

KI - was bleibt nach dem Hype? Status und Ausblick

Diskussion

19. APRIL 2024

Midjourney | "Generative AI"



Was wir heute mit Ihnen besprechen möchten

1 Was ist KI?



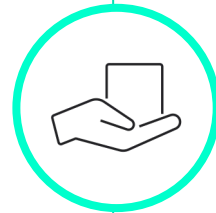
2 Die GenAI-Revolution



3 Die Herausforderung der Umsetzung

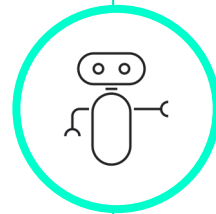


Moravecs
Paradoxon: Was
für Menschen
einfach ist, ist
für KI schwierig
und umgekehrt



Ein 17-Jähriger kann in 20 Fahrstunden lernen, ein Auto zu steuern

Wir haben immer noch kein unbegrenztes autonomes Fahren der Stufe 5!



Ein 10-Jähriger kann lernen, den Esstisch abzuräumen und die Spülmaschine in wenigen Minuten zu füllen (One-Shot-Example)

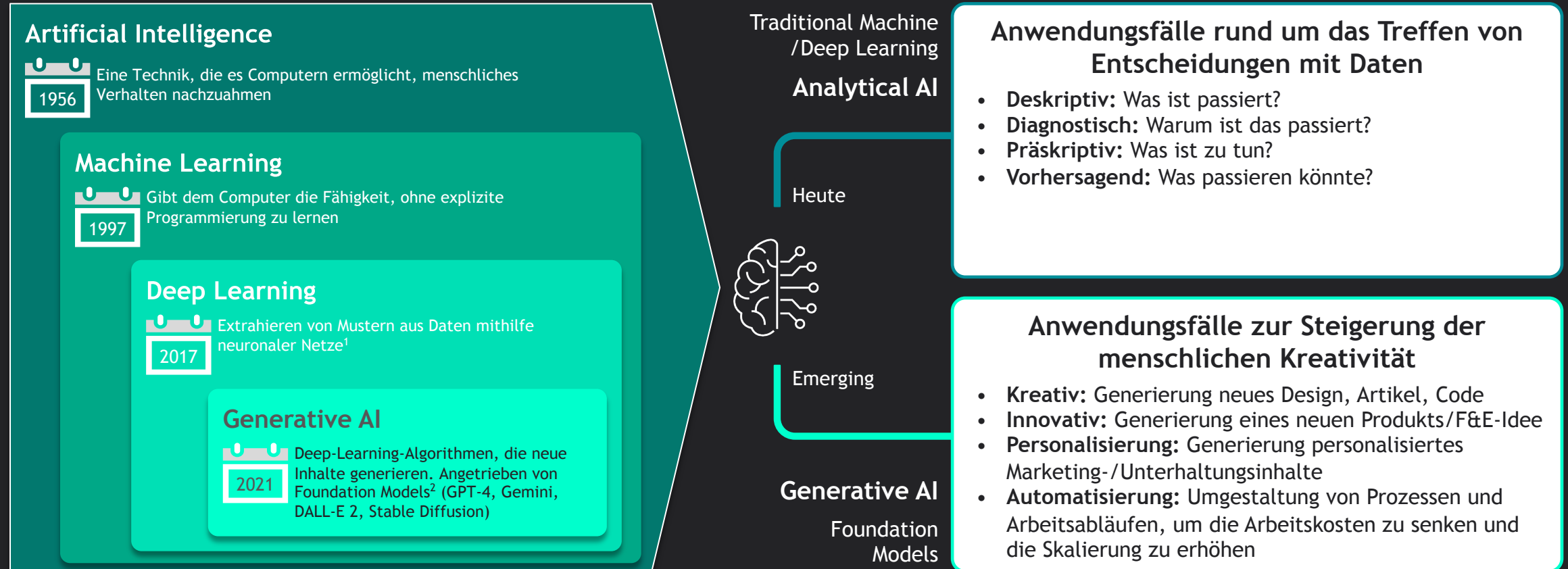
Wir haben derzeit keine Roboter, die dazu in der Lage sind!



Ein 4-Jähriger hat mit Erreichen dieses Alters in 300 Stunden "Training" 50-mal mehr Daten verarbeitet als ein LLM, das mit $1,0E13$ -Token ($0,75E13$ Wörter) trainiert wurde

Ein Mensch bräuchte 170.000 Jahre, um den LLM-Trainingsdatensatz durchzulesen!

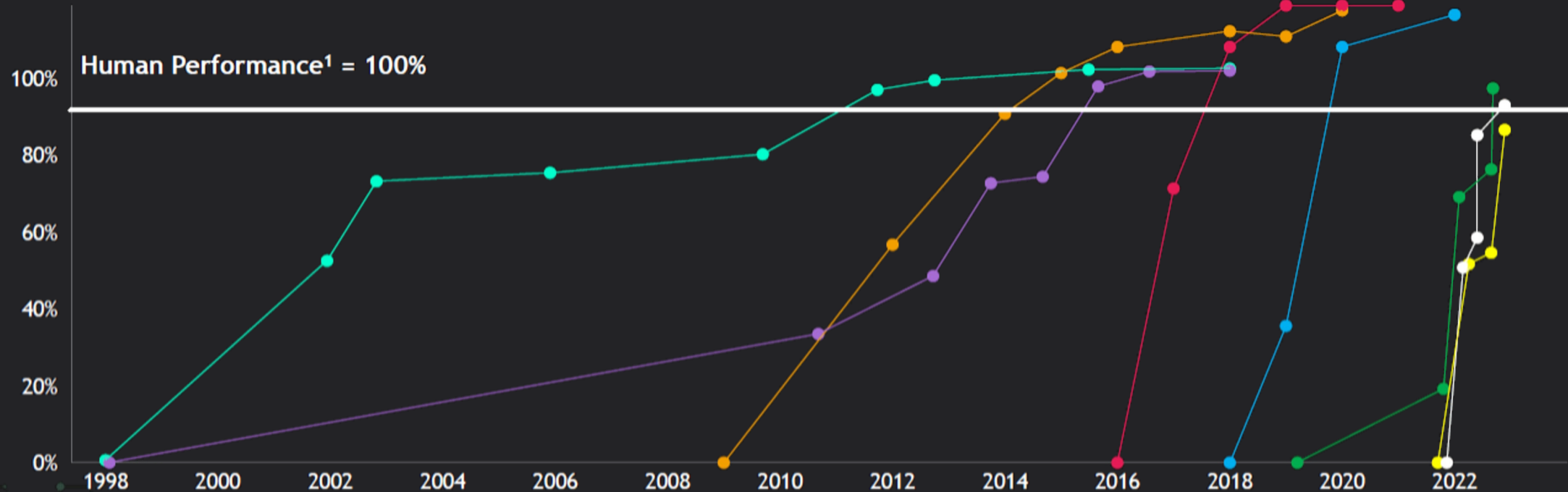
GenAI erweitert das bereits bestehende Anwendungsspektrum der analytischen KI



Quelle: Corporate Strategy . Note: (1) Neuronale Netze sind Computersysteme, die dem menschlichen Gehirn nachempfunden sind und darauf ausgelegt sind, Muster zu erkennen, Daten zu interpretieren und Entscheidungen zu treffen. (2) Ein Foundation Model ist ein fortschrittliches neuronales KI-Netzwerk, das auf Rohdaten trainiert wird, in der Regel mit unüberwachtem Lernen, und das angepasst werden kann, um eine breite Palette von Aufgaben zu erfüllen.

Die KI-Fähigkeiten werden weiter ausgebaut - und das in immer schnellerem Tempo

State-of-the-art AI performance on benchmarks, relative to human performance



- Handwriting recognition
- Image recognition
- Speech recognition
- Code generation
- Common Sense Completion
- Reading comprehension
- Grade school math
- Language understanding

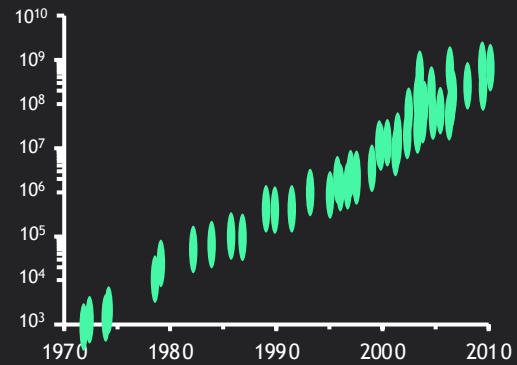
¹: Human performance defined as average performance of medium skilled worker

Note: Handwriting recognition = MINIST, Language understanding = GLUE, Image recognition = ImageNet, Reading comprehension = SQuAD 1.1 and 2.0, Speech recognition = Switchboard, Grade school math = GSK8k, Common sense completion = HellaSwag, Code generation = HumanEval

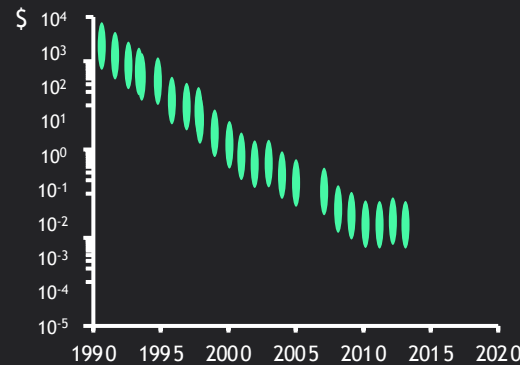
Source: Contextual AI; TIME

Die Beschleunigung der KI wurde durch anhaltende Fortschritte in der Hardware (einschließlich des Mooreschen Gesetzes) ermöglicht

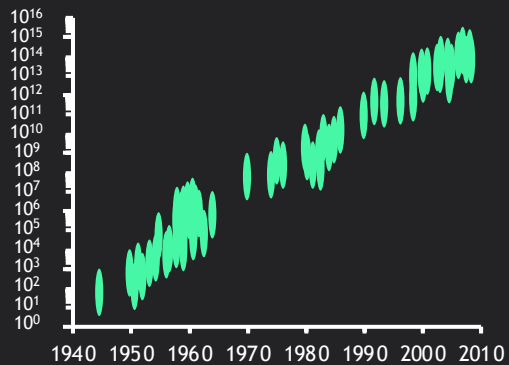
Transistors/chip



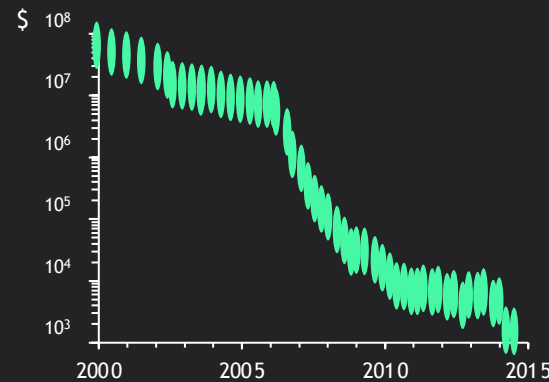
Storage cost/byte



Computes /KwH



Sequencing cost/genome



NVIDIA Marktkapitalisierung

+ 240 % im letzten Jahr und ein Wert von 2,2 Billionen US-Dollar (Stand Mitte April)



KI hatte bereits einen noch nie dagewesenen Automatisierungsgrad möglich gemacht – noch vor dem Aufkommen generativer Modelle



- Einer der weltweit größten Hersteller von Industrierobotern mit einer Produktion von 8.000 Einheiten pro Monat
- Seit 2001 betreibt Fanuc 22 teilautonome Werke zu 80 %
- “Lights-Out” mit bis zu 600 Stunden ununterbrochen mit menschlichem Eingreifen nur für Wartungs- und technische Problemlösungen



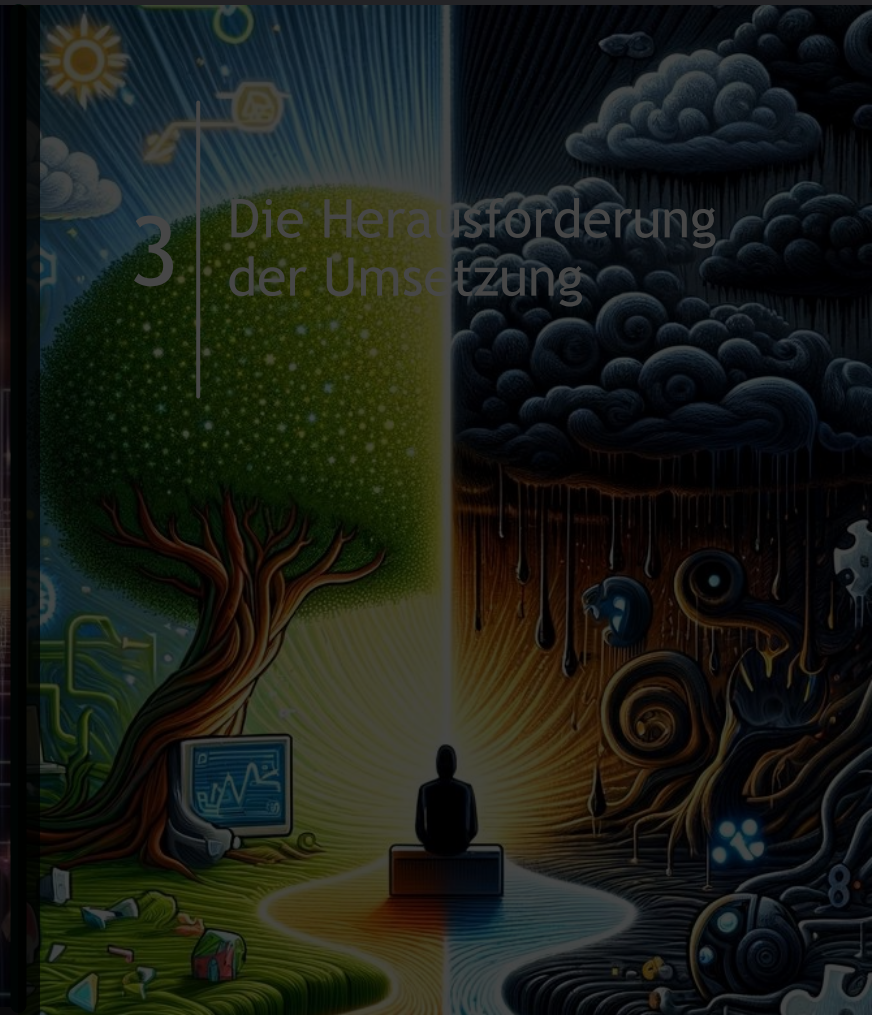
- Der siebtgrößte Hafen der Welt
- Das autonome Terminal (2021 eingeweiht) verfügt über 25 % des Personals, das in den anderen Terminals benötigt wird
- Das intelligente Planungssystem reduziert die Be- und Entladezeit von 8 Stunden auf 40 Minuten, wobei 76 Roboter die Last mit einer menschlichen Eingriffsrate von $<0,1\%$ bewegen



- Alibaba Bank mit 45 Mio. KMU-Kunden im Jahr 2021, 80 % davon mit ihrem ersten Kredit
- 55% der Mitarbeiter arbeiten im Technologiebereich. Es gibt keine Risikoanalysten oder Kreditsachbearbeiter
- Kreditgenehmigung in 1 Minute mit mehr als 100 Risikomodellen zu 0,1 % der Kosten des traditionellen Bankwesens und mit einer notleidenden Rate von 1,9 %



Was wir heute mit Ihnen besprechen möchten



Large Language Models sind nicht der Endzustand: Wir stehen vor einem Jahrzehnt der "permanenten Revolution"

Large Language Models
Möglichkeit, mehrere Aufgaben im Textformat zu erledigen

Multimodal Foundation Models
Möglichkeit, mehrere Aufgaben in verschiedenen Formaten zu erledigen (Text, Bilder, Videos, PDFs, Excel, IoT/Sensoren usw.)

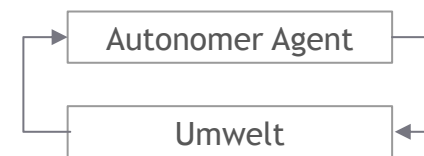
Autonomous Agents
Möglichkeit, Aktionen auszuführen, wie z. B. das Senden direkter Anweisungen an andere Systeme

“
With stand-alone large language models, you have access to a powerful brain; autonomous agents add arms and legs.

Human-level Intelligence
Fähigkeit, ein umfassendes Verständnis der Welt zu entwickeln, eine breite Palette von kognitiven Aufgaben auszuführen und fortgeschrittenes Denken zu entwickeln
Interaktionen mit der digitalen Welt, vermittelt durch KI-Assistenten über intelligente Schnittstellen (z.B. Sprache, Vision, Display)

Intelligente Assistenzen (beantwortet Fragen, hilft bei täglichen Aktivitäten, versteht Geschäftslogiken, Vorlieben und Interessen)
Neuartige KI-Modelle und Betriebssysteme

Time



Heute transformiert GenAI Schlüsselfunktionen innerhalb von Organisationen



Marketing



Kundenbetreuung



Unterstützende
Funktionen



Software
Entwicklung

Beispiel

Hyper-personalisierte Marketing-Materialien über alle Medien hinweg (z. B. Text, Bild, Video)

Kundenbetreuer Unterstützung mit einem Gen AI-basierten **Konversations-Assistent**

Überprüfen und Markup von Verträgen und Verhandlungsunterstützung

Entwicklung von Code durch textbasierte Aufgabenbeschreibung und **Korrekturvorschläge** für existierenden Code

Benefit

5-10x schnellere Generierung von Inhalten¹

14% Steigerung der Produktivität und Effektivität der Betreuer²

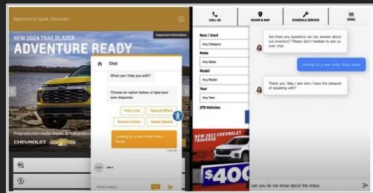
70% Produktivitätssteigerung für die Rechtsabteilung³

55% Steigerung der Produktivität⁵

Unternehmen testen GenAI-Lösungen primär für Text- und Code Generierung

fullpath

Nutzung von ChatGPT, um Käufer bei ihrer Suche und Händler bei der Verwaltung ihrer Kunden zu unterstützen



500+

North American Dealership

SIEMENS

Verwendung von GenAI zur Erweiterung der Erstellung, Optimierung und des Debuggings von Code in der Fabrikautomatisierung

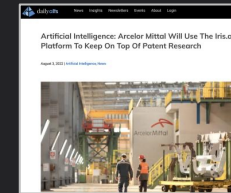


100%

Auswirkungen auf den gesamten Produktentwicklungslebenszyklus



Automatisierte Patentrecherche und -synthese durch Aufnahme und Zusammenfassung veröffentlichter Arbeiten



94%

Präzision der Extraktion

3M

Abfragen in natürlicher Sprache, um Informationen schnell zu finden und Forschung und Entwicklung zu beschleunigen



1200+

Neue Produkte erstellen

Activity4

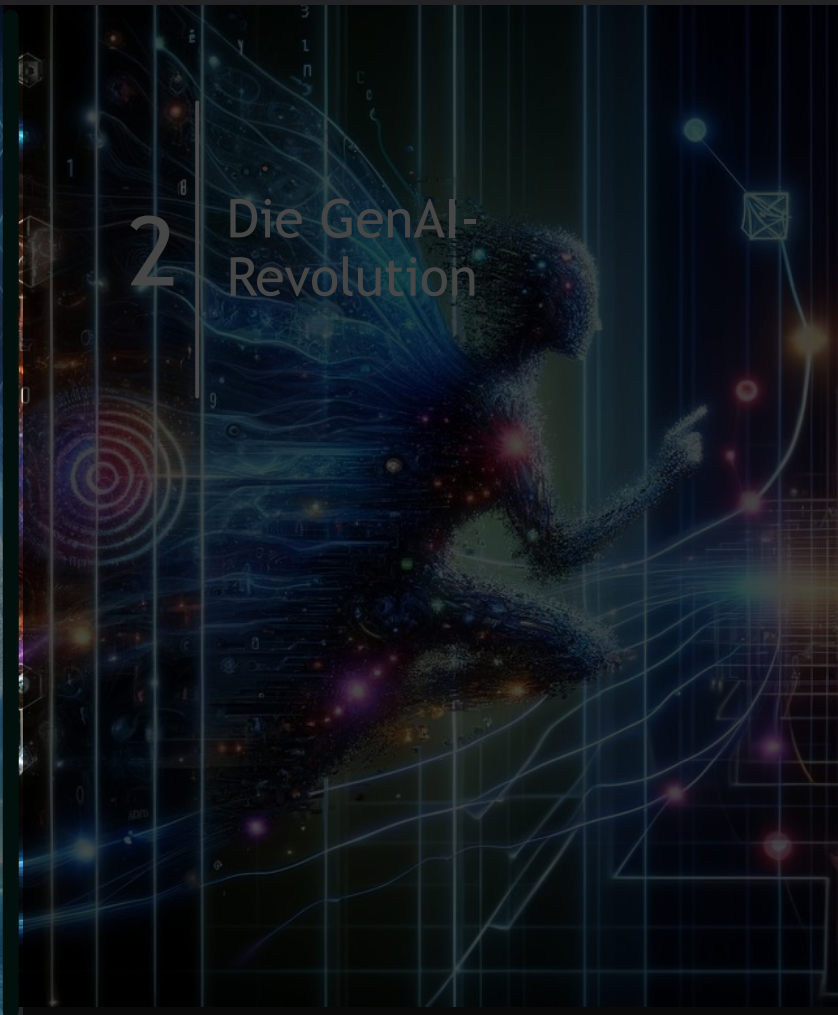


Question Answering



Content Generation

Was wir heute mit Ihnen besprechen möchten



GenAI erweist sich als Game Changer, aber der volle makroökonomische Effekt wird Zeit brauchen

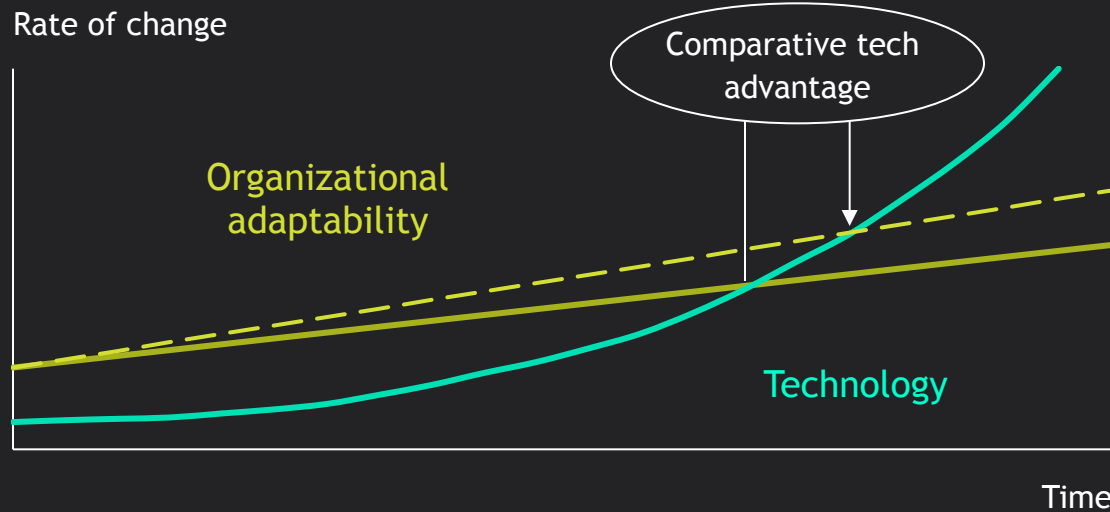
Echte und nachgewiesene Produktivitätssteigerungen

- 30%** Kostensenkung für den Kundenservice durch InsureTech-Plattform durch Nutzung von GenAI⁴
- 55%** Schnellere Erledigung von Codierungsaufgaben mit höheren Erfolgsquoten durch GitHub CoPilot⁵
- 2 Monate** to 100 Millionen Nutzer für OpenAI ChatGPT, die schnellste Produktannahme aller Zeiten²
- 20+ Mrd. \$** Of VC-Finanzierung für GenAI in den letzten drei Jahren¹

- Die vollen makroökonomischen Gewinne blieben oft Jahrzehnte hinter der Einführung von Allzwecktechnologien zurück
- “You can see the computer age everywhere but in the productivity statistics” (Robert Solow, 1987)
- Einige Berufsgruppen, die es heute gibt, werden verschwinden, aber neue Berufsgruppen werden entstehen
- Die GenAI-Transformation wird kommen, aber der Übergang wird länger dauern, als viele heute erwarten - heute haben nur ~25% der führenden US-Unternehmen traditionelle KI angewendet

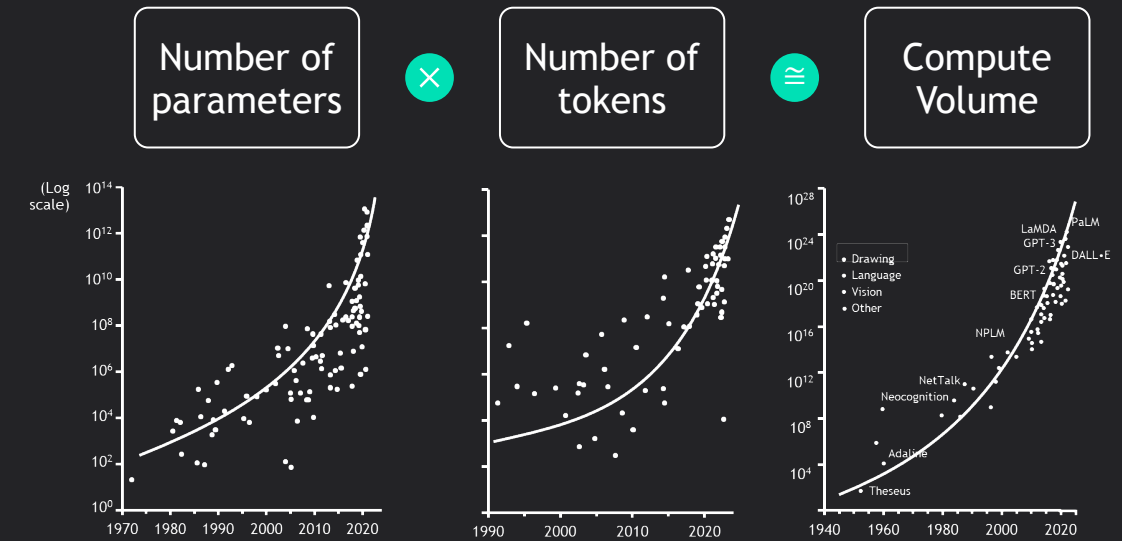
Warum die Einführung von GenAI für die meisten Unternehmen wahrscheinlich langsam voranschreiten wird

Nachfrage: Die Herausforderung der Innovationsrate



Während sich die Technologie exponentiell entwickelt, ist die Akzeptanzrate linear und beschränkt sich auf die Anpassungsfähigkeit des Unternehmens und die äußeren Umstände

Angebot: Die Kosten der Inferenz



Das Rechenvolumen und die Kosten steigen exponentiell mit Anzahl der Parameter und Token eines GenAI-Modells an. Die breite Einführung im Unternehmen verlangsamt sich ohne signifikante Kostendegression

Risiken und Einschränkungen ...

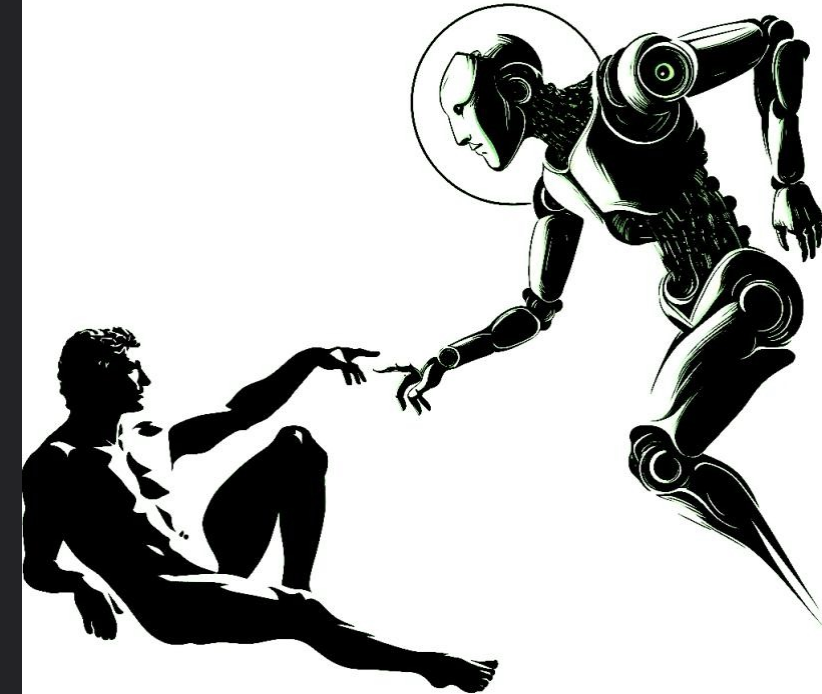
- ! Risiko logischer Fehler
- ! Eingeschränkte Argumentationsmöglichkeiten
- ! Kein gesunder Menschenverstand
- ! Kein Gedächtnis - damit begrenztes Lernen
- ! Antworten sind nicht planbar
- ! Toxizität und Bias
- ! Kein Verständnis der Welt ("strings" not "things")

... fordern eine verantwortungsvolle GenAI-Einführung

- > Bereitstellung domänenspezifischer Daten, Taxonomien und Regeln
- > Analyse domänenspezifischer Daten über Retrieval Augmented Generation (RAG)
- > LLM Agenten für Datenzugriff, Argumentation, Planung, Aufgabenausführung
- > In- und Output Guardrails - AI-Regeln
- > Letztendlich werden neuartige KI-Modelle benötigt, um die Lücke zwischen der Funktionsweise des menschlichen Gehirns in Bezug auf Lernen, Schlussfolgern und Planen zu schließen (z. B. über Joint Embedding Predictive Architecture)

Ausblick - die nächsten 20 Jahre

- KI wird die menschliche Intelligenz unterstützen, der Fortschritt wird sich beschleunigen
- Jeder wird Zugang zu superintelligenten KI-Assistenten haben
- Die Auswirkungen auf die Gesellschaft wird so tiefgreifend sein wie die Druckerpresse oder die Dampfmaschine
- Durch die Verstärkung der menschlichen Intelligenz wird KI eine neue Ära der Kreativität, Innovation und des Wohlstands einläuten
- KI hat das Potenzial, eine neue Renaissance für die Menschheit auszulösen (Yann LeCun)



BCG BOSTON
CONSULTING
GROUP



Midjourney | "Generative AI"