

Die besten KI-Tools

für Texte, Bilder, Audio, Video, Recherche,
Lernen und Workflows

» Hier geht's
direkt
zum Buch

DIE LESEPROBE

Einführung in KI-Tools

Keine Technologie hat in den letzten Jahren so rasant an Fahrt aufgenommen wie die künstliche Intelligenz. Was vor Kurzem noch nach Science-Fiction klang – Maschinen, die Texte verfassen, Bilder gestalten, Daten analysieren oder sogar programmieren – ist heute Realität und für jedermann zugänglich. Neuartige KI-Tools halten Einzug in nahezu alle Lebensbereiche: Sie verändern Lern- und Arbeitsprozesse, strukturieren Wissensarbeit neu und eröffnen Anwenderinnen und Anwendern ungeahnte kreative und produktive Möglichkeiten.

Dabei geht es längst nicht mehr nur um vereinzelte technische Spielereien oder komplexe Anwendungen für IT-Profis. Künstliche Intelligenz wirkt inzwischen tief in unseren Alltag hinein: Sie hilft beim Formulieren von Texten aller Art, schlägt passende Filme oder Produkte vor, beantwortet Fragen, fasst Informationen zusammen oder übersetzt Texte aus vielen Fremdsprachen in Sekundenschnelle. Unsere Art, zu lernen, zu kommunizieren, nach Wissen zu suchen oder Entscheidungen zu treffen, verändert sich grundlegend und oft, ohne dass wir es bewusst bemerken. KI ist damit nicht nur ein technologischer Trend, sondern ein kultureller Wendepunkt, der unser Denken, Handeln und Zusammenleben auf vielfältige Weise beeinflusst. Es ist jetzt schon abzusehen, wie groß der Vorsprung derer sein wird, die die neuen Werkzeuge mutig, reflektiert und gezielt einsetzen, und wie deutlich jene zurückfallen, die ihnen mit Ablehnung oder Unsicherheit begegnen. Das gilt im Übrigen in vielerlei Hinsicht: intellektuell, produktiv und nicht zuletzt sogar kreativ.

Was dabei oft unterschätzt und übersehen wird: Die Veränderung betrifft nicht nur spezialisierte Technologiebereiche oder große Konzerne. Viele KI-Tools beeinflussen schon heute ganz konkrete Alltagssituationen – im Büro, im Klassenzimmer, in der Werkstatt, in der Arztpraxis und sogar zu Hause. Sie unterstützen bei der Texterstellung, helfen bei der Planung verschiedenster Vorhaben, generieren Bilder, analysieren Daten, formulieren E-Mails, strukturieren Wissen und schlagen kluge Lösungen für fast alle Aufgabenstellungen vor. In vielen Berufen und beim Lernen übernehmen sie zunehmend die Rolle eines digitalen Assistenten, der rund um die Uhr verfügbar, unbegrenzt lernfähig und individuell einsetzbar ist. Wer die Potenziale dieser neuen Werkzeuge erkennt und souverän nutzt, verschafft sich nicht nur einen erheblichen Effizienzvorteil, sondern auch kreative Freiräume, mehr Überblick und oft sogar deutlich bessere Ergebnisse. KI wird

damit zur Schlüsselkompetenz einer neuen Arbeitswelt quer durch alle Branchen und Hierarchieebenen.

Doch was genau verbirgt sich hinter dem Begriff *KI-Tools*, der derzeit in aller Munde ist? Welche Arten von Anwendungen zählen dazu, wie funktionieren sie im Detail und worin unterscheiden sie sich eigentlich von herkömmlicher Software? Um das Potenzial dieser Technologie wirklich zu verstehen, lohnt sich ein Blick auf ihre Grundlagen und die Funktionsweise. Die folgenden Abschnitte und das nächste Kapitel liefern einen kompakten Einstieg in dieses Thema, klären zentrale Begriffe, zeigen, was KI-Tools ausmacht und warum sie in Beruf und Alltag immer wichtiger werden.

1.1 Was sind KI-Tools?

Um zu verstehen, was KI-Tools eigentlich sind und warum sie so wirkungsvoll eingesetzt werden können, braucht es zunächst ein grundlegendes Verständnis davon, was sich hinter dem Begriff *künstliche Intelligenz* verbirgt und wie sich KI-basierte Anwendungen von klassischer Software unterscheiden.

1.1.1 Das Grundprinzip künstlicher Intelligenz

Künstliche Intelligenz (KI) bezeichnet die Fähigkeit von Maschinen oder Programmen, Aufgaben zu lösen, die normalerweise menschliches Denken erfordern. Dazu gehört etwa das Verstehen von Sprache, das Erkennen von Bildern, das Lernen aus Beispielen oder das Treffen von Entscheidungen.

Im Unterschied zu klassischer Software folgt KI nicht nur starren Regeln, sondern kann aus Daten lernen, Muster erkennen und sich an neue Situationen anpassen. Sie ahmt dabei bestimmte Aspekte menschlicher Intelligenz nach, zum Beispiel logisches Denken, Sprachverständnis oder kreative Problemlösung. KI ist dabei kein *fühlendes Wesen*, sondern ein Werkzeug, das mit statistischen Methoden sehr wahrscheinliche und oft erstaunlich präzise Ergebnisse liefert.

KI begegnet uns heute ganz selbstverständlich im Alltag z.B. in der Autokorrektur beim Schreiben auf dem Smartphone, in Sprachassistenten wie Siri oder Alexa, in personalisierten Empfehlungen bei Netflix oder Spotify und in sogenannten *KI-Tools*, die diese neuen Fähigkeiten für bestimmte Anwendungsbereiche gezielt bündeln.

1.1.2 Abgrenzung zu traditionellen Software-Tools

Klassische Software arbeitet nach dem Prinzip fester Regeln: Sie tut genau das, was ihr zuvor von einem Programmierer vorgegeben wurde – nicht mehr und nicht weniger. Jede Funktion und jeder Ablauf sind exakt vorausgedacht und strikt geplant. Solche Programme sind zuverlässig und vorhersehbar, aber eben auch

starr: Wenn sich die Anforderungen ändern oder neue Situationen auftreten, muss der Mensch eingreifen und die Software entsprechend anpassen und erweitern.

KI-Tools funktionieren grundlegend anders. Sie basieren auf Daten statt fester Regeln und können flexibel auf neue Informationen reagieren. Ein KI-Tool kann z.B. aus vielen Beispieltextrn lernen, wie ein guter Artikel aufgebaut ist, ohne dass ihm jemand exakt vorgibt, wie dieser genau auszusehen hat. Es erkennt Muster, bildet Zusammenhänge und entwickelt eigenständig Vorschläge oder Lösungen, die nicht direkt einprogrammiert oder in anderer Weise vorgegeben wurden.

Ein weiterer wesentlicher Unterschied zwischen klassischer Software und KI-Tools liegt in ihrer Fähigkeit, mit Unschärfe und unvollständigen Eingaben umzugehen. Während herkömmliche Programme nur auf vollständige und klar definierte Eingaben reagieren können, sind KI-Systeme in der Lage, auch bei vagen oder unvollständigen Informationen passende Antworten zu generieren. Sie arbeiten mit Wahrscheinlichkeiten und orientieren sich an Mustern, die sie in großen Datenmengen erkannt haben, was sie sehr flexibel und erstaunlich anpassungsfähig macht.

Doch genau darin liegt auch ihre Schwäche: Wenn keine ausreichenden Informationen vorliegen oder ein KI-Modell kein passendes Wissen parat hat, kann es zu sogenannten *Halluzinationen* kommen, also zu frei erfundenen, aber sprachlich und inhaltlich überzeugend klingenden Aussagen. Dieser Umstand macht KI-Tools insbesondere dann fehleranfällig, wenn sie unter Unsicherheit agieren oder nur auf eine begrenzte Datenbasis zurückgreifen können. Die Fähigkeit, trotzdem eine Antwort zu liefern, ist deshalb zugleich Stärke und Risiko und erfordert ein wachsames, kritisch prüfendes Auge aufseiten der Nutzer.

Die Fähigkeiten zur Anpassung, zum Lernen und zur Interpretation machen KI-Tools zu einem ganz neuen Typ Werkzeug, das dynamischer, aber auch weniger vorhersehbar als herkömmliche Software ist. Gerade diese fehlende Vorhersehbarkeit führt zu einem anderen zentralen Problem: Als Blackbox sind KI-Systeme insofern problematisch, als dass sich ihre Ergebnisse nicht vollständig nachvollziehen lassen. Welche Daten herangezogen wurden und den größten Einfluss auf das Resultat hatten oder welchen internen Entscheidungsweg das Modell gegangen ist, bleibt in der Regel unklar.

Selbst Entwickler großer Sprachmodelle verstehen im Detail oft nicht mehr, warum das neuronale Netz, das sie für das KI-System entwickelt haben, eine bestimmte Ausgabe formuliert oder eine bestimmte Entscheidung getroffen hat. Die internen Prozesse sind hochkomplex, statistisch gesteuert und nicht nachvollziehbar im Sinne einer klassischen Logik. Diese Intransparenz erschwert nicht nur die Fehleranalyse, sondern birgt auch Risiken in Bezug auf Kontrolle. Wer KI-Tools bzw. Sprachmodelle nutzt, muss sich also bewusst sein: Man erhält fast immer ein Ergebnis – aber selten eine Erklärung dafür.

1.2 Typische Merkmale von KI-Tools

KI-Tools zeichnen sich also durch eine Reihe gemeinsamer Eigenschaften aus, die sie von klassischen Programmen unterscheiden. Diese Merkmale erklären, warum KI-Anwendungen so flexibel und oft überraschend leistungsfähig sind.

1.2.1 Lernfähigkeit

Statt nur vorgegebene Befehle auszuführen, können KI-Modelle aus Daten lernen. Das bedeutet: Sie analysieren viele Beispiele, erkennen Muster und verbessern ihre Leistung mit jeder neuen Information. Besonders deutlich wird das bei Sprachmodellen, die durch das Training mit Millionen von Texten eine erstaunlich menschliche Sprachkompetenz entwickelt haben und auf ein extrem breit gestreutes Wissen zurückgreifen können.

1.2.2 Automatisierung komplexer Prozesse

Moderne KI-Tools sind in der Lage, mehrstufige und durchaus anspruchsvolle Aufgaben selbstständig auszuführen z.B. Texte zusammenfassen, Bilder analysieren, Übersetzungen erstellen oder Programmcode generieren. Diese Automatisierung geht weit über das hinaus, was mit klassischen *Wenn-dann-Logiken* möglich ist, und kann ganze Arbeitsabläufe effizienter gestalten. So können KI-Systeme etwa automatisch E-Mails kategorisieren, Tabellen auswerten und visualisieren, Inhalte in andere Sprachen übertragen, Gesprächsprotokolle aus Audioaufnahmen erstellen, Chatverläufe analysieren oder sogar passgenaue Präsentationen aus Rohdaten generieren, ohne dass jeder Schritt manuell vorgegeben werden muss.

1.2.3 Mustererkennung

Ein zentrales Merkmal von KI ist die Fähigkeit, Muster in großen Datenmengen zu erkennen – sei es in Texten, Bildern, Sprache oder Zahlen. Dadurch kann die KI z.B. thematische Zusammenhänge verstehen, Stimmungen erkennen und Trends analysieren. Die von ihr getroffenen Vorhersagen sind deutlich präziser als die von menschlicher Intelligenz.

1.2.4 Anpassungsfähigkeit

Viele KI-Systeme sind nicht statisch, sondern passen sich dynamisch an neue Aufgaben oder Daten an. Sie können mit unvollständigen Informationen umgehen, Alternativen vorschlagen und aus dem Kontext heraus sinnvolle Schlüsse ziehen, auch wenn eine bestimmte Aufgabe zuvor nicht explizit programmiert wurde.

1.2.5 Natürlichsprachliche Interaktion

Moderne KI-Tools ermöglichen häufig eine Kommunikation in Alltagssprache. Nutzer müssen keine speziellen Befehle oder Programmierkenntnisse haben. Die Eingabe erfolgt einfach in Form von Fragen, Anweisungen oder Beschreibungen (den sogenannten *Prompts*), was die KI für viele Menschen leicht zugänglich macht. All diese Merkmale zusammen machen KI-Tools zu extrem leistungsstarken Werkzeugen, die in vielen Bereichen völlig neue Möglichkeiten eröffnen und gleichzeitig eine neue Form des Arbeitens und Denkens erfordern.

1.3 Die wichtigsten KI-Tools im Überblick

Künstliche Intelligenz zeigt sich heute in einer Vielzahl spezialisierter Werkzeuge, von Schreibassistenten über Bildgeneratoren bis hin zu Tools für Analyse, Übersetzung und Automatisierung. Die folgenden sieben Kategorien bieten einen kompakten Überblick über zentrale Einsatzfelder und die jeweils typischen KI-Tools, die dort zum Einsatz kommen. Pro Kategorie wird ein typischer Vertreter kurz vorgestellt und die wichtigsten Alternativen dazu genannt.

1.3.1 Textgenerierung

Textgenerierung ist ein zentrales Anwendungsfeld der künstlichen Intelligenz. Gemeint ist die Fähigkeit von KI-Systemen, automatisch sinnvolle, zusammenhängende und kontextbezogene Texte zu erstellen. Die Spannweite reicht dabei von kurzen Antworten oder Textvorschlägen über ganze Blogartikel und Geschäftskorrespondenz bis hin zu Gedichten, Dialogen und längeren Aufsätzen. KI-Systeme analysieren den eingegebenen Prompt inhaltlich (wenn auch nur auf oberflächlicher Ebene) und generieren daraufhin Texte, die in Struktur und Ausdruck an den jeweiligen Kontext angepasst sind. Typische Aufgabenbereiche der Textgenerierung sind:

- Formulieren und Zusammenfassen von Texten
- Umformulieren, Kürzen oder Verlängern
- Beantworten von Fragen und Recherchehilfe
- Ideengenerierung und Brainstorming
- Verfassen von E-Mails, Artikeln, Skripten oder Beiträgen

Beispiel: ChatGPT (OpenAI)

ChatGPT ist das wohl bekannteste KI-Sprachmodell seit seiner Vorstellung im Jahr 2022 und wird vom US-Unternehmen OpenAI in San Francisco entwickelt. Die aktuelle Version ChatGPT 5 wurde am 7. August 2025 offiziell veröffentlicht und bietet einige fortschrittliche Fähigkeiten, darunter Text- und Bildverarbeitung,

verbessertes Reasoning (logisches Schlussfolgern) und ein erheblich erweitertes Kontextfenster (bis zu 1 Million Tokens). ChatGPT 5 ist konversationsfähig, kontextbewusst und flexibel einsetzbar, vom privaten Alltagshelfer bis zur professionellen Textmaschine. Seine Stärke liegt in der Benutzerfreundlichkeit und der Fähigkeit, unterschiedlichste Textsorten in natürlicher Sprache zu erzeugen. Schwächen zeigen sich vereinzelt bei Faktenungenauigkeit (Halluzinationen) und Quellenlosigkeit bei sehr aktuellen oder fachlich tiefgreifenden Themen. Die Nutzung ist in der Basisversion kostenlos unter `chat.openai.com`.

Alternativen zu ChatGPT

Neben ChatGPT als führendes Sprachmodell gibt es eine wachsende Zahl leistungsstarker Alternativen:

- **Gemini (Google)**

Nahtlose Integration in Google-Dienste, echte Multimodalität (Text, Bild, Code, Tabellen, Audio)

- **Claude (Anthropic)**

Besonders stark im Umgang mit sehr langen Kontexten (mehrere Hundert Seiten), sicherheitsorientiert, präzise in Sprache und Tonfall

- **LLaMA (Meta)**

Hochentwickeltes Open-Source-Modell von Meta, mit wachsendem Einfluss in Forschung und Industrie

- **DeepSeek**

In China entwickeltes Sprachmodell, das sich besonders gut für das Codieren und wissenschaftliches Schreiben eignet

- **Perplexity LLM**

Kombination aus Sprachmodell und Suchmaschine mit Fokus auf aktuelle Informationen und transparente Quellen

1.3.2 Bildgenerierung

KI-Bildgeneratoren erstellen aus kurzen Texteingaben eigenständig Bilder. Die KI *versteht* den Inhalt des Prompts und setzt ihn grafisch um, z.B. fotorealistisch, im Illustrationsstil oder als Gemälde. Typische Anwendungsfelder für diese Tools sind:

Illustrationen und Grafiken für Social Media, Präsentationen, Werbung

Produktideen und Designs

Konzeptzeichnungen und Kreativ-Experimente

Beispiel: DALL-E (OpenAI)

DALL·E 3 ist die aktuelle Bild-KI von OpenAI und direkt in ChatGPT (ab GPT-4o und ab dem Plus-Abo) integriert. Das Tool erzeugt aus einfachen Spracheingaben hochwertige Bilder und erlaubt sogar einfache Bildbearbeitung durch sogenannte *Inpainting-Prompts*. Besonders einsteigerfreundlich ist die Nutzung im Browser, ohne separate Plattform oder zusätzliche Anmeldung. Die Bildgenerierung erfolgt im Gesprächsverlauf und lässt sich leicht mit weiteren Textaufgaben kombinieren. DALL·E 3 überzeugt durch seine verständliche Bedienung, solide Bildqualität und die Möglichkeit, auf einfache Weise nachzubessern. Einschränkungen zeigen sich bei der künstlerischen Vielfalt und bei der detaillierten Steuerung komplexer Bildinhalte. Der Zugriff erfolgt z.B. über `chat.openai.com`.

Alternativen zu DALL-E**■ Midjourney**

Künstlerisch anspruchsvolle, stilisierte Bilder – beliebt bei professionellen Designern, aber nur über *Discord* (Online-Kommunikationsplattform) nutzbar

■ Stable Diffusion

Open Source, flexibel anpassbar, lokal einsetzbar – aber technisch komplexer

■ Adobe Firefly

Rechtssichere Bildgenerierung für den professionellen Einsatz in Adobe-Umgebungen

1.3.3 Spracherkennung & Übersetzung

KI-gestützte Systeme können gesprochene Sprache erkennen und in geschriebenen Text umwandeln (*Transkription*), automatisch in andere Sprachen übersetzen oder umgekehrt auch Texte in künstlich erzeugte, gesprochene Sprache verwandeln (*Text-to-Speech*). Sie werden in Meetings, beim Diktat, im Kundenkontakt oder für barrierefreie Kommunikation eingesetzt. Typische Aufgabenbereiche sind:

- Übersetzung von Texten und Dokumenten
- Live-Übersetzung von Sprache
- Transkription von Audio- und Videoaufnahmen
- Text-to-Speech- und Speech-to-Text-Anwendungen

Beispiel: DeepL

DeepL ist wohl das bekannteste KI-Übersetzungstool. Es überzeugt durch eine hohe Übersetzungsqualität, einen flüssigen, natürlich klingenden Stil und eine intuitive Bedienung, die auch für Einsteiger leicht zugänglich ist. Besonders bei häufig verwendeten Sprachen wie Englisch, Deutsch, Französisch oder Spanisch liefert das Tool sehr zuverlässige Ergebnisse. Etwas eingeschränkter ist die Leis-

tung bei Fachtexten, weniger verbreiteten Sprachen oder bei freier, kreativer Textgestaltung, wo die Übersetzung mitunter an Präzision oder Nuancierung verliert. Die Nutzung ist in der Basisversion kostenlos über [deepL.com](https://www.deepL.com).

Alternativen zu DeepL

■ **Google Translate**

Bietet Übersetzungen in über 100 Sprachen mit Funktionen wie Text-, Sprach- und Kameraübersetzung sowie Offline-Modus

■ **LibreTranslate**

Open-Source und selbst auf eigenem Server hostbar, datenschutzfreundlich und ohne Tracking. Für Anwender, bei denen Datensicherheit oberste Priorität hat.

1.3.4 Video- & Audioproduktion

KI-Tools im Bereich Video und Audio können Sprache, Geräusche, Musik oder ganze Videos automatisch erzeugen, bearbeiten oder verändern. Dabei kommen Techniken wie *Text-to-Video*, *Voice-Cloning* oder automatische Schnittfunktionen zum Einsatz.

Typische Aufgabenbereiche dieser Tools sind:

- Erstellung kurzer Erklär- oder Werbevideos
- Generierung synthetischer Stimmen für Vertonung
- Bearbeitung und Transkription von Audio- und Videoinhalten
- Animation von Avataren oder Sprecherfiguren

Beispiel: Runway ML

Runway ML ist eine Plattform für KI-gestützte Videoproduktion, mit der sich per Texteingabe animierte Szenen erzeugen oder vorhandene Videos bearbeiten lassen. Sie ist leicht bedienbar und besonders für kreative Kurzformate geeignet. Einschränkungen bestehen bei der Auflösung und Detailkontrolle. Zugriff über runwayml.com.

Alternativen zu Runway ML

■ **Veo (Google DeepMind)**

Generiert Videos aus Textprompts – inklusive synchronem Audio mit Hintergrundmusik, Soundeffekten und sogar Dialogen

■ **Synthesia**

Spezialisiert auf KI-Avatare, die realistisch und lippensynchron sprechen. Ideal für Unternehmensvideos, Schulungen oder Präsentationen.

1.3.5 Analyse & Automatisierung

KI-gestützte Systeme können Daten auswerten, Inhalte strukturieren und Abläufe automatisieren. Sie erkennen Muster, verdichten Informationen, erzeugen To-do-Listen oder Empfehlungen und übernehmen so Aufgaben, die früher mühsam manuell erledigt wurden. Besonders im digitalen Arbeitsalltag – von Recherche bis Projektorganisation – steigern sie Effizienz und Übersicht. Typische Aufgabenbereiche sind:

- Datenanalyse und Visualisierung
- Auswertung von Tabellen, Texten und Formularen
- Automatisierung wiederkehrender Aufgaben und Workflows
- Planung und Organisation (z.B. Zusammenfassungen, To-do-Listen)

Beispiel: Notion AI

Notion AI ist in das bekannte Notiz- und Organisationstool *Notion* integriert ist. Die KI unterstützt beim Zusammenfassen, Strukturieren und Formulieren von Inhalten, erkennt automatisch Aufgaben in Texten, erstellt Gliederungen, beantwortet Fragen zu eigenen Notizen oder generiert Tabellen und Checklisten. Notion AI überzeugt durch seine intuitive Bedienung, seine klare Einbettung in den Arbeitskontext und durch den Fokus auf Produktivität statt bloßer Textgenerierung, eignet sich für Einzelpersonen ebenso wie für Teams und ist sowohl im Bildungs- als auch im Business-Kontext nutzbar.

Alternativen zu Notion AI

■ Zapier AI

Automatisiert Abläufe zwischen Webanwendungen, erstellt per natürlicher Sprache ganze Workflows (z.B. Anhänge neuer E-Mails in Google Drive speichern)

■ Microsoft Copilot

Unterstützt in Office-Programmen wie Excel oder Outlook bei der Datenanalyse, Textverarbeitung oder Planung (stark im Unternehmensumfeld)

1.3.6 Personalisierte Empfehlungen

Empfehlungssysteme nutzen künstliche Intelligenz, um Nutzerverhalten zu analysieren und daraus passende Vorschläge zu generieren – etwa für Musik, Filme, Produkte oder Nachrichten. Die Systeme lernen aus Vorlieben, Suchverläufen, Interaktionen und Kontextdaten, um Inhalte gezielt und individuell auszuspielen. Typische Aufgabenbereiche und Anwendungen sind:

- Musik- und Videoempfehlungen
- Produktempfehlungen im Onlinehandel

- Personalisierte Newsfeeds
- Vorschläge in Streamingdiensten und sozialen Netzwerken

Beispiel: Spotify

Spotify nutzt KI, um Musik- und Podcast-Empfehlungen individuell an das Hörverhalten anzupassen. Algorithmen werten u. a. Genre-Vorlieben, Hördauer und Tageszeit aus und stellen personalisierte Playlists zusammen. Die Empfehlungen wirken oft treffsicher und entwickeln sich mit der Nutzung permanent weiter. Kritisiert wird die so entstehende Filterblase, durch die neue Genres oder unbekannte Künstler weniger sichtbar werden. Zugriff über `spotify.com`.

Alternativen zu Spotify**■ YouTube**

Nutzt KI, um Videoempfehlungen auf Basis von Sehverhalten, Klicks und Interaktionen personalisiert zusammenzustellen

■ Netflix

Setzt KI ein, um Filme und Serien individuell vorzuschlagen, inkl. personalisierter Titelbilder und Reihenfolgen

1.3.7 Multimodale KI-Tools

Multimodale KI-Systeme können verschiedene Arten von Eingaben z.B. Text, Bilder, Tabellen, Code oder Audio gleichzeitig verarbeiten und in Beziehung zueinander setzen. Sie analysieren also nicht nur eine Form von Information, sondern verknüpfen Inhalte über verschiedene Medien hinweg. Typische Aufgabenbereiche sind:

- Gleichzeitige Verarbeitung von Text, Bild, Audio und Daten
- Analyse komplexer Dokumente mit eingebetteten Elementen
- Interaktive, kontextbasierte Problemlösungen
- Erstellen multimodaler Inhalte (z.B. Präsentationen)

Beispiel: Gemini (Google)

Gemini 1.5 ist das aktuelle multimodale KI-Modell von Google DeepMind. Es verarbeitet Texte, Bilder, Diagramme, Code, Webseiten und einfache Audioformate nahtlos in einem einzigen Eingabefeld, ohne dass zusätzliche Erweiterungen erforderlich sind. Das Modell erkennt komplexe Zusammenhänge zwischen verschiedenen Informationsarten, beantwortet Fragen zu Grafiken, analysiert Tabellen und generiert daraus visuelle Darstellungen und inhaltlich passende Beschreibungen. Zu den besonderen Stärken von Gemini zählen seine hohe Aktualität, die echte Multimodalität ohne Plugins sowie die enge Integration in bestehende Google-Dienste. Der Zugang erfolgt über die Plattform `gemini.google.com`.