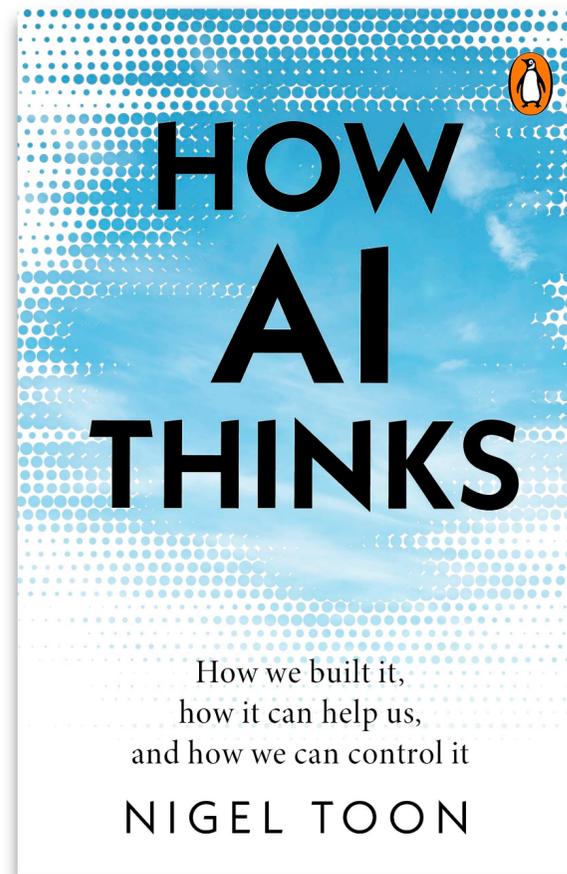


Wie künstliche Intelligenz „denkt“ !?



Wie künstliche Intelligenz „denkt“ !?

KI / AI

**I.—COMPUTING MACHINERY AND
INTELLIGENCE**

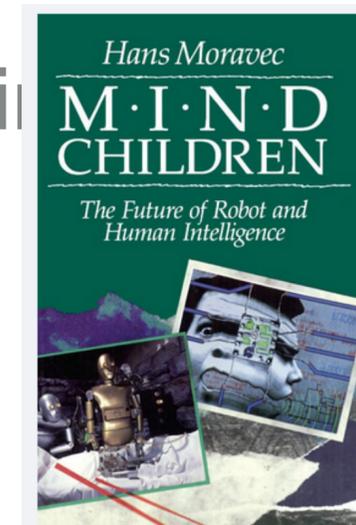
BY A. M. TURING

The Chinese Room

by John Searle

excerpted from: *Minds, Brains, and Programs* (1980)

- ▶ Forschung zu KI/AI seit den 1940er Jahren
- ▶ Alan Turing - 1950 - „The Imitation Game“
- ▶ John McCarthy - prägte den Begriff „Künstliche Intelligenz“ – 1955/1956
Dartmouth Conference 1956 – Förderantrag Rockefeller Foundation 1955
- ▶ ELIZA - Joseph Weizenbaum - 1966 der erste Chatbot
- ▶ John Searle - 1980 - Gedankenexperiment des Chinesischen Zimmers
- ▶ Hans Moravec 1988 - Moravec'sches Paradoxon



Intelligenz - Versuch einer Definition

- ▶ die Fähigkeit, rational zu denken, sinnvoll zu handeln und passend mit der Umwelt zu interagieren
- ▶ In der Psychologie unterscheidet man heute verschiedene Formen von Intelligenz, wobei jede Form eine spezifische Fähigkeit in den Vordergrund stellt. Häufig findet man die Ansicht, dass mittels einer Anzahl von sieben unterschiedlichen Faktoren ein Gesamtbild von Intelligenz dargestellt werden kann: (Süß, 2003)
- ▶ Raumvorstellung - räumlich-visuelle Aufgaben
- ▶ Erkennen der Relationen von Objekten
- ▶ mathematisch - rechnerische Fähigkeiten 
- ▶ mechanische Gedächtnisleistung 
- ▶ logisches Denken - Fähigkeit vom Einzelfall auf Allgemeines zu schließen
- ▶ Sprachverständnis - richtig verstehen und interpretieren
- ▶ Wortflüssigkeit - verwenden von Umschreibungen, Synonymen 

Machine Learning

- ▶ „Es ist unwürdig, die Zeit von hervorragenden Leuten mit knechtischen Rechenarbeiten zu verschwenden, weil mit dem Einsatz einer Maschine auch der Einfältigste die Ergebnisse sicher hinschreiben kann.“ Gottfried Wilhelm Leibniz (Poser, 2017)
- ▶ Pierre-Simon Laplace / Thomas Bayes
- ▶ Berechnung bedingter Wahrscheinlichkeiten
- ▶ Gesetz der großen Zahl (Füllsack, 2011 S 247)
- ▶ Streuungseffekt

Schwache künstliche Intelligenz

- ▶ Reactive AI / reaktive KI

KI Systeme ohne Speicherung und somit Gedächtnis
einfache Abarbeitung von vorgegebenen Regeln

z.B. Schachcomputer - > IBM Deep Blue - siegt gegen Garri Kasparov 1997

- ▶ Limited Memory AI / KI mit begrenzter Speicherung

mit Beispieldaten trainierte Systeme- z.B. Bilderkennung

die Klassifizierung eines neuen Bildes wird als weiterer Lernschritt bezeichnet

Schwache künstliche Intelligenz

- ▶ Weak AI - oder artificial narrow intelligence ANI
- ▶ basiert auf Regeln aus Mathematik, Statistik und Informatik
- ▶ schwache KI hat nicht den Anspruch menschliche Intelligenz vorzutäuschen
- ▶ Beispiele: Text-oder Schrifterkennung
- ▶ Bilderkennung und Analyse
- ▶ Spracherkennung
- ▶ personalisierte Werbung
- ▶ Systeme zur automatisierten Sprachübersetzung
- ▶ Navigationssysteme
- ▶ Chatbots und Steuerung von Telefondiensten
- ▶ Sprachausgabe etc....

Starke künstliche Intelligenz

- ▶ auch Superintelligenz, oder engl.: strong AI oder Artificial General Intelligence AGI
- ▶ Theory of Mind -
Erkennen von Bedürfnissen und Emotionen - Integration in den sozialen Alltag
- ▶ Fähigkeit, Gefühle, Wahrnehmungen und Gedanken anderer einzuordnen und deren Verhaltensweisen einzuschätzen. (Stangl, 2020)

TECH

It's becoming less taboo to talk about AI being 'conscious' if you work in tech

By [Effie Webb](#)



Three years ago, Google fired an engineer who claimed AI was "sentient." Google; Husam Cakaloglu/Getty, Tyler Le/BI

Starke künstliche Intelligenz

- ▶ **AI with Self-Awareness / Intelligenzen mit Selbstbewusstsein**
- ▶ gehören heute „noch“ in die Rubrik Science Fiction
- ▶ würde nicht ausschließlich reaktiv handeln, sondern aus eigenem Antrieb
- ▶ Eigenschaften:
 1. Logisches Denkvermögen
 2. Entscheidungsfähigkeit auch bei Unsicherheit
 3. Planungs- und Lernfähigkeit
 4. Fähigkeit zur Kommunikation in natürlicher Sprache
 5. Kombinieren aller Fähigkeiten zur Erreichung eines übergeordneten Ziels
- ▶ die essentielle ethische Frage - > **Zieldefinition**

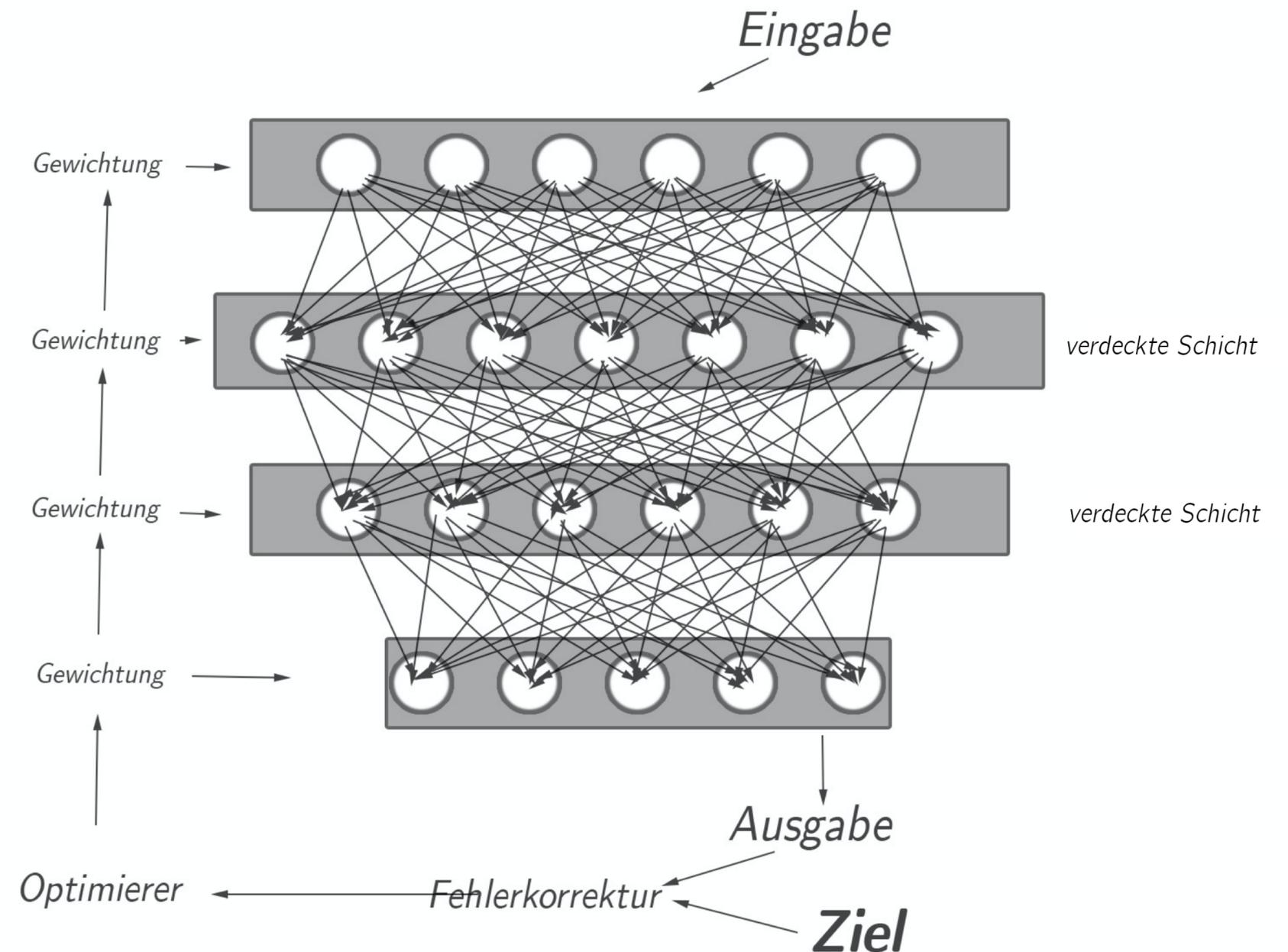
Deep Learning und Neuronale Netze

- ▶ erste Forschungsansätze zu neuronalen Netzwerken bereits in den 1940er Jahren - Marvin Minsky SNARC 1951 (Minsky 2006) (**Stochastic Neural Analog Reinforcement Calculator**)
- ▶ im Prinzip ein Nachbau der Vorgänge im menschlichen Gehirn (ca 86 Milliarden Neuronen)
- ▶ Kommunikation der Neuronen mittels elektrischer Signale
- ▶ Lernen durch Verstärkung oder Abschwächung
- ▶ bei Erreichen eines Schwellwertes „feuert“ das Neuron

**Reinforcement
Learning**

Deep Learning und Neuronale Netze

- ▶ symbolische Darstellung eines neuronalen Netzwerks
- ▶ die Bezeichnung des Deep Learning orientiert sich an der Tiefe (Anzahl der verdeckten Ebenen) der Struktur des Netzes



Training & Abfrageverfahren

KI Modelle funktionieren nicht mit Text, Bild, Audio oder Videodaten – sie benötigen numerische Werte, um damit rechnen zu können.

Diese numerischen Werte werden in Vektor Daten umgerechnet. Beim (sehr rechenintensiven) Training lernt das Modell iterativ. Das Ergebnis ist ein fertiges Modell (Vektor-Raum).

In der Folge werden bei Abfragen die Texte in sogenannte Token zerlegt, die ebenfalls einem numerischen Wert zugeordnet werden und nach der Umrechnung in Vektor Daten (Embedding) mit den Trainingsdaten verglichen werden. Es werden unterschiedliche Muster und verschiedene Kontexte dabei berücksichtigt.

Die Distanz zu den verglichenen Vektoren bestimmt die Ähnlichkeit. Ist der Abstand gering, so ist die Wahrscheinlichkeit einer Ähnlichkeit sehr hoch.

Generative Pretrained Transformer

GPT-Modelle sind Sprachvorhersagemodelle, die auf der **Transformer***-Architektur aufbauen. Sie analysieren Anfragen in natürlicher Sprache, so genannte Prompts, und sagen auf der Grundlage der Trainingsdaten die bestmögliche** Antwort voraus.

*Bei jedem Schritt ist das Modell autoregressiv, d. h. es verwendet die zuvor generierten Symbole als zusätzlichen Input für die Generierung des nächsten Symbols.

**abhängig von der Temperatur - d.h. Die Temperaturen in GPTs dienen als Kontrollmechanismus. Höhere Temperaturen führen zu Zufälligkeiten, was für kreative Aufgaben von Vorteil sein kann. Im Gegensatz dazu sorgt eine Temperatur von Null für konsistente Antworten um bestimmte Ergebnisse ohne Abweichungen zu erzeugen.

Wie künstliche Intelligenz „denkt“ !?

die Antwort auf die Frage, ob ChatGPT intelligent sei ;-)

Tokens	Characters
106	575

The LLM (Language Model) model, developed by OpenAI, is designed to generate human-like text based on the input it receives. While it can generate coherent and contextually relevant responses, it is important to note that this does not imply true intelligence. The LLM model does not possess consciousness or understanding in the way that humans do. It is merely trained through large-scale datasets to predict the likelihood of certain words or phrases following specific prompts. Therefore, while it can produce impressive results, it does not exhibit genuine intelligence.

TEXT

TOKEN IDS

Wie künstliche Intelligenz „denkt“ !?

die Antwort auf die Frage, ob ChatGPT intelligent sei ;-)

Tokens	Characters
177	637

Das von OpenAI entwickelte LLM-Modell (Language Model) soll auf der Grundlage der Eingaben, die es erhält, menschenähnlichen Text erzeugen. Es kann zwar kohärente und im Kontext relevante Antworten erzeugen, aber es ist wichtig zu beachten, dass dies keine echte Intelligenz darstellt. Das LLM-Modell verfügt nicht über Bewusstsein oder Verständnis, wie es Menschen haben. Es wurde lediglich anhand umfangreicher Datensätze darauf trainiert, die Wahrscheinlichkeit bestimmter Wörter oder Ausdrücke auf bestimmte Aufforderungen hin vorherzusagen. Daher kann es zwar beeindruckende Ergebnisse liefern, besitzt aber keine echte Intelligenz.

TEXT TOKEN IDS

Token ca. +67% Buchstaben ca. +11 %

Wie künstliche Intelligenz „denkt“ !?

die Antwort auf die Frage, ob ChatGPT intelligent sei ;-)

Tokens	Characters
106	575

```
[791, 445, 11237, 320, 14126, 5008, 8, 1646, 11, 8040, 555, 5377, 15836, 11, 374, 6319, 311, 7068, 3823, 12970, 1495, 3196, 389, 279, 1988, 433, 21879, 13, 6104, 433, 649, 7068, 56887, 323, 2317, 1870, 9959, 14847, 11, 433, 374, 3062, 311, 5296, 430, 420, 1587, 539, 34608, 837, 11478, 13, 578, 445, 11237, 1646, 1587, 539, 15575, 25917, 477, 8830, 304, 279, 1648, 430, 12966, 656, 13, 1102, 374, 16632, 16572, 1555, 3544, 13230, 30525, 311, 7168, 279, 29736, 315, 3738, 4339, 477, 32847, 2768, 3230, 52032, 13, 15636, 11, 1418, 433, 649, 8356, 16358, 3135, 11, 433, 1587, 539, 31324, 22785, 11478, 13]
```

TEXT

TOKENIDS

Wie künstliche Intelligenz „denkt“ !?

die Antwort auf die Frage, ob ChatGPT intelligent sei ;-)

Tokens	Characters
177	637

```
[33717, 6675, 5377, 15836, 81229, 301, 668, 445, 11237, 12, 4559, 616, 320, 14126, 5008, 8, 23410, 7367, 2761, 48327, 56854, 2761, 469, 287, 41408, 11, 2815, 1560, 60901, 42303, 11, 16434, 7674, 22243, 77, 22412, 2991, 2781, 38500, 268, 13, 9419, 16095, 83764, 597, 2319, 2357, 72823, 2073, 737, 37966, 428, 79415, 5048, 96917, 2781, 38500, 268, 11, 19649, 1560, 6127, 66833, 6529, 11573, 2002, 11, 15297, 8898, 32457, 31972, 668, 1357, 616, 6569, 89, 15627, 267, 33963, 13, 19537, 445, 11237, 12, 4559, 616, 2807, 31266, 5289, 8969, 14104, 38310, 355, 267, 49535, 12666, 6383, 48735, 26209, 11, 13672, 1560, 43886, 18096, 13, 9419, 27201, 6197, 343, 6915, 459, 10888, 4543, 43577, 265, 29164, 22362, 729, 90593, 75154, 5542, 17465, 11, 2815, 70690, 95313, 89285, 95818, 466, 468, 9603, 466, 12666, 20602, 3696, 19919, 68, 7367, 95818, 668, 362, 1386, 1382, 11856, 52821, 14230, 1964, 89, 355, 8703, 13, 76391, 261, 16095, 1560, 83764, 38328, 81687, 1983, 11178, 79896, 15760, 10457, 79833, 11, 9234, 11289, 83, 19649, 32457, 31972, 668, 1357, 616, 6569, 89, 13]
```

TEXT

TOKENIDS

► RLHF Reinforcement Learning from Human Feedback

Sprachmodelle haben in den letzten Jahren beeindruckende Fähigkeiten bewiesen, indem sie aus menschlichen Eingaben vielfältige und ansprechende Texte erzeugen. Was jedoch einen "guten" Text ausmacht, ist naturgemäß schwer zu definieren, da es subjektiv und kontextabhängig ist. Es gibt viele Anwendungen, wie z. B. das Schreiben von Geschichten, bei denen Kreativität gefragt ist, informative Texte, die wahrheitsgemäß sein sollten, oder Codeschnipsel, die ausführbar sein sollen.

▶ RLHF Reinforcement Learning from Human Feedback

≡ TIME

✉ SIGN UP FOR OUR IDEAS NEWSLETTER

SIGN IN

SUBSCRIBE



BUSINESS • TECHNOLOGY

Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic

Wie künstliche Intelligenz „denkt“ !?

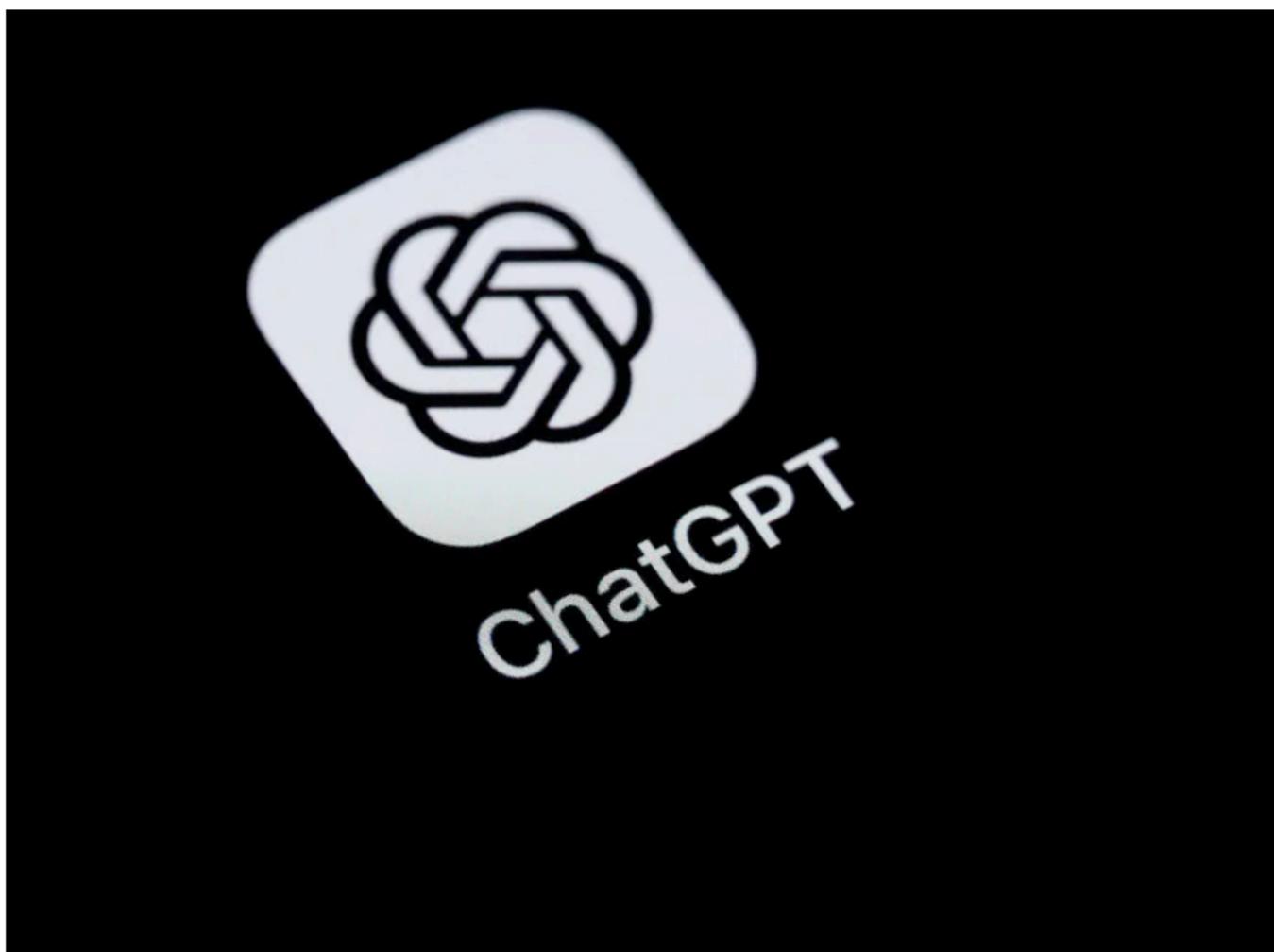


IMAGE CREDITS: SILAS STEIN / PICTURE ALLIANCE / GETTY IMAGES

AI



OpenAI pledges to make changes to prevent future ChatGPT sycophancy

Kyle Wiggers — 8:54 AM PDT · May 2, 2025

OpenAI [says it'll make changes](#) to the way it updates the AI models that power ChatGPT, following an incident that caused the platform to become overly sycophantic for many users.

Last weekend, after OpenAI rolled out a tweaked [GPT-4o](#) — the default model powering ChatGPT — users on social media noted that ChatGPT began responding in an overly validating and agreeable way. It quickly became a meme. Users posted screenshots of ChatGPT applauding all sorts of problematic, [dangerous decisions](#) and [ideas](#).

StrictlyVC 

May 13, 2025 | London, England

Get inside access to Europe's top investment minds — with leaders from Monzo, Accel, Paladin Group, and more — plus top-tier networking at StrictlyVC London.

REGISTER NOW >



Ethische Aspekte Künstlicher Intelligenz

AI-Ethik

EU – AI-Act → formuliert die Forderung nach vertrauenswürdiger KI

UNO - High-level Advisory Body on Artificial Intelligence

Ethische Aspekte Künstlicher Intelligenz?

- ▶ **Datenschutz und Überwachung:** KI-Systeme verarbeiten große Mengen personenbezogener Daten. Bedenken hinsichtlich der Datenschutzrechte und des Überwachungspotenzials liegen nahe.
- ▶ **Transparenz und Erklärbarkeit:** von KI-Systemen präsentierte Ergebnisse sind, auch wenn sie plausibel klingen, nicht im Detail nachvollziehbar! Nachvollziehbarkeit ist allerdings entscheidend für das notwendige Vertrauen.
- ▶ **Vorurteile und Diskriminierung:** KI-Algorithmen können bestehende Vorurteile in Daten, auf Basis derer sie trainiert wurden, verstärken.

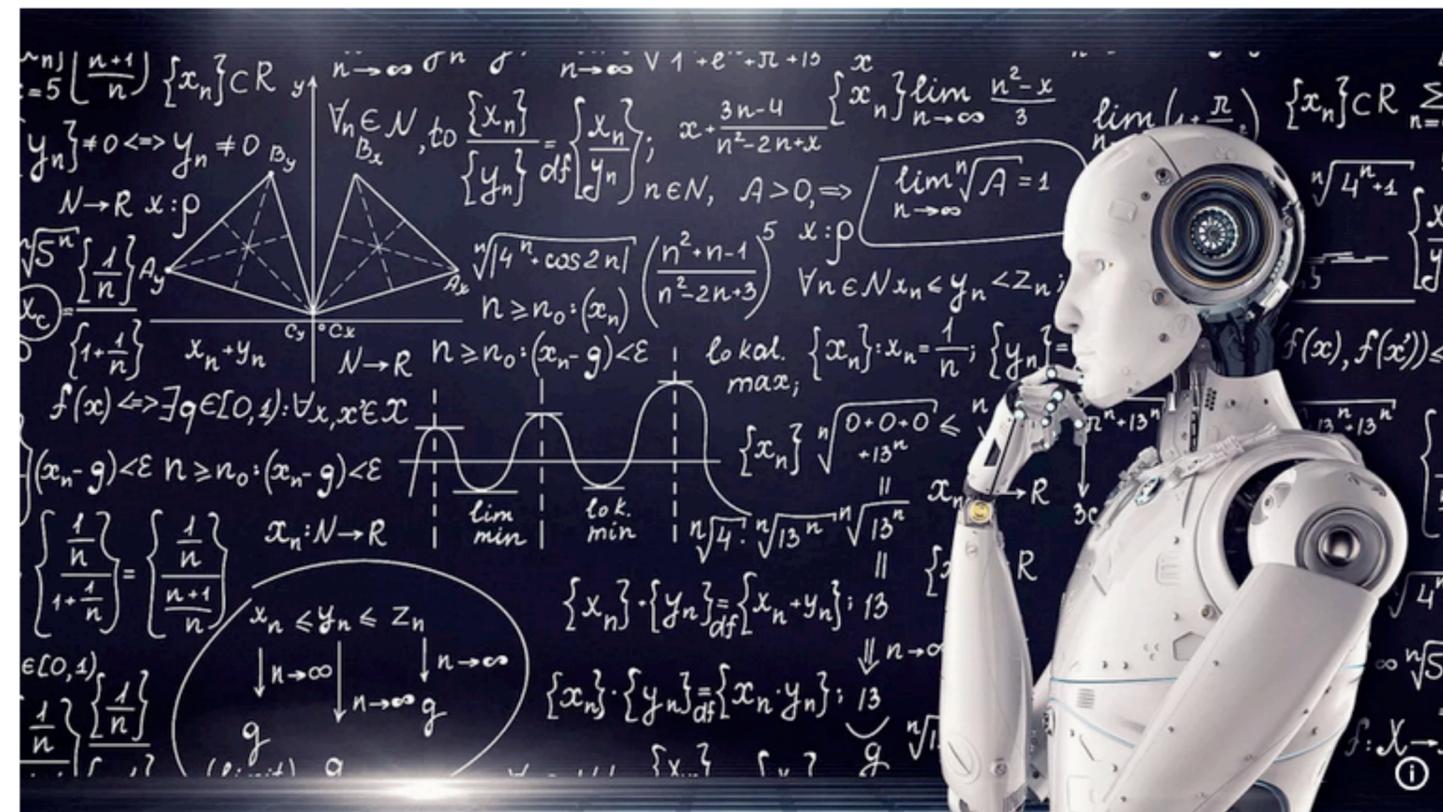
Ethische Aspekte Künstlicher Intelligenz

- ▶ **Wirtschaftliche Auswirkungen:** KI hat das Potenzial, Aufgaben zu automatisieren, die traditionell von Menschen ausgeführt werden. Dies kann zur Vernichtung von Arbeitsplätzen führen.
- ▶ **Menschliche Autonomie:** Einsatz von KI-Systemen in Kontexten, in denen Personen nicht wissen oder nicht vollständig verstehen, wie sie von KI-Systemen unterstützt oder auch gesteuert werden.
- ▶ **Rechenschaftspflicht:** Wer ist für die Handlungen von KI-Systemen verantwortlich?

Ethische Aspekte Künstlicher Intelligenz

- ▶ **Digital Divide:** befördert KI eine größere “digitale Kluft” als die Ungleichheiten bei der Nutzung von Netzdiensten und sozialen Medien der letzten Jahrzehnte?
- ▶ **Kulturelle Vielfalt:** tritt im Kontext des Wettbewerbs um die KI-Vorherrschaft in den Hintergrund. Lösungen für regionale und lokale Probleme erfordern kulturell unterschiedliche Ansätze.
- ▶ **Langfristige existenzielle Risiken:** Zu den Bedenken hinsichtlich der langfristigen Zukunft der KI gehören Szenarien, in denen die KI die menschliche Intelligenz übertrifft, was zu unvorhergesehenen und möglicherweise existenziellen Risiken führen kann.

Almost half of workers that use AI on the job dont trust it, new survey shows



Copyright mikemacmarketing/Creative Commons

By [Anna Desmarais](#)

Published on 05/05/2025 - 9:23 GMT+2 • Updated 9:23

[Share this article](#) [Comments](#)

A new study shows that most employees use artificial intelligence (AI) regularly but admit they don't fact-check its results before using it at work.

A vast majority of people use artificial intelligence (AI) every day, even though they don't trust it's outputs, according to a new study.

Researchers from the University of Melbourne in Australia and global consulting firm KPMG surveyed over 48,000 people in 47 countries from November to January 2025

Wie künstliche Intelligenz „denkt“ !?

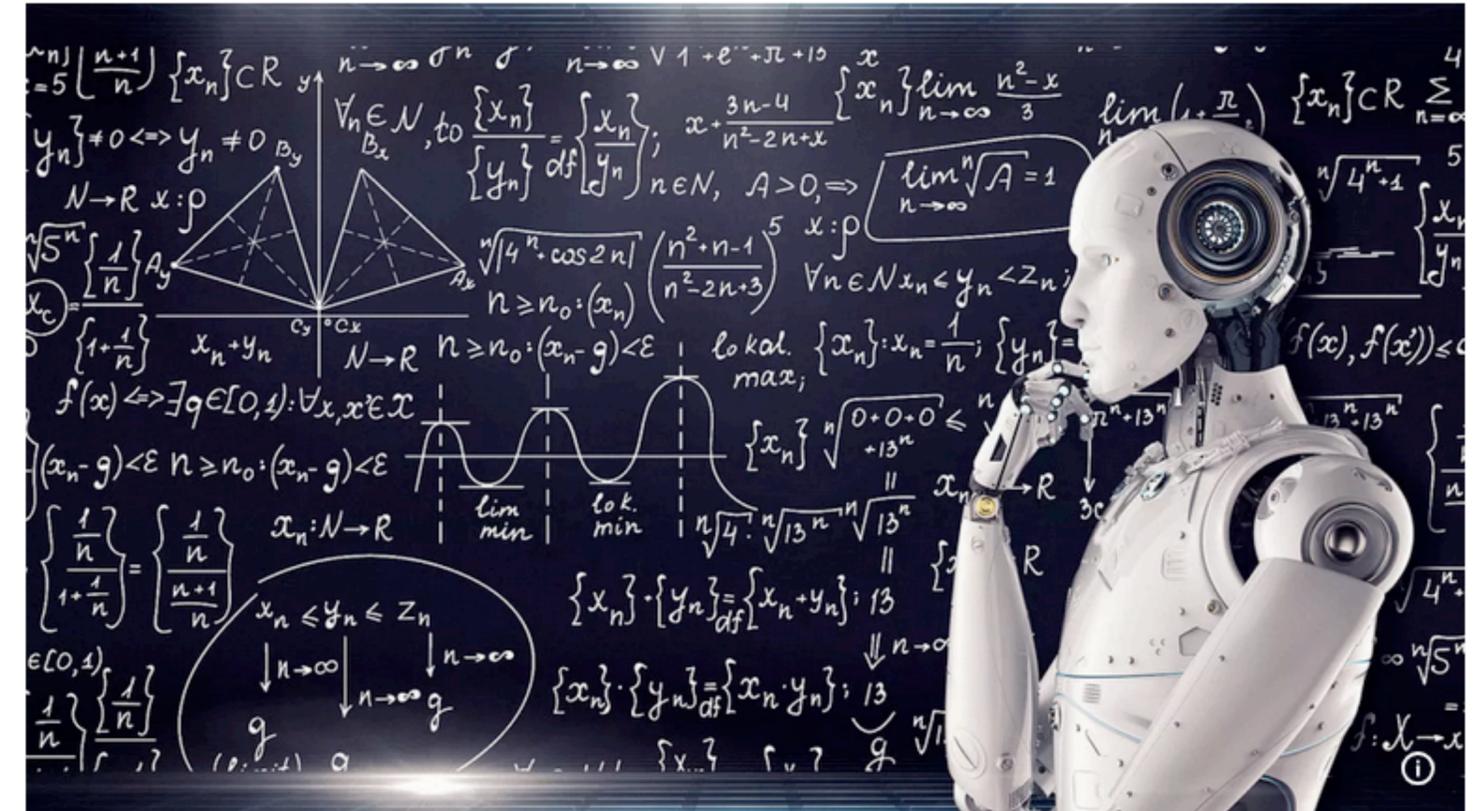
GIGO

Garbage in Garbage out

Vielen Dank für die Aufmerksamkeit!

Next > Work

Almost half of workers that use AI on the job dont trust it, new survey shows



Copyright mikemacmarketing/Creative Commons

By [Anna Desmarais](#)

Published on 05/05/2025 - 9:23 GMT+2 • Updated 9:23

[Share this article](#) [Comments](#)

A new study shows that most employees use artificial intelligence (AI) regularly but admit they don't fact-check its results before using it at work.

A vast majority of people use artificial intelligence (AI) every day, even though they don't trust it's outputs, according to a new study.

Researchers from the University of Melbourne in Australia and global consulting firm KPMG surveyed over 48,000 people in 47 countries from November to January 2025

