenarbeit
- **Digital Europe Programme:** € 7.5 Mrd. Investment in souveräne digitale Infrastruktur
- **EU Chips Act:** Halbleiter-Abhängigkeit reduzieren
- **5G Souveränitäts-Initiativen:** Europäische Alternativen für ausländische Netzwerk-Ausrüster

Beispiel für KI-Anwendung aus Europa: AMBOSS



Nachfrage nach vertrauenswürdigen Lösungen generieren



→ Qualitätsnormen, branchenspezifische Anwendungs-Guidelines

KI-Literacy in der breiten Bevölkerung erhöhen

Beispiel Faktentreue von LLM

Table 7: *SimpleQA Verified* results across the key metrics also computed in Wei et al. (2024a). All results are reported in percent (%) unless stated otherwise.

| Rank | Model | F1-Score | Δ SimpleQA (%pt) | Accuracy | Acc./Attempted | Attempted | Hedged |
|------|-----------------------|-------------|-------------------------|----------|----------------|-----------|--------|
| 1 | Gemini 2.5 Pro | 55.6 | 0.5 | 55.3 | 55.9 | 98.9 | 1.1 |
| 2 | GPT 5 | 52.3 | 1.8 | 50.9 | 53.8 | 94.6 | 5.4 |
| 3 | o3 | 51.9 | 1.9 | 51.6 | 52.0 | 99.3 | 0.7 |
| 4 | GPT 4.1 | 39.9 | -1.0 | 39.8 | 40.1 | 99.3 | 0.7 |
| 5 | GPT 4o | 34.9 | -3.5* | 34.4 | 35.5 | 97.0 | 3.0 |
| 6 | DeepSeek R1 (0528) | 33.3 | 1.1 | 32.7 | 33.9 | 96.4 | 3.6 |
| 7 | Claude Opus 4 | 28.3 | -4.0* | 19.2 | 54.1 | 35.5 | 64.5 |
| 8 | Gemini 2.5 Flash | 28.2 | -1.4 | 27.8 | 28.7 | 96.9 | 3.1 |
| 9 | GPT 5 Mini | 24.6 | 1.1 | 17.3 | 42.8 | 40.4 | 59.6 |
| 10 | o4-mini | 23.4 | 2.9* | 23.0 | 23.8 | 96.5 | 3.5 |
| 11 | Claude Sonnet 4 | 18.7 | -4.4* | 12.5 | 36.9 | 33.9 | 66.1 |
| 12 | GPT 5 Nano | 14.4 | 0.7 | 10.2 | 24.2 | 42.2 | 57.8 |
| 13 | Gemini 2.5 Flash Lite | 11.1 | -0.4 | 10.2 | 12.1 | 84.0 | 16.0 |

Notes: * $p < 0.05$; Δ SimpleQA is the F1-Score of SimpleQA Verified minus the F1-Score of SimpleQA. Both SimpleQA Verified and SimpleQA results use the same prompted autorater (gpt-4.1-2025-04-14).

Sogar die besten LLM können nur jede 2. Wissensfrage richtig beantworten

Danke für die Aufmerksamkeit!

Kontakt:
Gertraud.Leimueller@leiwand.ai
www.leiwand.ai

