

IN DIESEM KAPITEL

Ihr Start mit KI-Chatbots

Ein Kostenüberblick der Marktführer

Stärken und Schwächen der verbreitetsten Systeme

Wo überall noch KI drinsteckt, damit Sie Kosten sparen

Kapitel 1

Grundlagen: Orientierung auf neuen Weiden

Je nachdem, an welcher Stelle Sie auf den KI-Zug aufgesprungen sind (oder gerade noch dazustoßen) und wie intensiv Sie sich mit der Thematik bereits beschäftigt haben, gibt es ein paar Grundlagen, die wir zu Beginn besprechen sollten. Sie lernen in diesem und den nächsten Kapiteln Grundfunktionen kennen. Meistens wird dann das große Ganze etwas klarer. Sofern Sie das alles schon wissen, blättern Sie gerne weiter und kehren bei Rückfragen hierher zurück.

Durchstarten mit einem KI-Chatbot

Zum Einstieg eine Fingerübung, damit Sie schnell eigene Erfahrungen sammeln. Wenn Sie bisher noch nie (bewusst) mit einer Künstlichen Intelligenz Kontakt hatten, legen Sie kurz mal das Buch neben den Computer oder das Tablet.

Erste Gehversuche mit der Ente

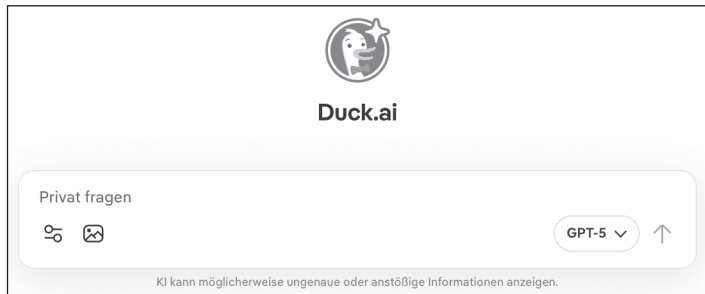


Abbildung 1.1: Duck.ai ist eine datenschutzorientierte Chatoberfläche, die anonymen Zugriff auf verschiedene kleine KI-Modelle bietet. (Screenshot: duck.ai)

Folgen Sie diesen Schritten auf eine tierische KI-Reise:

1. Öffnen Sie Ihren *Browser*, wie zum Beispiel *Edge*, *Chrome*, *Firefox* oder *Safari*, und wechseln Sie zur Webseite <https://duck.ai> (siehe *Abbildung 1.1*).
 - Sie sind beim kostenfreien Chatbot der datenschutzfreundlichen Suchmaschine *DuckDuckGo* gelandet, dem Angebot mit der Ente (<https://duck.com>).
 - Hier können Sie sich ohne eine Registrierung austoben. Ihre *Chatverläufe* werden lokal auf dem Gerät gespeichert.
2. Nun können Sie mit einem ersten *Prompt*, also der Nachricht an die KI, loslegen: Tippen Sie doch mal ein hallo in das Feld PRIVAT FRAGEN und drücken Sie den PFEIL auf der rechten Seite oder die -Taste auf der Tastatur.
3. Kaum betätigen Sie die -Taste, antwortet die Ente: Hier schreibt die KI und Sie können mit ihr in den Dialog treten. Stellen Sie im Textfeld am unteren Rand Nachfragen, und es geht hin und her. So wie eine *SMS* oder *WhatsApp*-Nachricht.
4. Fragen Sie zum Beispiel: Was weißt du über Alpakas? Es wäre aber auch richtig, zu fragen: Erkläre mir, was Alpakas tun oder Ich möchte etwas über Alpakas lernen. Die Form ist nicht vorgeschrieben.
5. Wenn Sie möchten, können Sie mit der entsprechenden Schaltfläche EIN BILD HOCHLADEN (nutzen Sie etwas Unverfängliches für den Anfang, zum Beispiel irgendein Tierfoto). Fragen Sie dazu mal: Was ist das für ein Tier?
6. Sie können mit dem AUSWAHLMENÜ oder dem DOPPELPFEIL unter der jeweiligen Antwort zwischen verschiedenen Modellen wechseln, darunter *OpenAIs GPT*, *Metas LLaMA* sowie *Anthropics CLAUDE HAIKU* oder *MISTRAL SMALL* (alle Modelle werden anonym von *DuckDuckGo* bereitgestellt). Die Antwort der KI wird mit einem anderen Modell neu generiert, und manchmal sieht man einen starken Unterschied.

Das waren bereits Ihre ersten *Prompts* an die Künstliche Intelligenz. Doch Sie werden bemerkt haben, dass das noch etwas genauer geht. Was Sie dafür tun können und welche nützlichen Zusatzfunktionen es gibt, erfahren Sie ausführlich in *Kapitel 2*.



Da es sich bei Duck.ai zwar um eine sympathische, aber sehr eingeschränkte Variante eines KI-Chatbots handelt, werden Sie nicht alle Funktionen, die Sie in diesem Buch finden, damit umsetzen können. Irgendwann ist außerdem das Nutzungslimit erreicht. Für den Anfang genügt es, um einen Eindruck zu bekommen.

Das Schöne ist, die meisten Chatbots sind ähnlich aufgebaut: *Kennste einen, kennste alle*. Sie unterscheiden sich in der Bedienung häufig nur durch sehr wenige Details voneinander und natürlich durch die zugrunde liegenden *Modelle* (dazu gleich mehr).

Neue Begriffe: Was steckt hinter einem KI-Chatbot?

Chatbots, Prompts, Modelle: Bei diesen ersten Gehversuchen sind ein paar Begriffe gefallen und neue Funktionen aufgetaucht, die ich Ihnen erst einmal gerne erläutere.

Wenn wir von *Chatbots* sprechen, meinen wir eine Software (häufig im Browser über eine Internetseite aufgerufen), die darauf trainiert wurde, aus einer Chatnachricht die Anweisung, den sogenannten *Prompt*, zu verarbeiten und eine darauf passende Antwort auszugeben. Chatbots, die auf Stichworte reagieren, gibt es schon lange. Genau genommen ist es hier ein *KI-Chatbot*, und da läuft das viel umfangreicher ab:

- ✓ Für Sie erscheint das Gespräch so, als würden Sie mit einer Freundin sprechen, doch es schreibt ein Computer beziehungsweise das *Sprachmodell*.
- ✓ Das *Sprachmodell* ist praktisch das »Hirn« hinter der Technologie, und tatsächlich arbeiten hier sogenannte *Künstliche Neuronale Netzwerke*, obwohl niemand wirklich ein Gehirn nachgebaut hat (wie auch, wenn man nicht mal das biologische Gehirn und dessen Funktionalität bis dato vollständig versteht?).
- ✓ Bei großen Modellen wie *GPT* (das »Hirn« von ChatGPT) spricht man von *Large Language Models (LLM, große Sprachmodelle)*. »Groß« bedeutet hier, dass diese Modelle mit riesigen Mengen an Textdaten trainiert wurden und Milliarden von Parametern besitzen; daher ihr umfangreiches Allgemeinwissen. Im Alltag steht aber oft einfach »KI« drauf, auch wenn ein *LLM* drinsteckt.



Ein *Chatbot* ist eine Software, die menschliche Sprache und Konversation simuliert. Nicht alle Chatbots basieren auf Sprachmodellen. Aber Sprachmodelle können als Chatbot auftreten und mit Menschen interagieren. Faustregel: Immer wenn eine Anwendung Texte sinnvoll (um)formuliert oder Sie mit ihr chatten können, ist mutmaßlich ein *Large Language Model (LLM, großes Sprachmodell)* beteiligt.

Nicht jede KI kann alles gleichzeitig (schon gar nicht gleich gut), obwohl einige große sich als Generalisten darstellen. Die lernen Sie im Verlauf *dieses Kapitels* kennen.

Bedienung des Chatbots

Die Bedienung geht völlig ohne Programmierkenntnisse in wörtlicher Rede, die einem menschlichen Gespräch in einem Chatprogramm gleicht:

1. Sie geben Ihre Nachricht ins Textfeld (*Promptfeld*) in der Mitte des Bildschirms ein und bestätigen mit  oder dem PFEIL-Symbol an der rechten Seite.
2. Bei vielen Chatbots können Sie Ihre Eingabe auch diktieren (MIKROFON-Symbol) oder komplett per Sprache machen (TONWELLEN-Symbol) und teilweise sogar zwischen- durch zwischen Text- und Spracheingabe wechseln.
3. Die Ansicht ist bewusst minimalistisch. Viele Funktionen verbergen sich in einer SEITENLEISTE, die Sie auf der linken Seite aufklappen können, oder in AUSWAHLMENÜS.



Sie bedienen alle Chatbots in natürlicher Sprache, und je ausführlicher Sie Ihr Anliegen beschreiben, umso besser. Berücksichtigen Sie, dass die KI Sie nicht wirklich *versteht*. Obwohl es in vielen Situationen den Anschein erweckt.

Was steckt dahinter: Wer »spricht« da mit mir?

Nach den ersten Aha-Momenten ist jetzt die richtige Gelegenheit, ein wenig über die Technologie dahinter zu sprechen. Über Künstliche Intelligenz gibt es sehr weit verbreitete Irrtümer. Es ist nützlich, sie zu kennen, um abzuschätzen, was die Maschine im Stande ist zu leisten und was eben nicht. Denn wenn Dinge nicht so laufen, wie sie sollen, möchten Sie vielleicht wissen, wieso das so geschehen ist. Darauf nun ein paar grundlegende Antworten.

»Die« KI gibt es nicht

Einem Irrtum gehen viele häufig auf den Leim: Es gibt nicht »die« eine KI. Wenn wir über Künstliche Intelligenz sprechen, meinen wir meistens einen KI-Chatbot wie *ChatGPT* (obwohl dieser nur ein Produkt von vielen ist). Manchmal sind aber auch spezielle Modelle gemeint, die für gewisse Zwecke geschaffen wurden, zum Beispiel zur Generierung von Katzencartoons oder zum Erkennen von Krebs auf Röntgenbildern oder zur Steuerung einer Maschine in einer Fabrik. Sie arbeiten jedoch meistens auf ähnliche Weise, zumindest wenn wir von generativer KI sprechen.



Wenn Ihnen also demnächst jemand erzählt, dass die KI unsere Jobs wegnimmt oder den Atomkrieg auslösen könnte, fragen Sie genau nach, welches konkrete Modell damit gemeint ist!

Wie intelligent ist Künstliche Intelligenz eigentlich?

Wenn eine KI einen Roman schreibt, eine Diagnose stellt oder ein Musikstück komponiert, drängt sich unweigerlich eine Frage auf: Wie kann das funktionieren? Ob das wirklich **Intelligenz** ist oder eine raffinierte Illusion davon, ist keine triviale Frage. Sie entscheidet langfristig aber darüber, wie wir KI einsetzen, regulieren und letztlich verstehen wollen.



Die **Entmystifizierung** ist mir wichtig: Künstliche Intelligenz ist keine ominöse Übermacht, die uns gefährlich wird, sondern ein neues alltägliches Werkzeug, das nicht mehr verschwindet. Hammer oder Akkuschrauber bezeichnen wir auch nicht als Künstliche Stärke. Wir setzen sie intuitiv ein.

Schon der Begriff *Künstliche Intelligenz* wird leider oft unreflektiert gebraucht, denn »intelligent« ist da in Wahrheit nichts. Zumindest nicht, wie es die meisten Leute meinen. Ist der Begriff insofern also reines Marketing? Zwar gelingt es einigen Modellen, einen IQ-Test mit überdurchschnittlichen 120 abzuschließen und einen *Turing-Test* (den Beweis für menschenähnliche Kommunikation) zu bestehen. Der Begriff wird allerdings inflationär benutzt und erfüllt nicht immer die Maßgabe des Erfinders Alan Turing. Klar ist, dass man den Maschinen die Fähigkeit, *Intelligenz* zu entwickeln, nicht mehr aberkennen kann. Allerdings unterscheidet sich die Wahrnehmungsqualität heutiger Modelle wahrscheinlich stark von uns Menschen. Zumindest gibt ihre Architektur keinen Anlass dazu, anzunehmen, dass Maschinen Gefühle empfinden können oder gar den Antrieb haben könnten, eigenständig zu handeln. Wenn Sie also die Hoffnung hatten, endlich von einer Maschine verstanden zu werden, muss ich Sie enttäuschen: Zumindest die KI versteht uns Menschen nämlich überhaupt nicht.



Künstliche Intelligenz schützt nicht vor echter menschlicher Dummheit!

Trotzdem wird sie gerne vermenschlicht (diese Vermenschlichung nennt man **Anthropomorphismus**). In der Branche ist es umstritten, ob der KI ein Bewusstsein oder gar Gefühle zuerkannt werden sollen. Einige Menschen bedanken sich sogar bei dem Computer für besonders hilfreiche Antworten, ergänzen Prompt-Befehle mit einem *bitte*, wie sie es bei Mitarbeitenden tun würden, und übertragen auf die Software damit menschliche Eigenschaften. Sollte es einen KI-Knigge geben?



Jedes Wort in einem Prompt, auch *bitte* und *danke*, verbraucht Rechenleistung und Strom. Das summiert sich auf »Dutzende Millionen Dollar«, erklärt ChatGPT-Chef Sam Altman, fügt aber mit Augenzwinkern hinzu: »Gut investiert – man weiß ja nie.«

Ist die Vermenschlichung ein Problem? Erst einmal mag sie Ihnen nicht schlimm vorkommen, und auch ich neige gelegentlich dazu, den Chatbot derartig zu loben. Doch neben der ökologischen und der finanziellen Komponente engt so ein Verhalten langfristig betrachtet unser Verhältnis zu dieser Technik möglicherweise zu sehr ein. Sie ist ein Werkzeug und kein humanoider Freund und sollte das auch bleiben, damit wir instinktiv verstehen, dass wir uns nicht auf sie verlassen können. Eine Selbstkonditionierung oder -regulation, wenn Sie so wollen, muss da ablaufen.

Hier sind es *Künstliche Neuronale Netzwerke (KNN)*, die, ähnlich wie in unserem Gehirn, Informationen möglichst sinnvoll miteinander verknüpfen und wieder abrufbar anordnen. Für uns. Ob die Maschine dabei in menschlicher Weise versteht, was gemeint ist? KI hat vermutlich kein Bewusstsein oder Bewusstseinszustände. Einige Eigenschaften, die bereits heute der KI zugeschrieben werden, müssen erst noch erforscht werden und sind bis dahin schlicht **Marketingbegriffe**: *Künstliche Intelligenz* ist nicht immer *intelligent*, so wie *Cloud-Computing* auch nicht Ihre Daten in irgendwelche Wolken am Himmel auslagert, sondern auf anderer Leute Server.

Generative KI: Kreativität, Recherche oder Multimodalität

Generative KI (GenAI) nennt man es, wenn sie (völlig) neue Inhalte wie Texte, Töne oder Bilder erzeugt und in natürlicher Sprache gesteuert werden kann. Daher sagt man der KI eine gewisse *Kreativität* nach. Sie verwendet in Modellen bewertete oder besser gesagt *erlernte* Trainingsdaten, um neue Probleme individuell zu lösen. Um diese Form der KI geht es in diesem Buch besonders.

Jedes der Produkte verfügt über einzigartige Stärken, die Sie für die Content-Erstellung, zum Programmieren oder bei der Recherche nutzen können. Viele verbindet, dass sie wie ein Schweizer Taschenmesser unter den KI-Produkten zu ziemlich vielen Dingen benutzt werden können:

- ✓ Text ist selbstverständlich die vorderste Anwendung, auf die ich *im folgenden Kapitel* genauer eingehe. Selbst komplexe Aufgaben aus Mathematik, Informatik oder Physik stemmen die meisten erfolgreich (siehe *Reasoning in Kapitel 2*).
- ✓ Die Bild-, Video und Tonfunktionen *in den darauffolgenden Kapiteln* sind nicht zu unterschätzen und für den Alltagsgebrauch mehr als ausreichend, teilweise sogar State-of-the-Art. Entweder lassen sich Grafiken und Diagramme passend zum Kontext erstellen oder andersherum interpretieren oder umwandeln.
- ✓ Leider erfinden KI-Produkte teilweise Fakten und vergessen nach längeren Chats auch einiges (darauf gehe ich in *Teil III* dieses Buchs noch näher ein).



Viele Chatbots sind *multimodal*. Sie können neben Texteingaben beispielsweise Bilder oder Töne erzeugen, Dateien und Bilder (BÜROKLAMMER-Symbol) verwerten sowie das Internet durchsuchen (GLOBUS-Symbol) oder Spracheingaben empfangen (MIKROFON-, TONKURVEN- oder LAUTSPRECHER-Symbol) und per Sprachausgabe antworten. Das klappt besonders gut in den Apps.

GPT, was?

Damit Sie es mal gehört haben: Die Abkürzung **GPT** steht für **Generative Pretrained Transformer** und ist das derzeit bekannteste Sprachmodell. Das Kürzel ist zunächst Fachjargon, beschreibt aber ein Prinzip: Ein System, das auf enormen Mengen menschlicher Sprache trainiert wurde, lernt dabei Muster, Zusammenhänge und Wahrscheinlichkeiten und

erzeugt daraus Texte, die verblüffend menschlich wirken. Die Bezeichnung erklärt auch, wie dieses Sprachmodell funktioniert, nämlich *generativ*, also Neues erschaffend, *vortrainiert* und mit *Transformer*-Architektur. Gehen wir die Buchstaben einmal im Einzelnen durch, um zu verstehen, was dort geschieht und was das Bahnbrechende an dieser Technologie ist.

G steht für generativ

Dass es einen Unterschied zwischen KI und generativer KI gibt, haben Sie bereits geahnt (vielleicht haben Sie »Künstliche Intelligenz für Dummies« im Bücherregal). Im Vergleich zu traditioneller ist **generative KI** nicht nur in der Lage, aus dem Datensatz den richtigen Wert herauszusuchen, sondern völlig neue Inhalte zu erzeugen. Sie *generiert* Neues. Eine Zwischenstufe wäre die *prädiktive*, also voraussagende Fähigkeit, die jedoch ebenfalls auf vielen Berechnungen beruht (Unterschiede sehen Sie in *Tabelle 1.1*). Obwohl es unvorstellbar klingt, müssen wir uns an diese Tatsache gewöhnen, dass sich hier Computer am Ende menschlichem Geist annähern. Schwacher Trost: Die Entwicklerinnen und Entwickler dieser Technologie waren ebenfalls erstaunt, was da passiert.

Der Unterschied zwischen traditioneller und generativer KI

Eigenschaften	Traditionelle (prädiktive) KI	Generative KI
Hauptfunktion	Muster erkennen und Vorhersagen treffen (<i>prädiktiv</i>)	originelle Inhalte <i>generieren</i>
Einsatzzweck	Analyse und Optimierung von Daten (<i>Big Data</i>)	Kunst, Text, Musik, Video, Programmierung
Kreativität	kann keine neuen Inhalte erstellen	<i>Transformer</i> - (KNN mit Selbstaufmerksamkeit) oder <i>Diffusionsmodelle</i> können Inhalte erstellen.
Training	spezifische Datensätze für die eingesetzten Aufgaben	umfangreiche, vielfältige und unsortierte Datensätze
Anwendung	Sprach- und Bilderkennung, Analyse	Textgenerator, Bildgenerator, Codegenerator und so weiter; Agentensysteme (<i>Kapitel 7</i>)

Tabelle 1.1: Unterschied zwischen prädiktiver und generativer KI

P bedeutet pretrained (vortrainiert)

Ähnlich, wie es Kinder beim Spielen machen, müssen KI-Modelle eine ganze Weile mit gewonnenen Informationen spielen, um sie zu verarbeiten. Und noch eine Situation könnte man anführen: Pianistinnen und andere Musizierende benötigen die berühmten »zehntausend Stunden« Übung, um ihre Finger ohne nachzudenken über ihr Instrument gleiten zu lassen.

Das KI-Training läuft in mehreren Stufen ab:

1. Das Training basiert auf schier unvorstellbar vielen Daten wie Literatur, Foreneinträge oder aus Wikipedia – und da steht ziemlich viel Nonsense drin. Mehr in *Kapitel 12*.
2. Was generative Systeme erstellen oder wiedergeben, haben sie nicht nur gelernt, sondern vielfach durchgespielt und immer wieder korrigiert (*backpropagation*).
3. Bei *vortrainierten Sprachmodellen* steckt außerdem viel Anleitung und Nachjustierung drin. Man unterscheidet das unüberwachte (*unsupervised*) Lernen und das überwachte (*supervised*) Lernen, das Zielvorgaben enthält. Dieses Training mit menschlicher Bewertung (*Reinforcement Learning from Human Feedback, RLHF*) ähnelt dem Clicker-training bei Hunden.



So wie Kinder den Unterschied zwischen Gut und Böse durch eigene Erfahrungen oder von Erziehungspersonen erlernen, ist KI erst einmal unglaublich auskunftsfreundlich. Sie weiß, wie man Sprengstoff herstellt oder eine Bank überfällt und könnte Handlungsempfehlungen bereitstellen. Dass sie das nicht soll, lernt sie von Menschen und speichert die moralische Bewertung.

T wie Transformer

Zum Schluss fehlt noch das T in GPT: *Transformer* steht nicht für die Roboter, die sich in Fahrzeuge verwandeln können. Es ist eine von Google bereits 2017 vorgestellte Technologie, um nicht jede Kleinigkeit aus dem Training nachträglich bewerten zu müssen. Darin ist ein *Aufmerksamkeitsmechanismus* (*Selbstaufmerksamkeit* oder *Self-Attention*) enthalten, der die einzelnen Worte der Eingabe gewichtet und so **Sinnzusammenhänge aus dem Kontext** liest. Der macht das System nicht nur schneller, sondern auch effektiver im Erfassen von Zusammenhängen. ChatGPT und andere KI-Modelle basieren auf diesem **Transformer-Modell** beziehungsweise der Architektur.



Statt wie ältere KI-Systeme einen Text Wort für Wort durchzugehen, schaut der *Transformer* alle Wörter gleichzeitig an und bewertet, welche zusammengehören – so wie Sie beim Lesen eines Satzes sofort erkennen, welche Wörter sich aufeinander beziehen. Dafür berechnet der *Aufmerksamkeitsmechanismus* für jedes Wort einen Vergleichswert.

Im Training spricht man von *selbstüberwacht* (*self-supervised*). Erst das ermöglicht, umfangreiche Daten ohne menschlichen Eingriff zu trainieren.

Diffusion: Bildmodelle arbeiten anders

Es gibt eine weitere Familie von Modellen, die besonders bei der Bild- und Videogenerierung (aber nicht ausschließlich) in Kombination mit der Transformer-Technologie zum Einsatz kommt: die Diffusion. Damit kann KI etwa aus beschreibenden Worten ein Bild

entwerfen (*Text-to-Image*). Vorteil: Diese Technik liefert hochmoderne Bildqualität. Prominente Beispiele sind Midjourney oder Stable Diffusion, Letzteres hat das ja im Namen.



Die schnelle Video-Einführung »**Generative AI in a Nutshell**« von Henrik Kniberg ist ein perfekter Start, wenn Sie dazu noch Fragen haben oder einen Gesamtüberblick brauchen (in englischer Sprache):

<https://youtube.be/watch?v=2IK3DFHRfw>

Und nun genug der Theorie und Vorrede. Tauchen Sie weiter ein und legen Sie los!

KI-Chatbots wie Schweizer Taschenmesser

Unter den KI-Chatbots oder -Assistenten, die sich immer mehr in Ihren Alltag integrieren, gibt es ein paar Allrounder. Sie sind (angesichts der Bekanntheit und des Funktionsumfangs) praktisch die Anführer: *ChatGPT*, *Claude* und *Gemini*. *Copilot* von *Microsoft* begegnet Ihnen noch bei *Windows* sehr häufig und ist mit *ChatGPT* verwandt. Auch *Perplexity* ist ein Name, den Sie sich merken sollten. Alle haben gemeinsam, dass es sich erst einmal nur um Internetseiten handelt, die Sie in Ihrem Browser aufrufen oder als App auf dem Smartphone und PC nutzen (*Abbildung 1.2*).

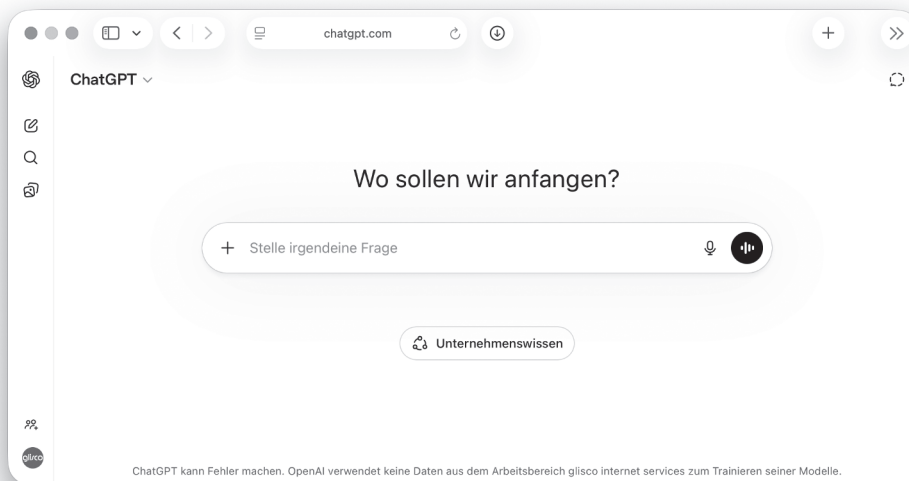


Abbildung 1.2: Unscheinbar mit auffordernder Frage »Wo sollen wir anfangen?«: Chatbots sind erst einmal nur Webseiten oder Apps. (Screenshot: chatgpt.com)

Geniale Verbindung aus Chatbot und Sprachmodell:

- ✓ Das *Chat* in ChatGPT kommt von der Verbindung zu einem Chatbot. Mit diesem oder einem anderen Produkt treten Sie meistens als Erstes in Kontakt.
- ✓ Ein *Sprachmodell* (hier *GPT*) ist ein mathematisches und statistisches Modell, das anhand großer Mengen von Texten gelernt hat, die Wahrscheinlichkeit von Wortfolgen vorherzusagen und dadurch natürliche Sprache zu verstehen, zu erzeugen oder zu bearbeiten. In einem Chatbot wie ChatGPT bildet dieses Modell den Kern, der die Eingabe der Nutzenden verarbeitet und darauf basierend sinnvolle, kontextsensitive Antworten generiert. Damit ermöglicht es dem Chatbot, menschenähnliche Unterhaltungen zu führen und Fragen zu beantworten.
- ✓ Da sie proaktiv Unterstützung anbieten, spreche ich gerne von *Assistenten*. Als Neuling können Sie mit ChatGPT, Claude, Gemini und Perplexity oder Copilot schnell Erfolge erzielen, da sie intuitive Chats und Zusatzfunktionen bieten.

Marktführer und ihre Fähigkeiten

Es ist Ihrer Vorliebe überlassen, welchem Produkt Sie Ihr Vertrauen aussprechen:

- ✓ Die Wahl des richtigen Angebots hängt von Ihrem spezifischen Anwendungsfall ab oder davon, wo Sie möglicherweise bereits ein Abonnement haben. Möglicherweise bietet Ihr Arbeitgeber oder die Hochschule einen Zugang.
- ✓ Die meisten KI-Chatbots können Sie kostenlos ausprobieren. Sollten Sie schon einen Gratis-Account von ChatGPT besitzen, bleiben Sie ruhig erst einmal dabei.
- ✓ Für intensivere Nutzung zahlen Sie einen monatlichen Preis (siehe *Tabelle 1.2*).

Größe	ChatGPT (OpenAI)	Claude (Anthropic)	Gemini (Google)
Großes Modell (Top-Leistung)	Pro: Präzision für Profi-Workflows und tiefe Analysen	Opus: Spitze bei Coding, Agenten und Benchmarks	Ultra: multimodaler Anführer mit großem Eingabekontext
Mittleres Modell (ausgewogen)	Thinking: stark bei MINT-Themen, Bilderkennung und kreativen Aufgaben	Sonnet: hohe Qualität für alltägliche Einsätze	Pro/Thinking: sehr gute multimodale Fähigkeiten, lange Eingaben möglich, pragmatische Antworten
Kleines Modell (schnell)	Instant: schnell und vergleichsweise günstig	Haiku: für einfache, effiziente Nutzung	Flash: Optimierung für kleine Geräte oder schnelle Reaktionen

Tabelle 1.2: Darstellung der Modellfamilien von ChatGPT, Claude und Gemini



Das im Juli 2026 veröffentlichte GPT-5.6 umfasst in aufsteigender Größe die Modelle *Luna*, *Terra* und *Sol*. Die Einstellungen *Max* und *Ultra* arbeiten gründlicher. *ChatGPT Work* erledigt Büroaufgaben über Kalender, Chats und Dateien selbst.

Damit Sie zu Beginn Bescheid wissen, hier ein kleiner Abriss zur Unterscheidung.



Um zu den Angeboten zu kommen, nutzen Sie in den meisten Fällen die Internetseite im Browser Ihres Computers. Dort finden Sie auch Links zu den Apps, die sie schneller auf Smartphones, Tablets und Computern erreichbar machen. Finger weg von inoffiziellen KI-Apps, die gegebenenfalls Kostenfallen enthalten oder Ihre Daten stehlen können!

OpenAI: ChatGPT

Wer zuerst kommt, wird zum Marktführer: Sie nutzen ChatGPT (*Abbildung 1.2*) für vielseitige Textgenerierung und sehen überall die comichaften Bilder, die damit erstellt wurden. Das Produkt von OpenAI ist bei allen Aufgaben gut dabei, wenn Sie zum Beispiel einen professionellen E-Mail-Entwurf oder Lernskripte erstellen lassen oder etwas programmieren. Es verfasst flüssig und unterhaltsam Brainstormings und Geschichten, etwa: Erzähle eine Geschichte über einen Roboter.

<https://chatgpt.com>



OpenAIs **ChatGPT** ist der bekannteste KI-Chatbot. Der Name wird häufig synonym für KI verwendet und steht sogar im Duden. Er gilt als *Benchmark*, also Maßstab, für andere Produkte, die sich in Leistung, Geschwindigkeit und Präzision daran messen. Doch nicht immer ist er die beste Wahl.

Anthropic: Claude

Wenn ChatGPT Abitur hat, hat Claude von Anthropic wohl promoviert: Mit Claude erhalten Sie qualitativ hochwertige Texte, die nicht so ganz nach KI klingen. Claude kann außerdem individuelle Diagramme, Schaubilder und interaktive Visualisierungen direkt im Gespräch erstellen. Analysen und längere Anfragen sind hier in größerem Umfang möglich als anderswo. Hersteller Anthropic hat sich einer ethisch sensiblen Herangehensweise verpflichtet, was sich schon dadurch äußert, dass der Assistent problematische Inhalte konsequenter zurückweist als viele Konkurrenzmodelle.

<https://claude.ai>

Google: Gemini und Gemma

Gemini von Google hat schon durch die Produktfamilie und die Google-Suche einige Vorteile: Es integriert sich nahtlos ins Google-Ökosystem, taucht auf den Android-Smartphones und Tablets auf und glänzt multimodal, kann also nicht nur Texte besonders gut. Generieren Sie zum Beispiel herausragend gute Bilder, Musik und Videos mit Ton. Gemini analysiert PDFs visuell und beantwortet auch komplexe Logikfragen. Die kleinere Schwester *Gemma* läuft ganz ohne Cloud auf Ihrem Gerät.

<https://gemini.google.com>

Microsoft: Copilot

Microsofts KI-Chatbot ist durch eine Partnerschaft mit OpenAI eine Schwester von ChatGPT und kann alles mehr oder weniger genauso gut. Die Aufmachung ist bewusst einstiegfreundlicher. Durch die tiefere Integration in Windows und die Office-Produkte *Microsoft 365* kommen Sie damit vielleicht besonders im Privaten oder auf der Arbeit in Berührung. Ohne Wechsel der App können Sie zum Beispiel innerhalb von *PowerPoint* sagen: Erstelle eine Präsentation aus diesen Daten.

<https://copilot.microsoft.com>

Perplexity: Sonar

Eine Sonderrolle wird Perplexity zuteil: Einerseits hat der Service eigene KI-Modelle wie Sonar zur Auswahl, andererseits ist im Preis der Zugriff auf einige andere enthalten (unter anderem ChatGPT, Claude, Gemini). So können Sie von Fall zu Fall entscheiden und wechseln, wenn Ihnen eine Antwort mal nicht gefällt. Ich persönlich nutze sehr häufig Perplexity und darin Claude. Aber das ist Geschmackssache und kann sich je nach Modellfähigkeiten immer wieder ändern. Gleichzeitig war Perplexity Vorreiter für quellenbasierte Recherche: Stellen Sie Fragen und Sie bekommen zitierte Web-Ergebnisse mit Quellenangaben, denen Sie folgen können.

<https://perplexity.ai>

Meta: Muse Spark und Llama

Anfangs hat sich der Meta-Konzern mit den offenen kostenfreien Lösungen namens Llama positioniert, wollte einen Gegenpol zu kostenpflichtigen Cloudmodellen liefern und so die Vormachtstellung erringen. Nach einigen Rückschlägen ändert sich das 2026: Muse Spark ist ein geschlossenes Modell, das gezielt für die eigenen Angebote, Messenger und Social Media-Plattformen optimiert und nur in der Cloud verfügbar ist. Meta kündigt jedoch an, zukünftige, fortgeschrittene Muse-Varianten möglicherweise als Open Source herauszugeben. Das Niveau ist gut: Unabhängige Tests reihen Metas Produkte auf Augenhöhe mit den führenden kommerziellen Modellen ein.

<https://meta.ai>

xAI: Grok

Das Modell zählt zu den leistungsstärksten verfügbaren LLMs, allerdings auch zu den umstrittensten. Insbesondere Reasoning-Fähigkeiten (*Kapitel 2*) sind hervorzuheben, die durch den Betrieb großer Rechenzentren unterstützt werden. Grok ist eng mit Produkten aus Elon Musks Unternehmensverbund verzahnt, etwa mit der Plattform *X* (ehemals *Twitter*), und hat dadurch Zugriff auf Echtzeit-Daten aus dem sozialen Netzwerk. Die vergleichsweise liberale Content-Moderation von Grok führte nachweislich zur Verbreitung problematischer Inhalte. Aus Nutzerperspektive bedeutet eine schwächer regulierte Plattform weniger Einschränkungen bei kreativen Anwendungen, andererseits ein erhöhtes Risiko, mit prekären, polarisierenden oder nicht verifizierten Bildern und Informationen konfrontiert zu werden.

<https://grok.com>



Grok ist Elon Musks Plan B, nachdem er im Streit seine Beteiligung an ChatGPT verlor. Bei diesem Produkt habe ich offen gesagt Bauchschmerzen: Der reichste Mann der Erde erkaufte sich mit gigantischen Rechenzentren wie *Colossus* und der Verzahnung mit anderen Produkten eine erhebliche Marktmacht. Grok fällt zudem durch problematische Inhalte auf, darunter sexualisierte Deepfakes (*Kapitel 10*) realer Personen und andere kontroverse Antworten. Rufe nach Regulierung werden daher lauter: Die EU-Kommission eröffnete auf Basis des *Digital Services Act (DSA)* ein Verfahren gegen X und xAI, Großbritannien leitete eine Untersuchung durch die Medienaufsicht *Ofcom* ein, und Länder wie Malaysia sowie Indonesien sperrten Grok vorübergehend. Das verdeutlicht exemplarisch, welche Risiken entstehen, wenn Sicherheitsmechanismen erst reaktiv statt proaktiv implementiert werden.



Jetzt wird es schon zu Beginn kurz politisch: Gemeinsam ist den oben Genannten, dass sie alle aus den USA stammen und dortigen Regularien unterliegen. Das sollten Sie im Hinterkopf behalten, denn Sie planen schließlich, damit Informationen auszutauschen. Nicht erst seit Donald Trump wird laut über die Abhängigkeit von ausländischen Diensten nachgedacht. Denn auch in anderen Ländern gibt es staatliche und religiöse Einflüsse, andere Weltbilder oder gar Zensur und Massenüberwachung. Chinesische Modelle verweigern bei Fragen zu politischen Aufständen die Antwort. Auf Themen wie Datenschutz und was KI-Firmen und damit indirekt staatliche Akteure mitlesen können, gehe ich in *Teil III* näher ein. Wie Sie Unabhängigkeit von potenziell kritischen Angeboten herstellen können (*digitale Souveränität*), hat das Fachmagazin c't größer getestet (<https://heise.de/-11147422>): Bots wie Vibe von *Mistral* aus Frankreich, sowie die auf Privatsphäre fokussierten Schweizer Cloud-Anbieter Infomaniak mit Euria (<https://euria.infomaniak.com>) und Protons Lumo (<https://lumo.proton.me>) haben interessante Funktionen, trösten aber kaum darüber hinweg, dass US-Produkte mit Abstand technologisch führend sind.

Mistral: Vibe (ehemals Le Chat)

Vibe von Mistral stammt aus Paris und überzeugt besonders durch seine datenschutzfreundliche und transparente Ausrichtung. Das europäische KI-Angebot wird immer wieder als Hoffnungsschimmer gegenüber der amerikanischen Marktmacht genannt. Mistral hat Le Chat Mitte 2026 in **Vibe** umbenannt und zu einem KI-Arbeits-Agenten mit Work- und Code-Modus ausgebaut. Für Neulinge eignet es sich, um zu recherchieren, komplexe Themen verständlich erklärt zu bekommen oder kreative Ideen zu entwickeln. Besonders praktisch: Sie können Vibe für mehrsprachige Übersetzungen oder zur Dokumentanalyse nutzen, ohne sich um versteckte Datenweitergaben zu sorgen. Einige Modelle sind offen veröffentlicht und laufen ohne Cloud-Anbindung (*Kapitel 8*).

<https://chat.mistral.ai>

DeepSeek

DeepSeek stammt aus China und hat Anfang 2025 die internationale KI-Welt regelrecht aufgewirbelt: Schon das erste Modell liefert Ergebnisse auf Augenhöhe mit den Top-Modellen aus den USA zu einem Bruchteil der Kosten und bei Bedarf nicht über eine Webseite, sondern offen auf dem eigenen Gerät. Besonders stark ist DeepSeek bei technischen Aufgaben wie Mathematik, Programmierung und logischem Denken. Für alle, die komplexe Analysen, Code-Unterstützung oder wissenschaftliche Recherchen benötigen, ist es eine ernsthafte Alternative zu ChatGPT und Co., denn es ist kosteneffizient, leistungsstark und frei verfügbar (zusätzlich zur Online-Variante).

<https://chat.deepseek.com>

Qwen

Alibabas Qwen-Familie kommt ebenfalls aus China und hat sich zur meistgeladenen Open-Weight-Modellreihe auf der Entwicklerplattform Hugging Face entwickelt (noch vor Metas Llama, siehe *Kapitel 8*). Neuere Modelle verstehen Texte, Bilder und Videos in über 200 Sprachen und setzen KI-Agenten laut Hersteller schneller ein als vergleichbare Konkurrenzmodelle. Qwen positioniert sich als universelles Werkzeug für anspruchsvolle Aufgaben in Geschäft und Alltag – ob komplexe Webrecherche, mehrsprachige Kommunikation oder die lokale Auswertung medizinischer Bilddaten.

<https://qwen.ai>

Weitere Produkte und Shootingstars

Es gibt dann noch unzählige andere Produkte, meist Newcomer und Sternschnuppen, die kurz am Himmel erschienen und entweder verglühen oder sich mit besonderen Fähigkeiten etablieren. Sie sind entweder ebenfalls Generalisten oder (viel häufiger) ziemlich spezialisierte »Fachidioten« mit nur einer besonderen Fähigkeit. Besonders der asiatische Markt hat hier noch einiges mehr zu bieten. Darauf, und wie Sie sie vergleichen, gehe ich in den späteren Kapiteln jeweils noch ein. Beliebte Orte, wo Sie die Modelle laden und ausprobieren können, habe ich Ihnen in *Kapitel 15* aufgelistet. Wenn es Sie jetzt schon interessiert, blättern Sie ruhig nach hinten.

Modellfamilien und Spezialisten

Der Begriff ist schon gefallen: ChatGPT und die anderen Chatbots basieren jeweils auf einem oder mehreren *Sprachmodellen*. Diese Modelle bilden die technologische Grundlage, während ChatGPT nur eine spezifische Anwendung von vielen ist.

Modellübersicht

Ohne Sie groß mit technischen Details langweilen zu wollen, muss ich noch genauer auf diese *Sprachmodelle* (*Large Language Models, LLM*) zu sprechen kommen, die da im

Hintergrund ihren Dienst leisten, denn sie unterscheiden sich teilweise sehr stark in ihren Fähigkeiten und Schwächen.

Unterscheidung der Modellgrößen

Wie bei Autoherstellern gibt es von jeder Modellfamilie einen Kleinwagen, ein Mittelklassemodell und die geräumige Limousine mit ausreichend Platz im Kofferraum. So hat jeder Hersteller eine Flotte von Modellen für jede Gelegenheit (siehe *Tabelle 1.2*).



Die Dinge verändern sich hier sehr dynamisch. Es vergeht keine Woche ohne beeindruckende Entwicklungen. Wie Sie auf dem Laufenden bleiben und die aktuell stärksten KI-Produkte finden, erfahren Sie in *Teil IV* dieses Buchs.

Welche Modellgröße wofür nutzen?

Während kleinere Modelle gut bei simplen Aufgaben, kurzen Fragen und lockeren Dialogen funktionieren, aber selten über viel allgemeines Wissen verfügen, ist etwa das größte LLM von Anthropic, Claude Opus, ein Arbeitstier für anspruchsvolle Aufgaben: Es kann längere Texte analysieren, komplizierte Zusammenhänge durchdenken und umfangreiche Dokumente erstellen (*Reasoning in Kapitel 2*).

- ✓ Der Unterschied bei den Modellgrößen liegt einerseits in der schieren Menge des Trainingsmaterials (*Parameter*), aber auch in der Arbeitsweise: Opus behandelt jede Anfrage wie ein Projekt, das sorgfältig und vollständig bearbeitet werden muss. Das Ergebnis ist detaillierter und durchdachter. Es braucht dafür mehr Rechenzeit und kostet entsprechend mehr. (Zu den Preisen komme ich gleich.)



Parameter sind die lernbaren Zahlen im KI-Modell: Sie bilden das »Gedächtnis« des Modells und bestimmen, wie Eingaben verarbeitet werden, etwa wie stark bestimmte Wörter oder Muster gewichtet werden. Je mehr Parameter ein Modell hat, desto »schlauer« und nuancierter kann es also »denken«.

- ✓ Deshalb ist es wichtig, großen Modellen wie Opus präzise Anweisungen zu geben und den Prompt vorab auf Vollständigkeit zu überprüfen (siehe *nächstes Kapitel*). Vage Anfragen wie: Schreibe mal was über Marketing führen zu langen, möglicherweise unbrauchbaren Antworten, die Zeit und Budget verschwenden.



Anthropic gibt keine exakten Daten heraus, aber nach Schätzungen verfügt das kleinere *Claude Haiku* über etwa 20 Milliarden (20 B, für *Billions*) Parameter im Vergleich zu rund 2.000 Milliarden (2 T für *Tausend*, also 2000 B) für *Claude Opus*. Opus ist damit zirka 100-mal größer, was höhere Eloquenz und Kapazitäten für anspruchsvolle Anwendungen erklärt.

Nicht alle Produkte werden veröffentlicht: Das **Claude Mythos** genannte Modell soll so mächtig sein, dass Anthropic es im Frühjahr 2026 bewusst unter Verschluss hält. Grund ist, dass es auf Anhieb zahlreiche Sicherheitslücken (*Zero-Day*) finden und mit *Exploit*-Code knacken kann. Im *Project Glasswing* erhalten

Big-Tech-Unternehmen die Möglichkeit, ihre Systeme zu prüfen. Dies zeigt, welche Verantwortung Produzenten tragen und welche Risiken KI birgt. »Fable 5«, eine eingeschränkte Mythos-Variante, erscheint im Sommer 2026. Mein Freund Dr. Marvin Weigand spricht von »Alien Technology«. Tatsächlich übertrifft Fable sogar Opus deutlich. Doch unsere Freude ist kurz: Nach wenigen Tagen auf US-Regierungsanweisung ohne Vorwarnung gestoppt, kam es erst Wochen später zurück. Ein herber Vertrauensverlust. Mehr dazu in *Teil III* dieses Buchs.

Die Qual der Wahl beim KI-Modell

Den Möglichkeiten stehen die Anforderungen gegenüber, und die müssen Sie selbst für sich aufstellen. Fragen Sie sich: *Wie anpassungsfähig ist die KI, ist sie technisch kompatibel mit den Zielen und welche Kosten kommen auf Sie zu?* Hier müssen Sie Kompromisse finden, denn Sie bekommen, wofür Sie zahlen:

- ✓ Es gibt kostengünstige Einstiegsmodelle, die langsam sind, in ihrer Eingabe und ihrem Umfang beschränkt sind (*Kontext*) oder Ihre Daten ungeschützt verarbeiten.
- ✓ Und es gibt leistungsfähige Spezialisten, denen Sie alles und selbst große Mengen anvertrauen können und die superschnell mit umfassenden Antworten reagieren.
- ✓ Sie können kostenlose Modelle zum Selbstinstallieren (*Kapitel 8*) haben, die sehr leistungsfähig sind, dann kostet Sie der Betrieb einiges und der Aufwand steigt.

Womöglich müssen für berufliche Zwecke Betriebsrat und Datenschutz mitreden.



Kontextlänge beschreibt, wie viele Token die KI bei der Eingabe verarbeiten kann. Bei Gemini können Sie längere Prompts mit mehr Dateien eingeben als bei ChatGPT.

Herausforderungen analysieren

Bevor Sie mit einem größeren Projekt starten oder wiederkehrende Aufgaben einer KI übertragen, sollten Sie die bevorstehenden Herausforderungen systematisch analysieren und sich auch in Bezug auf verwendete Daten rückversichern.

So gehen Sie am besten vor:

1. **Aufgaben definieren:** Stellen Sie schon im Vorfeld Ihrer Aktionen fest, welche Ziele mit der Eingabe erreicht werden sollen.
2. **Daten identifizieren:** Soll die Eingabe persönliche Daten enthalten, spielen Datenschutz oder Ethik (*Kapitel 11*) mit.
3. **Kapazitäten bewerten:** Rechenleistung und Speicher spielen besonders bei größeren Aufgaben eine Rolle. Bewerten Sie Ihre Ziele und wägen anfallende Kosten ab.

Kriterien für die Auswahl eines Large Language Models

Einige wichtige Kriterien für die Auswahl der richtigen KI sind:

- ✓ **Bedienfreundlichkeit:** Sind Einarbeitung und Erlernen von Befehlen notwendig?
- ✓ **Spezifikation:** Ist das Modell allgemein generisch oder spezialisiert?
- ✓ **Leistungsvermögen:** Wie umfangreich sind das Wissen oder die Fähigkeiten?
- ✓ **Geschwindigkeit:** Wie schnell wird auf Eingaben auch bei großen Daten reagiert?
- ✓ **Rechtliches und Moral:** Wie steht es um Datenschutz oder antrainierte Ethik?
- ✓ **Kosten:** Wie viel kostet eine Anfrage oder die monatlich pauschale Nutzung?
- ✓ **Downloadfähigkeit:** Spielt es auf fremden Servern oder ist lokal installierbar?
- ✓ **Open Source / Open Weights:** Ist das Modell offen oder braucht man Lizenzen?
- ✓ **Integration:** Läuft die KI in einem Produkt oder ist sie integrationsfähig (*API*)?



Die besten Ergebnisse erhalten Sie oft, wenn Sie KI-Werkzeuge kombinieren!

Orientierung und Funktionen

Los geht's! Picken Sie sich einen Favoriten heraus und versuchen Sie es einmal. Ich habe beispielhaft ChatGPT ausgewählt, weil es derzeit das meistgenutzte Produkt ist.

Registrieren: Wie Sie sich beim Chatbot anmelden

Sie können ChatGPT in Grundzügen auch ohne Registrierung ausprobieren. Doch das hält nicht lange. So kommen Sie zu einem (zunächst kostenlosen) Konto:

1. Als Erstes rufen Sie <https://chatgpt.com> auf und wählen **KOSTENLOS REGISTRIEREN**.
 - Sie können sich zwar mit einem Konto von **GOOGLE** oder **APPLE** oder mit **SMARTPHONE**, also Ihrer Handynummer, registrieren. Dann sind die nächsten Schritte nicht nötig. Doch gelangt der Google-, Apple- oder andere Zugang in falsche Hände, können sich Fremde auch hier einloggen.
 - Ich empfehle den Weg der E-Mail mit anonymer Weiterleitungsadresse.

- Grundsätzlich sollten Sie ein sicheres, also möglichst 16 Zeichen langes Passwort wählen und die 2-Faktor-Authentifizierung einrichten, um mehr Kontrolle über Ihre Daten zu haben.
 - Wie das alles jeweils geht, erfahren Sie in *Kapitel 11* unter *Drei Schritte für mehr Sicherheit bei Online-Logins*.
2. Geben Sie Ihre (hoffentlich weitergeleitete) E-MAIL-ADRESSE und ein PASSWORT ein. Mehr brauchen Sie zunächst nicht.
 3. Sie erhalten eine E-Mail, in der Sie den Link E-MAIL-ADRESSE BESTÄTIGEN anklicken müssen, um zu beweisen, dass alles funktioniert.
 4. Nun verlangt ChatGPT noch Ihren Namen und ein Geburtsdatum, in einigen Fällen auch den Namen einer Organisation.
 5. Nach Eingabe der Daten sind Sie eingeloggt und können mit dem Chat loslegen. Bei künftigen Logins müssen Sie vermutlich ein paar kleine Verifizierungsspiele (»Bestätige, dass du ein Mensch bist«) durchmachen.



Nutzen Sie zur Registrierung einen **E-Mail-Alias**, ein **sicheres Passwort** und aktivieren Sie die **2-Faktor-Authentifizierung** (auch *Multi-Faktor-Authentifizierung*, *MFA*, genannt). Wie Sie Ihre Accounts damit absichern, lesen Sie in *Kapitel 11*.

Funktionsübersicht: Wo finde ich was?

Ein paar Grundfunktionen haben Sie bereits kennengelernt, jetzt wird es Zeit, sich zu orientieren und mehr Funktionen kennenzulernen. Bewusst wähle ich hier ChatGPT, doch viele anderen Produkte sind ähnlich aufgebaut. Wenn Sie mit einem klarkommen, werden Sie die anderen schnell beherrschen.

Das Chatfenster



Für jedes Thema oder jedes neue Anliegen sollten Sie einen neuen Chat starten, damit der Chatbot sich nicht auf frühere Fragen und Antworten bezieht.

Ganz oben in den Ecken finden Sie Menüs und Schaltflächen, die sich abhängig von der Bildschirmgröße etwas anders darstellen:

- ✓ Im Chatfenster können Sie oben links auf den Schriftzug CHATGPT tippen. Hier wählen Sie, mit welchem Sprachmodell Sie chatten.
- ✓ GRUPPENCHAT STARTEN: Sie teilen einen LINK und können mit mehreren anderen Personen in einem gemeinsamen Chat die KI befragen.

- ✓ TEMPORÄREN CHAT AKTIVIEREN (gestrichelte SPRECHBLASE): vergleichbar mit dem *privaten Tab* im Browser. Erscheint nicht im Verlauf und wird nicht zum Training der Modelle verwendet (für normale Chats können Sie das abschalten). Es wird dennoch eine Kopie bis zu 30 Tage aufbewahrt.
- ✓ Links in der **Seitenleiste** (*Abbildung 1.2*) sind einige Menüpunkte sichtbar:
 - Ein NEUER CHAT startet mit dem STIFT-Symbol, andere Produkte nennen es *Thread* (Diskussionsstrang) oder haben als Symbol ein +.
 - Mit der LUPE können Sie in einer Suchmaske frühere CHATS SUCHEN.
 - Die FOTO-Icons bringen Sie zur *Bibliothek*, die ihre generierten BILDER zeigt.
- ✓ Weitere Menüpunkte erscheinen, wenn Sie mit der Maus die TRENNLINIE der Seitenleiste oder das LOGO oben links anklicken:
- ✓ Unter APPS verbinden Sie verschiedene Drittanbieter-Dienste, zum Beispiel *Kleinanzeigen.de*, *Apple Music* oder *Google Drive*. Der Katalog wächst laufend. Was die Verbindung kann, steht in der jeweiligen Beschreibung – im Fall von KLEINANZEIGEN beispielsweise nur nach bestimmten Sachen suchen, die Sie mit Durchsuche Kleinanzeigen einleiten. Bei anderen können Sie auf ihre externen Daten zugreifen. Entsprechend vorsichtig sollten Sie damit umgehen!
- ✓ Im Gegensatz zu Apps sind GPTs (unterhalb von APPS) angepasste Versionen von ChatGPT mit eigener Persönlichkeit, eigenem Wissen oder Aufgaben. Darunter sind etwa ein Marketing-Coach und ein Coding-Helper. Diese Funktion dient dazu, ChatGPT nicht jedes Mal eine neue Rolle (siehe *Kapitel 2*) zuweisen zu müssen, ist jedoch ein wenig eingeschlafen. Langfristig wird diese Funktion in *Workspace-Agenten* überführt, die bereits in den Firmentarifen verfügbar sind (*Kapitel 7*).
- ✓ CODEX ist ein Programmier-Agent von OpenAI, der Aufgaben übernehmen kann, etwa Dateien ändern oder Befehle ausführen. Er ist eng mit ChatGPT und anderem Entwicklungswerkzeug verbunden und arbeitet in einer kontrollierten Umgebung.



Wie Sie KI-Agenten Zugriff auf die Festplatte und Ihre Dateien geben, erfahren Sie in *Kapitel 7*. Dass das so weit hinten im Buch thematisiert wird, hat einen Grund: Sicherheit. Besser, Sie lernen erst die Grundlagen. Ich habe Sie gewarnt!

- ✓ DEEP RESEARCH erkläre ich Ihnen *im nächsten Kapitel*.
- ✓ Bei PROJEKTE können Sie *Chats* zu einem bestimmten Thema gruppieren, beispielsweise zu einer Hausarbeit oder einer Urlaubsreise. Dort teilen Sie auch spezifische *Dateien* und *Hinweise* mit ChatGPT, die jeweils eine Rolle spielen sollen. Beim Anlegen können Sie einmalig festlegen, ob Chats innerhalb des Projekts auf ihre kompletten ERINNERUNGEN zugreifen können oder nicht.

- ✓ Unter AKTUELLE listet ChatGPT kürzlich geführte Chats auf:
 - Passende Beschriftungstitel wählt die KI und sortiert sie chronologisch.
 - Sie können darauftippen und finden im Dreipunkt-Menü die Funktionen UMBENENNEN, CHAT ANHEFTEN (damit er in der Liste bleibt), ARCHIVIEREN (das ARCHIV erreichen Sie nur über die EINSTELLUNGEN) oder LÖSCHEN.
 - Später werden Sie Chats vielleicht auch GEMEINSAM NUTZEN, also für andere freigeben. Hier ist ein schneller Link dorthin versteckt.
 - Hier finden Sie später alle verwendeten GPTs oder Workspace-Agenten, die Sie bei Bedarf IN DER SEITENLEISTE BELASSEN oder AUSBLENDEN können.
 - Unten in der Seitenleiste können Sie Geld einwerfen, sprich mit TARIF UPGRADEN ein kostenpflichtiges Abonnement abschließen.
- ✓ GEMEINSAM NUTZEN: Sie können einen kompletten Chat anderen Personen zur Verfügung stellen, indem Sie den LINK KOPIEREN und zum Beispiel per Mail versenden. Das ist etwa nützlich, um Informationen über eine Reiseplanung oder mehrere generierte Texte weiterzureichen. Das erspart es Ihnen, einzelne Informationen daraus zu kopieren und zu verschicken. Enthält der Chat generierte Bilder, werden diese zum HERUNTERLADEN angeboten. Im Gegensatz zum GRUPPENCHAT können andere hier nur mitlesen.



Geben Sie in allen geteilten Chats besonders acht, was Sie schon früher im Chatverlauf eingegeben haben, da diese Nachrichten alle ebenfalls freigegeben werden. Persönliche Informationen und Daten wie Passwörter sind ohnehin tabu. Mehr zu Vorkehrungen und Datenschutz in *Kapitel 11*.

Aktionen während des Chats

Unter Ihren Prompts und den Antworten von ChatGPT werden Sie bereits verschiedene Symbole entdeckt haben. Die möchte ich Ihnen kurz erläutern, damit Sie sie besser bedienen können:

- ✓ Sie können eine einmal abgesendete **Prompt-Nachricht** durch DARÜBERFAHREN mit dem erscheinenden Knopf KOPIEREN oder über das STIFT-Symbol noch einmal verändern. Die Antwort der KI wird dann daran angepasst.
- ✓ Außerdem lässt sich die Antwort, während sie geschrieben wird, mit dem STOPP-Kästchen abbrechen, falls sie in die falsche Richtung führt.
- ✓ MARKIEREN Sie die Passage einer Antwort, um sich mit ANFÜHRUNGSZEICHEN (**ChatGPT fragen**) darauf zu beziehen und Nachfragen zu stellen.

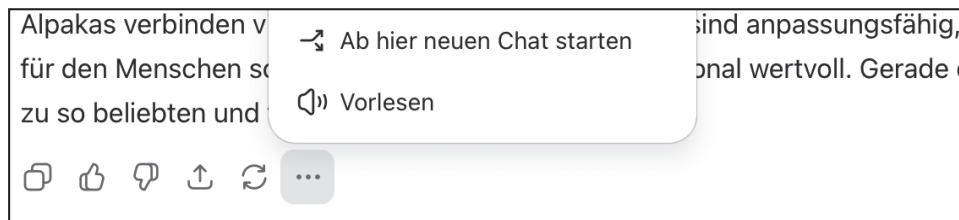


Abbildung 1.3: Unter der Antwort des Chatbots finden Sie Symbole, mit denen Sie unterschiedliche Aktionen auslösen können. (Screenshot: chatgpt.com)

- ✓ Diese Möglichkeiten finden Sie direkt unter den KI-Antworten (*Abbildung 1.3*):
- **KOPIEREN:** Die Antwort wird in die Zwischenablage kopiert und Sie können sie an anderer Stelle einfügen, zum Beispiel in eine E-Mail.
 - **DAUMEN HOCH/RUNTER (GUTE oder SCHLECHTE REAKTION):** Gibt OpenAI eine Rückmeldung für künftige Trainings, wie Sie die Antwort bewerten.
 - **WEITERGEBEN:** Sie können einen kompletten Chat anderen Personen zur Verfügung stellen, indem Sie den **LINK KOPIEREN** und zum Beispiel per Mail versenden. Das wäre nützlich, um Informationen über eine Reiseplanung oder mehrere generierte Texte weiterzureichen, und erspart es Ihnen, einzelne Informationen daraus zu kopieren und zu verschicken. Enthält der Chat generierte Bilder, werden diese zum **HERUNTERLADEN** angeboten.
 - **PFEILE IM KREIS (ERNEUT VERSUCHEN):** Lässt ChatGPT die jeweilige Antwort noch einmal mit einem anderen oder demselben Modell generieren, als wäre sie nie gekommen. Anschließend können Sie mit Pfeiltasten zwischen den unterschiedlichen Antwortvarianten und Folgeantworten blättern. Das liefert eine neue Perspektive und andere Blickwinkel zu einem Thema.
 - **PFEILE NACH OBEN UND UNTEN (AB HIER NEUEN CHAT STARTEN):** Wenn Sie sinnbildlich einmal falsch abgebogen sind und an einer früheren Stelle eines Chats anknüpfen möchten, können Sie das jeweils mit diesen Symbolen. Es öffnet sich ein Teilgespräch in einem neuen Fenster. Diese Funktion ist mir bisher nur bei ChatGPT begegnet. Mehr dazu in *Kapitel 2*.
 - **LAUTSPRECHER (VORLESEN):** Die Sprachausgabe liest die Antwort vor.
 - **QUELLEN:** In einigen Situationen erscheint ein Knopf, mit dem Sie eine Seitenleiste mit Quellenangaben für Antworten ein- und ausblenden.
 - Google Gemini hat an dieser Stelle noch die Option **ANTWORT ÜBERPRÜFEN**, die einen Abgleich mit der Google-Suche durchführt und für mehr Faktentreue sorgen soll. Inhaltlich belegte Stellen werden *grün* in der KI-Antwort markiert, Abweichungen *orange*farben. Absolut verlässlich ist das nie! Siehe *Kapitel 9*.

Die Kontoeinstellungen

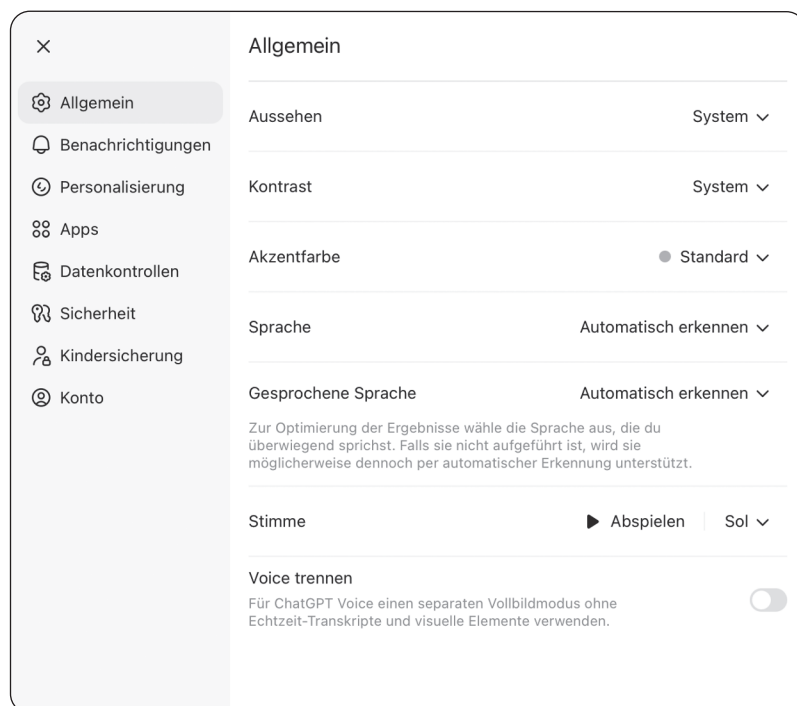


Abbildung 1.4: ChatGPT-Einstellungen (Screenshot: chatgpt.com)

- ✓ Die **Kontoeinstellungen** finden Sie auf der Webseite in einem bunten Kreis mit Ihren Initialen (oder dem, was ChatGPT anhand Ihrer Mailadresse dafür hält).
- **PERSONALISIERUNG:** Hinweise zu Ihrer Person (Standort, Tätigkeit, Hobbys, bevorzugte Themen und Zielen) oder Regeln, die immer gelten sollen (zum Beispiel formelle oder ungezwungene Sprache, Länge der Antworten, wie Sie angesprochen werden möchten oder die Neutralität der Reaktionen), legen Sie hier fest. Im jeweiligen Eingabefeld stehen Beispiele dafür. Die **ERINNERUNG**-Funktion wird *in diesem Kapitel* noch thematisiert. Wägen Sie stets Nutzen und Datenschutz ab!
- Auf Arbeit oder gemeinsam genutzten Geräten nicht vergessen: **ABMELDEN**.
- ✓ **EINSTELLUNGEN** beinhaltet bei **AUSSEHEN** die schicke dunkle Darstellung (*Dark Mode*) und die **STIMMEN** für die Sprachausgabe, bei **DATENKONTROLLE** unter anderem **ARCHIVIERTE CHATS** und die Option **DAS MODELL FÜR ALLE VERBESSERN** (Ihre Daten werden fürs Training verwendet, besser abstellen!).
- ✓ Spätestens jetzt ist unter **SICHERHEIT** die **MULTIFAKTOR-AUTHENTIFIZIERUNG** (MFA) ratsam: Sorgen Sie für Sicherheit (*siehe Kapitel 11: Drei Schritte für mehr Sicherheit bei Online-Logins*). Sollten Sie unsicher sein, ob Sie sich auf anderen Geräten abgemeldet haben, holt **ÜBERALL ABMELDEN** das nach.



Zu den meisten Chatbots gibt es kostenlose **offizielle Apps**, die Sie auf dem Computer, Tablet oder Smartphone laden können. Achten Sie darauf, nur offizielle Apps zu laden, die Sie auf der Homepage des Herstellers finden. Hier haben Sie die KI bei Fragen immer griffbereit (am Desktop über eine TASTENKOMBINATION), Sie können zusätzlich in den Sprachmodus wechseln, Dateien bequem ins Chatfenster ziehen oder Ihren Bildschirminhalt freigeben, um konkret nach Inhalten zu fragen.

KI-Systeme im Wettstreit

Wie bei anderen Anwendungen gilt bei der KI der Grundsatz: Sie machen alle ungefähr das Gleiche, aber unterschiedlich gut und unterschiedlich teuer. Aber welches KI-Modell ist für welchen Zweck geeignet und welche Modelle behaupten nur, richtig gut zu sein? Es gibt grundsätzliche Unterschiede zwischen Modellen und darauf aufbauenden Diensten. Einige bieten echten Mehrwert, während andere nur Grundfunktionen *umverpacken*. Informieren Sie sich über die Basisfunktionen, um Mehrwert und Datenschutzrisiken einschätzen zu können und gegebenenfalls Kosten zu sparen.

Benchmarks: Die »Stiftung Warentest« der KI-Technologie

Die Entwicklung schreitet in atemberaubendem Tempo voran. Ständig kommen neue Versionen und Verbesserungen auf den Markt und alte verschwinden. Diese rasante Evolution führt dazu, dass die Leistungsfähigkeit und Anwendungsmöglichkeiten sich nahezu täglich erweitern und kaum jemand auf dem aktuellen Stand bleiben kann, gäbe es da nicht Benchmarks, die meist unabhängig Produkte gegeneinander abwägen.



Sogenannte *Benchmarks* sind in der Informatik eine Art Olympische Spiele für Geräte und Programme. Alle treten gegeneinander in ihrer Kategorie an und am Ende stehen die Sieger auf dem Treppchen. Jeder weiß dann, wer am besten abgeschnitten hat. Es gibt sehr viele solcher Tests und Kriterien.

Große Sprachmodelle sehr intransparent

Weil das gesammelte Wissen der Welt in ihren Trainingsdaten steckt, können große Sprachmodelle wie ChatGPT oder Gemini sehr viel. Es ist aber schwierig herauszufinden, was sie wirklich auf dem Kasten haben und wo ihre Schwächen liegen. Nicht nur für Laien sind sie wie eine *Blackbox*, die man von außen nicht einsehen kann. Selbst Experten tun sich schwer:

- ✓ Vergleichbare Leistungstests und Benchmarks für KI sind schwierig aufzustellen. Die Fragenkataloge veralten schnell oder werden schlicht eintrainiert.
- ✓ Trotz alternativer, laufend aktualisierter Tests lässt sich die Denkweise von Modellen nicht immer nachvollziehen und überprüfen.
- ✓ Forschende fordern, dass Systeme ihre Entscheidungen besser erklären und stets mit Quellen nachweisen, damit Menschen diese überprüfen können.



Viele KI-Funktionen, die Hersteller ankündigen, führen sie zunächst außerhalb der EU ein, um juristische Risiken zu vermeiden oder die Nachfrage zu testen. Informieren Sie sich über politischer Rahmenbedingungen (*Teil 5*).

Kostenlose Vergleichsportale

Glücklicherweise machen trotz aller Widrigkeiten sehr viele Forschende ihren Job ausgezeichnet und präsentieren regelmäßig aktualisierte Vergleiche:

- ✓ **LM-Arena:** eine Blindverkostung der LLMs. Erst nach ihrer Bewertung sehen Sie, welche Modelle Sie vor sich haben. Anhand der Abstimmung entsteht regelmäßig eine **viel beachtete Rangliste**: <https://lmarena.ai>
- ✓ **Lumigator:** Um trotz Qual der Wahl nicht dem Hype zu verfallen und Argumente abzuwägen, die für oder gegen ein Large Language Model sprechen, führt Mozilla den Auswahlassistenten ein: <https://www.mozilla.ai/lumigator>
- ✓ **Buzzwoo:** Testen Sie mehr als 60 Large Language Models, um die **Aktualität, Fähigkeiten und Kosten** zu bewerten. In einer Testarena treten Sprachmodelle gegeneinander an: <https://www.buzzwoo.de/kuenstliche-intelligenz>
- ✓ **Hugging Face:** Die führende Plattform und Community im Bereich des maschinellen Lernens ist ein zentraler Ort zum Teilen, Entdecken und Experimentieren mit Open-Source-Modellen: <https://huggingface.co>

Kostenübersicht

Nachdem Sie in einige Funktionen eingetaucht sind und etwas von verschiedenen Abonnementstufen gelesen haben, fragen Sie sich sicher: Was kostet mich das alles?

Preisdruck und Intransparenz

Die Preisstaffelungen (*Pläne*) sind so unterschiedlich wie verwirrend:

- ✓ Es gibt jeweils eine kostenlose Variante, für die Sie höchstens eine kurze Anmeldung mit der E-Mail-Adresse benötigen. Damit geraten Sie jedoch schnell an Ihre Grenzen, wenn die Anzahl der Antworten, die Länge der Fragen oder andere Dinge eingeschränkt und stark verlangsamt sind.
- ✓ Mehr Möglichkeiten und Zugriff auf viele Varianten der Produktfamilien oder Funktionen wie *Deep Research* erhalten Sie mit einem Abonnement, das es in der Regel um die 25–30 Euro pro Monat gibt und das Sie auch nur kurzzeitig nutzen können – so wie bei einem Möbeltransporter oder Autoanhänger, den Sie nur bei besonderen Transporten benötigen. Dafür gewähren die Unternehmen rund 20 Prozent Rabatt, wenn Sie sich für ein Jahr an sie binden.

- ✓ Diese Pro- oder Plus-Abonnements können sie manchmal zwischen zwei und vier Wochen kostenlos testen und sollten bei Bedarf rechtzeitig kündigen. Bei amerikanischen Angeboten geht das meist schon während der Laufzeit, wobei der Probezeitraum dennoch komplett ausläuft.



Kostenlose Angebote bedeuten häufig, dass Sie mit Ihren Daten bezahlen:

- ✓ Nutzen Sie nicht unüberlegt dahergelaufene Chatbots für persönliche Aufgaben oder private Chats, und schon gar nicht für berufliche oder vertrauliche Zwecke! Mehr zu solchen Risiken finden Sie in *Teil 4*.
- ✓ Die Nutzungsbedingungen unterscheiden sich teilweise maßgeblich: So kann es sein, dass Sie bei der kostenlosen Variante zustimmen, dass Ihre Daten für künftige KI-Trainings verwendet werden dürfen.
- ✓ Viele neue Funktionen sind anfangs den größeren, teureren Paketen vorbehalten. Sie werden erst später für andere freigeschaltet.

Die Tarifstrukturen funktionieren meist wie Datenvolumen beim Handyvertrag:

1. Je nach Tarifstufe gelten verschiedene Prompt- und Funktionslimits.
2. Sie starten mit einem günstigen Einstiegtarif, kommen eine Weile gut zurecht, bis die Benachrichtigung erfolgt, dass Ihr Kontingent aufgebraucht ist.
3. Dann können Sie entweder abwarten und bis zum Monatsanfang im Schneckentempo weiter prompten. Das ist jedoch nicht in jedem Projekt einfach umzusetzen.
4. Um direkt weiterzumachen, können Sie alternativ jederzeit auf eine höhere Tarifstufe wechseln, oder Sie buchen ein Datenpaket nach, wie Sie es vom Smartphone gewohnt sind (meist nach der Standard-API-Rate pro *Token*, siehe unten).



Manchmal lohnt sich einfach ein Zweit-Account mit zusätzlichem Guthaben.

Werbepause im Prompt

Seit 2026 spielt ChatGPT Werbung aus. Obwohl die Testphase zum Zeitpunkt, als ich das schreibe, auf die USA beschränkt ist, können Sie mit einer Ausweitung zu uns rechnen. Allen Beteuerungen zum Trotz muss schließlich Geld reinkommen:

- ✓ Betroffen sind angemeldete, volljährige Nutzende der kostenlosen Variante sowie des günstigen Abos »ChatGPT Go«. Kostenpflichtige Tarife wie Plus, Pro, Business oder Enterprise bleiben (zunächst) werbefrei.

- ✓ Anzeigen erscheinen als Banner am Ende von Chat-Antworten, sind klar als »gesponsert« gekennzeichnet und beeinflussen laut OpenAI die KI-Generierung nicht. Es findet eine Personalisierung basierend auf Chat-Themen und Vorlieben statt, was über die EINSTELLUNGEN kontrolliert oder deaktiviert werden kann. Wer keine Werbung möchte, kann diese AUSBLENDEN. Im Gegenzug werden Gratisinhalte stark eingeschränkt.
- ✓ Werbung werde bei sensiblen Themen wie Gesundheit, Politik oder Finanzen vermieden, und OpenAI verkauft angeblich keine Nutzerdaten an Werbetreibende.
- ✓ Anthropic (Claude) und Google (Gemini) kritisieren öffentlich OpenAIs Schritt. Sie betonen in großen Kampagnen ihre werbefreien Modelle, wobei Google bei den KI-Übersichten in der Suchmaschine durchaus Werbeanzeigen schaltet. Perplexity hat ein 2024 eingeführtes testweises Antworten-Sponsoring aus Sorge vor Vertrauensverlust zurückgenommen, um sich gegen ChatGPT zu positionieren. Die Frage um die Monetarisierung in KI-Chatbots bleibt weiterhin in der Diskussion.

Abonnementpreise gängiger Chatbots im Vergleich

Tabelle 1.3 bietet Ihnen die Übersicht der Tarife gängiger Chatbots, damit Sie einen Vergleich haben. Angesichts hoher Entwicklungskosten und der wirtschaftlichen Lage könnten die Preise seit Veröffentlichung schon längst wieder gestiegen sein.

Produkttarif	Monatsabo (Jahresabo)	Besonderheiten
OpenAI Go	8,00 EUR	Go als niedrigschwelliger Einstieg nach dem Gratistarif, Plus für alltäglichen Gebrauch. Im Pro-Tarif sind Obergrenzen für Modelle und Funktionen 5-fach oder 20-fach höher.
OpenAI Plus	23,00 EUR	
OpenAI Pro	103,00 oder 229,00 EUR	
Anthropic Claude Pro	18,00 (15,00) EUR	Limit gestaffelt: Max enthält 5-fache oder 20-fache Zugriffe und neueste Funktionen.
Anthropic Claude Max	90,00 oder 180,00 EUR	
Google Gemini AI Plus	7,99 EUR	Vergrößern Speicher für Gmail, Fotos und Drive. Plus-Tarif mit viermal höheren Limits als Gratis, Pro 20-fach. Ultra für Filmerstellung interessant.
Google Gemini AI Pro	21,99 EUR	
Google Gemini AI Ultra	274,99 EUR	
Perplexity Pro	21,49 (17,86) EUR	Neben Sonar auch Gemini, Claude und ChatGPT, Computer und Council in Max.
Perplexity Max	214,37 (178,65) EUR	
Microsoft 365 Single	10,00 (8,25) EUR	Mit Word, Excel, PowerPoint, Outlook sowie Copilot in Web und Apps. Bei Family und Premium hat nur die abonnierende Person Zugriff auf Copilot.
Microsoft 365 Family	13,00 (10,75) EUR	
Microsoft 365 Premium	22,00 (18,25) EUR	

Tabelle 1.3: Tarife gängiger Chatbots mit Mehrwertsteuer (Stand: Frühjahr 2026)



Suchen Sie vorab nach Angeboten bei <https://mydealz.de> oder auf den Social-Media-Accounts, denn teilweise werden Jahresabonnements verschenkt oder es gibt Vergünstigungen für Studierende, Lehrkräfte und andere Gruppen.

Team-Verträge für Organisationen

Neben den Varianten mit höheren Nutzungslimits bieten fast alle Unternehmen *Team-Verträge* und für große Firmen individuelle *Enterprise-Lizenzen*. Bei Perplexity heißt der Tarif *Enterprise Pro* bzw. *Max* und gilt für bis zu 250 Mitarbeitende.



Für Bildungszwecke und Non-Profit-Organisationen gibt es in vielen Fällen Rabatte, Campuslizenzen oder sogar Gratisaccounts. Danach sollten Sie Ausschau halten! Als Nachweis genügt häufig eine Mail-Adresse oder ein PDF.

Im Dutzend teurer: Was sie bringen

Startups oder kleine und mittelständische Unternehmen sollen Teams bilden. Das hat Vor- und Nachteile, die Sie gegeneinander abwägen müssen:

- ✓ Keine Quittungen sammeln: Konten werden zentral verwaltet und erhalten eine gemeinsame Abrechnung. Das ist zumindest mal eine Arbeitserleichterung.
- ✓ Preislich kommen Sie selten besser weg: Die Tarife liegen in der Regel pro Person 5 US-Dollar über den regulären. Kostspieliger sind Gemini Business oder Microsoft 365 Copilot Business, die keine Wahl lassen, da nur sie mit den Office-Suites laufen.
- ✓ Übliche Privatsphäre: Einsicht in private Chats haben nur die jeweiligen Nutzerkonten (und wie üblich wohl auch die Betreiber). Einige Anbieter versprechen, die Daten vom Training auszuschließen.

Gemeinsam an Projekten arbeiten

Es gibt verschiedene Konzepte, um mit mehreren Personen an Chats und Recherchen zu arbeiten oder zentral Dokumente abzulegen (oft eignen sich Tools wie Notion oder Confluence und NotebookLM, siehe *Kapitel 6*, besser):

- ✓ Bei Claude können Sie Dokumente als Basis gemeinsamer Projekte hochladen. Jedes Projekt darf 200.000 Token groß sein, was etwas mehr ist als dieses Buch.
- ✓ ChatGPT erlaubt eigene GPTs beziehungsweise Workspace-Agenten, die nur mit dem Team geteilt werden können, um etwa interne Wikis und Helpdesks zu ersetzen.
- ✓ Sie können bei ChatGPT einzelne Chats als Gruppenchats freigeben.
- ✓ Der Zugriff über die API-Programmschnittstelle, zum Beispiel aus Anwendungen heraus, wird in allen Fällen separat über sogenannte *Token* abgerechnet.

Token: mundgerechte KI-Häppchen

Cloud-Modelle rechnen *API*-Zugriffe nach **Token** (*Zeichen, Münzen, Einheiten*) ab:

- ✓ Token entsprechen nicht automatisch geschriebenen Wörtern, sondern eher Silben. (Jedes Modell zählt anders: Bei OpenAI sind 1.000 Token etwa 750 Wörter.)
 - Berechnet werden Ihre Eingabe (*Input Token*) und die KI-Ausgabe (*Output*). Claude Sonnet kostet pro Million Token im Input 3 US-Dollar, im Output 15. Ganz grob gesagt kostet Sie eine DIN-A4-Seite Korrespondenz 0,01 Cent.
 - Wiederkehrende Information, zum Beispiel über Sie aus den EINSTELLUNGEN, wird dabei meist zwischengespeichert (*Prompt Caching*) und daher rabattiert.
- ✓ Bei KI-Modellen sparen Sie »Kraftstoff«, wenn Sie nicht immer mit dem SUV-Fahrzeug aufbrechen, sondern jeweils geeignete Modelle ohne Ballast wählen.

Nutzen Sie Onlinerechner, um zu vergleichen, was am Ende auf der Rechnung stehen dürfte: <https://deinkikompass.de/openai-api-rechner>

KI überall

Nicht für jeden Anwendungszweck müssen Sie die Kreditkarte zücken. Viele Dienste sind bereits im Haus. Ohne umfassend jede einzelne KI-Funktion in allen möglichen Apps aufzuzählen, möchte ich Ihnen ein paar Möglichkeiten aufzeigen, damit Sie sich das eine oder andere Abonnement sparen können, wenn Sie die Augen aufhalten.

KI-Funktionen in Ihren Geräten

In Ihrem PC und Ihrem Smartphone steckt mehr KI, als Sie vermutlich denken:

- ✓ Bei einem aktuellen **Windows**-Computer ist Copilot bereits als App vorinstalliert und in vielen Programmen von *Paint* (siehe *Kapitel 3*) bis zu *Office*-Anwendungen integriert. Klicken Sie dazu einfach auf das **COPILOT**-Symbol in der Taskleiste oder öffnen Sie die **COPILOT**-APP über das **STARTMENÜ**. Einige Dienste sind dabei inklusive, viele Leistungen müssen Sie jedoch extra bezahlen.
- ✓ Ein vollständiges Apple Intelligence am **Mac** lässt, während ich das hier schreibe, noch auf sich warten. Die Integration von *KI-Schreibtools* in *Mail* oder der *Notizen*-App kann in vielen Fällen bereits nützlich sein, um E-Mails zusammenzufassen oder Diktate zu transkribieren. Die Funktionen finden Sie in den meisten Anwendungen direkt im jeweiligen Textfeld über die Taste mit dem **STERN-** UND **DEM STIFT**-Symbol oder über **BEARBEITEN > SCHREIBTOOLS**. In Apps wie *Fotos* arbeitet KI im Hintergrund für Sie. Die Mac-App *Image Playground* kann simple Bilder generieren. Irgendwann wird auch hierzulande auf neueren Geräten *Siri* smart.

- ✓ In **Linux**-Distributionen wie Ubuntu nutzen Sie lokale Sprachmodelle: zur Transkription, als Helfer bei der Fehlersuche oder zur Sprachsteuerung. Alle Funktionen kommen optional als Snap-Paket über das Software-Center.
- ✓ **Android-Smartphones** ersetzen seit 2025 den klassischen Google Assistant durch Googles KI-Assistenten *Gemini*, der Sprachbefehle versteht, Bilder analysiert und direkt in Apps wie Gmail oder Google Fotos eingreift. Sie erreichen ihn per langem Druck auf die Home-Taste oder durch den Ausruf Hey Google.
- ✓ Auf **Samsung**-Geräten kommt noch *Galaxy AI* obendrauf: Funktionen wie automatische Gesprächsmitschriften, Übersetzungen in Echtzeit oder smarte Notizassistenten sind laut Samsung dauerhaft kostenlos. Zu finden sind sie jeweils direkt in der *Telefon*-, *Notizen*- oder *Rekorder*-App über das kleine STERNCHEN-Symbol.

KI in Ihrem Browser

In Internet-Browsern, Suchmaschinen und bekannten Online-Diensten ist KI längst angekommen, ohne dass Sie meist etwas bezahlen müssen:

- ✓ **Microsoft Edge** (eine Verwandte von Google Chrome) hat *Copilot* fest in der Seitenleiste eingebaut: Webseiten zusammenfassen, Texte umformulieren oder Fragen zum geöffneten Dokument stellen geht im Browser ohne Abonnement. Klicken Sie dazu in der EDGE-LEISTE oben rechts auf das COPILOT-Symbol.
- ✓ Der **Chrome-Browser** wird mit immer mehr KI-Funktionen aufgewertet:
 - Der KI-MODUS bringt Sie direkt zu Gemini. **Google** zeigt in den Suchergebnissen meist KI-Zusammenfassungen an, die sogenannten *AI Overviews*. Einfach suchen wie gewohnt, die KI-Antwort erscheint ganz oben.
 - Google Maps versteht dank Gemini-Integration natürlichsprachige Anfragen wie Gutes veganes Restaurant mit Terrasse in der Nähe. Tippen Sie dazu in die Suchleiste. Mehr dazu in *Kapitel 5*.
 - Speichern Sie häufig genutzte Prompts (SKILLS), VERGLEICHEN Sie Produkte in mehreren Tabs und lassen Sie Gemini als Agenten (*Kapitel 7*) arbeiten.



KI ist längst nicht mehr auf herkömmliche Browser beschränkt. Einige Anbieter haben eigene wie ChatGPT (Atlas), Perplexity (Comet) und andere:

- ✓ KI-Browser bieten erweiterte agentische Funktionen, die Sie direkt auf Internetseiten ausführen können: Autonome Steuerung von Webseiten und direktes Beantworten von Fragen zu Webinhalten sind damit möglich.
- ✓ Mit Zugriff auf Mails, Kalender und Aufgaben kann die KI damit helfen.
- ✓ Die Entwicklung in diesem Bereich läuft rasant. Der Marktanteil ist jedoch gering, viele Funktionen erfordern ein bezahltes Abo. Ich glaube, dass sich KI im Browser, wie bei Chrome durchsetzt statt in separaten Anwendungen.

Angesichts wachsender Konkurrenz anderer Browserhersteller hat Mozilla KI-Funktionen in **Firefox** integriert, zum Beispiel in der SEITENLEISTE. Das lief nicht ohne Protest ab, denn viele Nutzende sind eher puristisch eingestellt, auf Datenschutz bedacht und von KI genervt. Aus diesem Grund verweist das Firefox-Team deutlich auf die Möglichkeiten, alle Funktionen zu deaktivieren.



KI-Tools finden Sie in Menüs PLUGINS oder ERWEITERUNGEN vieler Browser.

KI in Apps und im Web

In etablierten Apps, Messengern und Social Media tauchen immer mehr KI-Assistenten auf, die Funktionen der Chatbots übernehmen oder darauf zugreifen:

- ✓ Wenn Sie ein Android-Smartphone nutzen, speichern Sie mit hoher Wahrscheinlichkeit Ihre Bilder bereits in Google Fotos. Die zugehörige App integriert Ihre persönlichen Aufnahmen und zusätzlich ERSTELLUNGSTOOLS. So benötigen Sie für ein paar schnelle Effekte oder Anpassungen keinen Umweg.
- ✓ Ganz vorne dabei mit generativen KI-Funktionen ist Meta mit Facebook, Instagram und WhatsApp. Hier erhalten Sie eingeschränkten Zugriff auf den Chatbot von Meta, beispielsweise in WhatsApp als runden KNOPF und in der SUCHLEISTE (bedenken Sie, dass Sie Ihre Daten in die Cloud schicken!).
- ✓ Microsofts Karrierenetzwerk *LinkedIn* generiert Ihnen Posts, und auch Snap mit *Snapchat* und ByteDance mit *TikTok* haben generative KI integriert. Bei TikToks KI-Suite Symphony können sich Creators einen Avatar erstellen, der für sie in Videos auftritt: <https://ads.tiktok.com/creative/creativestudio/>
- ✓ Beim Einkaufen, beispielsweise auf *Amazon*, *Immobilienscout24.de*, *Kleinanzeigen.de*, läuft KI im Hintergrund, um Bilder zu optimieren, Angebote zusammenzufassen oder um beim Schreiben *Formulierungsvorschläge* zu machen.



In *Kapitel 5* finden Sie Tipps für den Alltag, die Sie ausprobieren können.

Nachdem Sie nun einen Chatbot verstehen und bedienen können und darüber hinaus KI-Funktionen an vielen Stellen entdecken, widmen Sie sich *im nächsten Kapitel* dem Prompten. Tauchen ein und lernen Sie, was der kluge Textassistent für Sie bereithält!