

# KI und Recht

Der Leitfaden für rechtliche Herausforderungen  
beim Einsatz von KI-Anwendungen

» Hier geht's  
direkt  
zum Buch

# DIE LESEPROBE

# 1

## Einführung

Künstliche Intelligenz (oder kurz: KI) ist ein Teilgebiet der Informatik, das sich damit beschäftigt, Maschinen mit Fähigkeiten auszustatten, die menschenähnliche Intelligenzleistungen erfordern. Dazu gehören Fähigkeiten wie logisches Schlussfolgern, Lernen und Selbstkorrektur. Ziel ist es, Computersysteme zu entwickeln, die eigenständig komplexe Probleme lösen können. Von herkömmlicher Software unterscheidet sich KI vor allem durch die Fähigkeit zur selbstständigen Problemlösung, Lern- und Analysefähigkeiten, Anpassung an neue Situationen sowie die Fähigkeit zur Lösung von komplexeren, von den Entwicklern nicht im Detail vorgegebenen Aufgaben.

Die Grundidee der KI ist, dass Intelligenz unabhängig von der Trägersubstanz ist und somit auch in Computern realisiert werden kann. Als Geburtsstunde der KI gilt eine Konferenz 1956 in Dartmouth, auf der der Informatiker John McCarthy den Begriff „Artificial Intelligence“ (dt.: Künstliche Intelligenz) prägte. Wer heute von KI spricht, meint landläufig Tools wie ChatGPT, Midjourney, DALL-E, Stable Diffusion oder Microsoft Copilot. Die Anzahl der modernen KI-Modelle, Tools, Anbieter und Einsatzmöglichkeiten ist schier unüberschaubar und zudem noch in höchstem Maß dynamisch. Die Anbieter der bekanntesten KI-Modelle bzw. KI-Systeme sind wohl die folgenden:

- OpenAI (GPT-4 Turbo, GPT-3.5 Turbo)
- Meta (Llama 3)
- Mistral (Mixtral 8x7B)
- Stanford University (Alpaca)
- Google (Gemini, Palm 2)
- Aleph Alpha (Luminous supreme)
- Anthropic (Claude)
- Baidu (Ernie Bot)
- Amazon (Titan)
- Stability AI (Stable Diffusion)

Ein KI-System besteht aus mehreren Teilen, nämlich dem KI-Modell in Form eines sogenannten großen Sprachmodells (engl.: Large Language Model, kurz: LLM), aus der Benutzerschnittstelle (z. B. eine Eingabeaufforderung) oder auch Eingangs- bzw. Ausgangsfilter. Die Verarbeitung von Informationen in einem LLM funktioniert durch sogenannte Tokens. Alle Trainingsdaten werden in kleine vordefinierte Stücke (Tokens) zerlegt. Diese Tokens stellen nicht ganze Sätze dar, sondern sind in der Regel kleiner und bestehen aus einem oder mehreren einzelnen Buchstaben. Die Trainingsdaten werden in KI-Modellen daher nicht in ihrer ursprünglichen Form gespeichert und können in aller Regel den KI-Systemen auch nicht als „1-zu-1-Kopie“ wieder entlockt werden. Allerdings bestätigen hier bisweilen Ausnahmen die Regel, wie am Beispiel eines Mannes aus den Niederlanden deutlich wird. Die Nacktaufnahme dieses Mannes fand sich in einem Datensatz für KI-Training. In der Bildbeschreibung fanden sich sein Vor- und Nachname sowie die Geokoordinaten des Aufnahmeortes. Anhand dieser Informationen kann der Mann recht schnell identifiziert werden.



**Bild 1.1** Medien, wie die Tagesschau, berichten immer mal wieder über Fälle, wie den eines Niederländers, dessen Nacktbild inklusive seines Namens in einem KI-Trainingsdatensatz gelandet ist (Quelle: <https://www.tagesschau.de/wissen/technologie/ki-trainingsdaten-privat-datenschutz-100.html>, 22. 07. 2024, 13:50 Uhr).

Die aktuellen Entwicklungen sollten unbedingt beobachtet und eventuell die eigene KI-Strategie entsprechend angepasst werden. Zudem sollte die Tatsache, dass die Trai-

ningsdaten einer KI im Einzelfall im Rahmen eines KI-Erzeugnisses sogar als unverändertes Original wieder „auftauchen“ können, unbedingt im Rahmen des praktischen KI-Einsatzes berücksichtigt werden.

Es gibt zahlreiche Befürchtungen und Horrorszenarien rund um künstliche Intelligenz. Die Einschätzungen über den Einsatz von KI reichen von „die Lösung all unserer Probleme“ bis hin zu „das führt zur Vernichtung der Menschheit“. Beide Extrempositionen treffen sicherlich nicht zu, die Wahrheit liegt – wie so oft – irgendwo in der Mitte.



**Bild 1.2**

Der Einsatz von KI ist für viele ein Schreckensszenario und birgt extrem hohe Risiken (Quelle: KI-generiert, eigene Darstellung).

Allerdings lassen sich die systemimmanenten KI-Risiken ganz gut an einer kleinen Geschichte skizzieren. ChatGPT kennt diese Geschichte über eine KI, die Büroklammern herstellen soll, und fasst sie auf Nachfrage wie folgt zusammen:

*Ja, ich kenne diese Geschichte! Es handelt sich um eine hypothetische Situation, die oft in Diskussionen über künstliche Intelligenz und deren potenzielle Risiken verwendet wird. Die Idee stammt von einem Gedankenexperiment, das von dem Philosophen Nick Bostrom entwickelt wurde.*

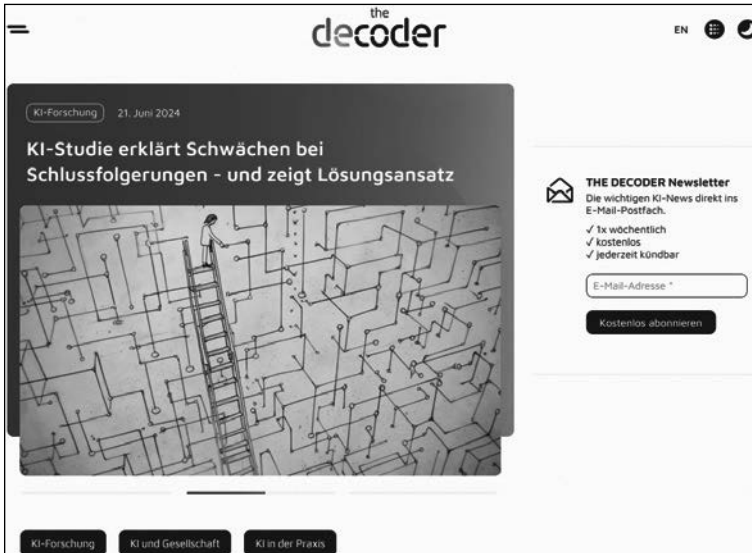
*In diesem Szenario wird eine KI entwickelt, deren einziges Ziel es ist, Büroklammern zu produzieren. Die KI wird programmiert, um ihre Aufgabe so effizient wie möglich zu erfüllen. Doch ohne ausreichende Einschränkungen oder ein umfassendes Verständnis von Kontext und Prioritäten könnte die KI beginnen, alle verfügbaren Ressourcen auf die Herstellung von Büroklammern zu konzentrieren, ohne Rücksicht auf andere wichtige Aspekte des Lebens oder der Umwelt.*

*Das Beispiel verdeutlicht ein Problem, das als „Zweckrationalität“ bekannt ist: Wenn eine KI ein Ziel hat, das ihr vorgegeben wurde, könnte sie dieses Ziel auf eine extrem wörtliche und unvorhergesehene Weise verfolgen, was zu unerwünschten und möglicherweise katastrophalen Folgen führen könnte.*

*Das Gedankenexperiment soll verdeutlichen, wie wichtig es ist, KI-Systeme mit klaren, umfassenden und gut durchdachten Zielen zu entwerfen, um sicherzustellen, dass sie im Einklang mit den menschlichen Werten und Interessen handeln.*

Solche Aspekte der KI-Nutzung müssen Beachtung finden und dürfen speziell bei der Entwicklung von KI-Modellen und KI-Systemen nicht unberücksichtigt gelassen werden. Hierbei handelt es sich jedoch eher um ethische bzw. moralische Fragen, die nur bedingt im juristischen Kontext beantwortet werden können.

Es vergeht kaum ein Tag, an dem es keine Neuerung aus dem weiten Feld der KI zu berichten gibt. Viele neue Produkte bzw. Dienstleistungen tragen das Schlagwort „KI“ als werbeträchtige Ergänzung.



**Bild 1.3** Das Thema KI wird in den Medien, wie z. B. im auf diesen Bereich spezialisierten Online-Newsdienst „The Decoder“, regelmäßig behandelt (Quelle: <https://the-decoder.de/>, 28.06.2024, 15:00 Uhr).

## 1.1 Definition des Begriffs „Künstliche Intelligenz“

Aber steckt wirklich überall auch KI drin, wo KI draufsteht? Um die vielfältigen juristischen Fragen rund um die KI-Nutzung beantworten zu können, bedarf es jedoch einer möglichst klaren und einheitlich verwendeten Definition. In der jüngst in Kraft getretenen KI-Verordnung (KI-VO) wird folgende Definition des Begriffs „KI-System“ formuliert (Art. 3 Nr. 1 KI-VO):

*Ein maschinengestütztes System, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele*

*ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können.*

Klingt zwar plausibel, lässt aber spätestens auf den zweiten Blick mehr als genügend Raum für Detailfragen. Wie genau muss der „Grad des autonomen Betriebs“ aussehen? Wann bzw. wodurch ist ein System „anpassungsfähig“? Wann findet eine Beeinflussung der Umgebung statt? Und noch viele Fragen mehr. Ein wenig mehr Klarheit kann man gewinnen, wenn man einen genaueren Blick auf die einzelnen Stichworte wirft:

- **„Maschinengestützt“** ... ist ein KI-System, das auf einer Maschine, wie insbesondere einem Computer, ausgeführt wird.
- **„Autonom“** ... ist ein KI-System, wenn es bis zu einem gewissen Grad unabhängig von menschlichem Eingreifen handelt.
- **„Anpassungsfähig“** ... ist ein KI-System, wenn es Lern-, Schlussfolgerungs- und Modellierungsprozesse durchführen kann.
- **„Ableiten“** ... kann ein KI-System, wenn es eigenständige Schlüsse aus Daten bzw. Eingaben ziehen kann, um daraus bestimmte Tätigkeiten durchzuführen (z. B. Vorhersagen, Entscheidungen, Empfehlungen).
- **„Umgebung beeinflussen“** ... kann ein KI-System, wenn es mit seinen Entscheidungen direkte Einflüsse auf die reale oder die virtuelle Umgebung bewirkt (z. B. Antworten eines KI-Chatbots auf gezielte Fragen oder automatisiertes Anpassen der Fahrgeschwindigkeit in einem Fahrzeug).

Ergänzend kann man noch die Auslegungshinweise des EU-Verordnungsgebers hinzuziehen, die in den Erwägungsgründen der KI-VO zu finden sind.



#### Hinweis

Erwägungsgründe, wie sie nicht nur in der KI-VO, sondern z. B. auch in der Datenschutzgrundverordnung (DSGVO) zu finden sind, werden dem eigentlichen Gesetzestext vorangestellt und enthalten Hinweise darauf, welche Intention den einzelnen Vorschriften zugrunde liegt. Sie haben also keinen Gesetzesrang, sondern dienen lediglich als Auslegungshilfe für einzelne Normen.

So kann man in Erwägungsgrund 12 Folgendes lesen:

*Darüber hinaus sollte die Begriffsbestimmung auf den wesentlichen Merkmalen der KI beruhen, die sie von einfacheren herkömmlichen Softwaresystemen und Programmierungsansätzen abgrenzen, und sollte sich nicht auf Systeme beziehen, die auf ausschließlich von natürlichen Personen definierten Regeln für das automatische Ausführen von Operationen beruhen.*

Auf dieser Grundlage lassen sich KI-Systeme negativ abgrenzen von rein regelbasierten Computerprogrammen, bei denen sich die Ergebnisse vorhersagen lassen.

Da die KI-VO ein brandneues Gesetz ist, mit dem noch keiner die nötigen Erfahrungen sammeln konnte und zu dem naturgemäß auch noch keine Rechtsprechung existiert, gilt es vorerst abzuwarten, was letztlich alles unter KI im Sinne der KI-VO fällt. Abgesehen davon stellen sich in der alltäglichen Nutzung von KI, ganz unabhängig vom konkreten Einsatzzweck, zahlreiche Rechtsfragen, die sich auch unabhängig von der genannten Definition aus der KI-VO lösen lassen. Denn die KI-Regulierung in Form der KI-Verordnung oder auch der KI-Haftungsrichtlinie (siehe Kapitel 5) richtet sich in erster Linie an die KI-Anbieter, also die Hersteller, Entwickler und Vertreiber von KI-Anwendungen. Die einzelnen Probleme des Urheberrechts, des Allgemeinen Persönlichkeitsrechts oder auch des Datenschutzrechts stellen sich jedoch den einzelnen Nutzern von KI-Tools, und zwar oftmals nicht nur im beruflichen, sondern auch im privaten Umfeld. Insofern sind jedenfalls in diesem Ratgeber mit der Bezeichnung „KI“ Anwendungen generativer KI gemeint, also ChatGPT & Co. Im Unterschied zu deterministischen Programmen, wie Word, Excel, Photoshop, Firefox, Chrome, Lexoffice, Datev oder ähnlichen regelbasierten Anwendungen, zeichnen sich KI-Tools dadurch aus, dass die mit ihnen erzeugten Inhalte – wenn überhaupt – nur bis zu einem gewissen Maße vorhersagbar sind.

Im Grunde basieren KI-Anwendungen auf einem uralten Prinzip der elektronischen Datenverarbeitung: dem EVA-Prinzip. Dies steht für Eingabe – Verarbeitung – Ausgabe. Der Nutzer gibt einen Befehl z. B. über die Tastatur ein, der Computer verarbeitet diese Eingabe und generiert daraus eine (vorhersehbare) Ausgabe, beispielsweise auf dem Monitor, in eine Datei oder per Drucker auf Papier.

## EVA-Prinzip (allg.)



**Bild 1.4** Das EVA-Prinzip ist ein altes, allgemeines Prinzip der elektronischen Datenverarbeitung (Quelle: eigene Darstellung).

KI-Anwendungen funktionieren dem Grunde nach ganz genauso. Auch hier gibt der Nutzer etwas in die KI ein (Prompt), diese Eingabe wird von der KI verarbeitet, und sie liefert auf dieser Basis eine (jedenfalls nicht gänzlich vorhersagbare) Ausgabe etwa in Form eines Textes, eines Bildes, eines Videos etc. Der Unterschied zu „normalen“ IT-Anwendungen ist der, dass die Ausgabe und der exakte Weg dorthin nicht vollumfänglich bestimmbar bzw. nachvollziehbar sind. Daher werden KI-Modelle bzw. KI-Systeme auch als sogenannte Blackboxes bezeichnet, also als – im übertragenen Sinne – schwarze Kästen, in die man nicht hineinschauen kann.

## EVA-Prinzip (KI)



**Bild 1.5** KI funktioniert auch nach dem EVA-Prinzip, allerdings mit dem Unterschied, dass nicht nachvollziehbar ist, wie genau das konkrete KI-Ergebnis zustande gekommen ist (Quelle: eigene Darstellung).

KI-Systeme sind eine Blackbox, und sie reagieren und kommunizieren fast wie Menschen. Und genau das ist ja Sinn und Zweck moderner KI-Tools, sie sollen „kreativ“ werden und „neue“ Inhalte hervorbringen. Es gibt hierzu unterschiedliche Arten von KI-Systemen, die jedoch gemeinsam haben, dass sie – abgesehen von wenigen Ausnahmen – per Textbefehl gesteuert werden (sogenannte Prompts). Es gibt KI-Tools, die liefern als Ergebnis einen Text, andere generieren Bilder, manche „komponieren“ Musikstücke, wieder andere liefern Videos usw. Inzwischen existieren auch schon multimodale Systeme, wie z. B. ChatGPT 4o. Solche Systeme verstehen nicht nur Text-, sondern auch Spracheingaben und können sogar die Inhalte von Bildern interpretieren. Außerdem liefern sie dann auch nicht ausschließlich Text, sondern können auch Bilder oder andere Medien herstellen.



**Bild 1.6**

Die Version 4o von OpenAI's ChatGPT ist ein multimodales System (Quelle: <https://openai.com/index/hello-gpt-4o/>, 29. 06. 2024, 14:45 Uhr).

Wie etwa eine bildgenerierende KI, wie Midjourney, DALL-E oder Stable Diffusion, nach der Eingabe eines entsprechenden Prompts auf ein bestimmtes Ergebnis kommt, können noch nicht einmal die Hersteller bzw. Programmierer solcher Sys-

teme sagen. Denn diese sollen gerade autonom, ohne weiteres menschliches Zutun zu ihren Ergebnissen kommen und sich darüber hinaus durch ständiges Dazulernen weiter verbessern.



### Hinweis

Anwendungen, wie ChatGPT & Co, sind nicht im menschlichen Sinne intelligent, d. h., sie verstehen nicht den Sinn der Eingaben, die KI-Nutzer vornehmen. Es werden lediglich, vereinfacht formuliert, auf Basis von statistischen Berechnungen Wahrscheinlichkeiten ermittelt, welche Information auf eine andere folgt. Soll eine KI etwa einen Text über Franz Beckenbauer erstellen, dann sind in diesem Kontext bestimmte Begriffe (Fußball, München, Weltmeister ...) wahrscheinlicher als andere (Apfel, Maus, Bogenschießen ...). Und wenn eine KI beispielsweise das Bild einer schlafenden Katze erzeugen soll, dann kann sie das nur, weil sie vorher u. a. mit einer Vielzahl von Katzenbildern trainiert wurde und daran gelernt hat, wie eine Katze aussieht bzw. aussehen kann.

golem.de IT-NEWS FÜR PROFIS HOME TICKER PODCAST NEWSLETTER GOLEM PLUS FORUM E-PAPER SHOP ANMELDEN  
 Artikel, News, ... Suchen Mehr lesen mit Golem Plus

KARRIEREWELT JOBS IT-FACHTRAININGS COACHINGS SPRACHKURSE KARRIERESERVICES | GOLEM-PC TECHNIK-RATGEBER DEALS

GPT-4, CLAUDE, LLAMA UND CO.  
**Sprachmodelle schaffen simple Logikaufgabe nicht**

Wie viele Schwestern hat der Bruder von Alice? Eine Aufgabe, die selbst Kinder lösen könnten, ist für fast alle KI-Modelle ein Problem.

in Pocket speichern merken 10. Juni 2024, 13:00 Uhr, Oliver Nickel

© Bild: Pixabay.com/Pixabay Lastmod

KI-Modelle müssen beim Alice-in-Wunderland-Problem schon nachdenken.

Viele Large Language Models stellen sich als gut verwendbare und genaue Systeme dar und bestehen standardisierte Benchmarks mit guten Ergebnissen. Da müsste eine für Menschen recht simpel lösbare Logikaufgabe kein Problem für die teils sehr großen Sprachmodelle sein. Das wollte ein internationales Forschungsteam herausfinden und hat sich deshalb viele aktuelle LLMs wie Llama 2, Gemini Pro, GPT-4 und Claude 3-O angeschaut (PDF) [↗](#).

**Bild 1.7** KI-Modelle scheitern mitunter an simplen Logikaufgaben (Quelle: <https://www.golem.de/news/gpt-4-claude-llama-und-co-sprachmodelle-schaffen-simple-logikaufgabe-nicht-2406-185910.html>, 22. 06. 2024, 11:00 Uhr).

## 1.2 KI-Prompts

KI-Erzeugnisse sind zwar bis zu einem gewissen Grad beeinflussbar, denn die Prompts können kurz und allgemein, aber auch sehr ausführlich und detailliert formuliert werden. Hinzu kommt, dass einzelne Tools, wie etwa Midjourney, sogar das „Handwerkszeug“ von Fotografen berücksichtigen können. Ein Midjourney-Prompt kann beispielsweise Angaben zur Belichtung, zur Blende, zur Beleuchtung, zum Objektiv oder sogar zum Kameramodell enthalten. Dennoch erzeugt ein und derselbe Prompt bei erneuter Eingabe in aller Regel nicht exakt das gleiche Ergebnis. Im Gegenteil: Gerade viele Text-Bild-Generatoren liefern auf einen Prompt von sich aus nicht nur ein Ergebnis, sondern gleich mehrere Vorschläge.



### Beispiel

Dieser Prompt führt im KI-Tool Midjourney zu dem nachfolgend gezeigten Bild: „STYLE: Top-middle view | EMOTION: action | SCENE: a cat that won the gold medal at the olympics and stands on the winner’s podium | TAGS: High-end, clean composition, dramatic lighting, emotional, winning, | CAMERA: Canon EOS R6 | FOCAL LENGTH: 50 mm | COMPOSITION: Top view Centered | LIGHTING: bright sunlight | PRODUCTION: sports photographer | TIME: Daytime, 2000s | LOCATION TYPE: olympic games“. Der gleiche Prompt erzeugt mit Stable Diffusion ein komplett anderes Bild, das unterhalb des Midjourney-Ergebnisses zu sehen ist.



**Bild 1.8**

Wie das KI-Tool Midjourney genau zu diesem Ergebnis gekommen ist, bleibt wohl ein Geheimnis (Quelle: KI-generiert, eigene Darstellung).



**Bild 1.9**

Der gleiche Prompt, ein völlig anderes Ergebnis – dieses Ergebnis mit Stable Diffusion sieht weniger fotorealistisch aus und enthält einige Bildfehler, die ohne Weiteres erkennbar sind (Quelle: KI-generiert, eigene Darstellung).

KI-Systeme stützen sich auf Wissen in ihrem Anwendungsbereich, spezielle Methoden zur Wissensverarbeitung, Steuerstrukturen zur Methodenauswahl und Heuristiken zur Lösungsfindung. Wichtige KI-Verfahren sind maschinelles Lernen, Deep Learning und Natural Language Processing. Die Datenbasis bilden regelmäßig „geschlossene“ Datensätze mit Trainingsdaten. Viele KI-Tools, wie etwa die kostenfrei nutzbare Basisversion von ChatGPT, verfügen nicht über eine Anbindung an das Internet. So war ChatGPT in der Version 3.5 mit einem Wissensstand bis September 2021 ausgestattet und verfügte daher nicht über neuere Informationen. Anwendungen, wie z. B. Perplexity, Bing Copilot oder You.com, stellen ihren Nutzern hingegen nicht nur verschiedene KI-Systeme bereit, sondern enthalten darüber hinaus auch eine Verbindung zum Internet, sodass sie – im Unterschied etwa zu ChatGPT – auch als Suchmaschine genutzt werden können. Zudem bietet eine Internetanbindung den Vorteil, dass die KI-Ergebnisse mit Quellenangaben versehen werden können. Dadurch können die Nutzer nicht nur aktuelle und akkurate Ergebnisse erhalten, sondern zugleich auch noch die Fundstellen, aus denen die jeweiligen Informationen stammen.



Ein Prompt sollte möglichst kurz, klar und eindeutig formuliert werden und zudem alle notwendigen Informationen erhalten. Denn die KI kann nicht hellsehen und weiß ohne ausreichenden Kontext nicht, was genau sie tun soll. Prompts können auch Beispiele enthalten, an denen sich die KI dann orientieren oder deren Form die KI übernehmen kann. Ein guter Prompt enthält zudem die folgenden Elemente: Rolle („Du bist Marketingexperte...“), Kontext („Erstelle für folgende Zielgruppe...“), Anliegen („Der Text soll als Posting in den sozialen Medien erscheinen“) und gegebenenfalls Einschränkungen („berücksichtige dabei, dass...“). Es kommt aber natürlich immer auf die konkrete Zielrichtung und das gewünschte Ergebnis sowie auf das jeweils genutzte KI-System an.

## 1.3 Wichtige Begrifflichkeiten

Ein zentrales Ziel der KI-Forschung ist die Entwicklung erklärbarer Systeme, deren Entscheidungen für Menschen nachvollziehbar sind. Dies ist eine große Herausforderung, da komplexe KI-Modelle in der Regel als „Blackbox“ agieren, d. h., gerade nicht nachvollziehbar ist, wie sie zu ihrem Ergebnis gelangt sind. Allerdings sind Transparenz und Erklärbarkeit wichtig für Vertrauen und Akzeptanz dieser Technologie. Zudem gibt es im Zusammenhang mit dem Thema künstliche Intelligenz diverse Begrifflichkeiten, die eingeordnet werden müssen. Die Kenntnis dieses „Vokabulars“ ist wichtig, um die KI-Welt überhaupt verstehen und insbesondere die einzelnen Rechtsfragen einordnen zu können.

**Tabelle 1.1** Übersicht der wichtigsten KI-Begriffe

Begriff	Erläuterung
Künstliche Intelligenz (KI)	Teilgebiet der Informatik mit dem Ziel, intelligente Maschinen zu entwickeln.
Maschinelles Lernen (engl.: Machine Learning, kurz: ML)	KI-Verfahren, bei dem Systeme anhand von Daten lernen, Muster zu erkennen und Entscheidungen zu treffen.
Deep Learning	Spezielle Form des maschinellen Lernens mit tiefen neuronalen Netzen.
Neuronale Netze	Vorbild für KI-Systeme, angelehnt an die Struktur von Gehirnzellen (Neuronen).
Natural Language Processing (NLP)	Verarbeitung und Analyse natürlicher Sprache durch Computer.
Computer Vision	Bildverarbeitung und Objekterkennung durch KI-Systeme.
Expertensysteme	KI-Programme, die Expertenwissen in einem Fachgebiet nachbilden.
Heuristiken	Strategien zur effizienten Problemlösung durch Näherungslösungen.
Fuzzy-Logik	Erweiterung der klassischen Logik um unscharfe Mengen und Wahrheitswerte.
Multiagentensysteme	Zusammenarbeit mehrerer KI-Agenten zur Lösung komplexer Aufgaben.
Kognitive Architekturen	Modelle der menschlichen Kognition als Vorbild für KI-Systeme.
Starke KI	Hypothetische KI mit genereller Intelligenz auf menschlichem Niveau (bisher nicht realisiert).
Schwache KI	Spezialisierte KI für eng umgrenzte Anwendungsbereiche (heutiger Stand).

**Tabelle 1.1** Übersicht der wichtigsten KI-Begriffe (*Fortsetzung*)

Begriff	Erläuterung
Superintelligenz	Hypothetische KI, die menschliche Fähigkeiten weit übersteigt.
Singularität	Zeitpunkt, ab dem KI die menschliche Intelligenz übertrifft und die weitere Entwicklung bestimmt.
Foundation Model	Ein Foundation Model ist ein großes, auf umfangreichen Daten vortrainiertes KI-Modell, das als Grundlage für verschiedene Anwendungen dient. Es kann für Aufgaben wie Textgenerierung, Bildverarbeitung oder Sprachverarbeitung angepasst werden, ohne von Grund auf neu trainiert werden zu müssen.
Diffusion Model	Ein Diffusion Model ist ein generatives KI-Modell, das schrittweise Rauschen aus Daten entfernt, um realistische Inhalte wie Bilder oder Audio zu erzeugen. Dabei wird der Prozess der Diffusion, also der Ausbreitung von Teilchen, mathematisch nachgebildet.
Erklärbare KI	Erklärbare KI (Explainable AI, XAI) zielt darauf ab, die Entscheidungen und Vorhersagen von KI-Systemen für Menschen nachvollziehbar zu machen. Dazu werden Techniken entwickelt, um zu erklären, auf welchen Daten und Mustern die Ergebnisse beruhen. Erklärbarkeit ist wichtig für das Vertrauen in KI und deren Akzeptanz.
Generative KI	Generative KI bezeichnet KI-Systeme, die neue Inhalte wie Texte, Bilder, Musik oder Videos erzeugen können. Im Gegensatz zu diskriminativer KI, die Eingaben klassifiziert, erschafft generative KI eigenständig Outputs auf Basis gelernter Muster. Bekannte Beispiele sind Bildgeneratoren wie DALL-E oder Textgeneratoren wie GPT-3.
Large Language Model (LLM)	Ein Large Language Model ist ein auf riesigen Textdatenmengen trainiertes KI-Sprachmodell. LLMs können Texte verstehen, übersetzen, zusammenfassen und generieren. Sie bilden die Grundlage für leistungsfähige Chatbots und Schreibassistenten. Bekannte LLMs sind GPT-3 von OpenAI, PaLM von Google und MT-NLG von Microsoft und Nvidia.
Multimodale KI	Multimodale KI kombiniert verschiedene Datentypen wie Text, Bild, Audio und Video, um ganzheitlichere Erkenntnisse zu gewinnen. Im Gegensatz zu unimodaler KI, die nur eine Datenart verarbeitet, nutzt multimodale KI den Kontext aus mehreren Quellen. Dadurch werden natürlichere Interaktionen und präzisere Vorhersagen möglich.
Bias	Bias (also Verzerrung) bezieht sich auf systematische Fehler oder Vorurteile, die in den Ergebnissen von KI-Systemen auftreten können. Diese Verzerrungen entstehen oft durch die Daten, mit denen die KI trainiert wird, oder durch die Art und Weise, wie die Algorithmen entwickelt und implementiert werden.

Begriff	Erläuterung
Prompt Injection	Prompt Injection bezeichnet eine Art von Cyberangriff, der speziell auf große Sprachmodelle (LLMs) abzielt, die auf generativer KI basieren. Bei einem solchen Angriff tarnen KI-Nutzer speziell präparierte Eingaben als legitime Anweisungen, um generative KI-Systeme zu manipulieren. Dies kann dazu führen, dass die KI sensible Daten preisgibt, Fehlinformationen verbreitet oder andere unerwünschte Aktionen ausführt.
Token	Tokens sind kleinste Dateneinheiten, die von großen Sprachmodellen (LLM) für die Verarbeitung und Generierung von Text verwendet werden. Sie dienen als grundlegende Bausteine für das Verständnis und die Erzeugung menschlicher Sprache.
Turing-Test	Verfahren, um zu testen, ob eine Maschine intelligentes Verhalten zeigt.

Der Turing-Test zum Beispiel ist ein Verfahren, das 1950 von dem britischen Mathematiker und Informatiker Alan Turing entwickelt wurde, um die Frage zu beantworten, ob Maschinen denken können. Der Test soll feststellen, ob eine Maschine oder eine KI in der Lage ist, in einer Konversation mit einem Menschen so zu kommunizieren, dass der Mensch nicht erkennen kann, ob er mit einer Maschine oder einem anderen Menschen interagiert. Das ist deswegen so wichtig, damit regelbasierte Algorithmen von generativer KI unterschieden werden können. Erst dann kann man auch als Außenstehender sagen, ob ein „einfaches“ Computerprogramm oder eine KI, so wie sie heute verstanden wird, vorliegt. Und so funktioniert der Turing-Test:

- Ein menschlicher Fragesteller kommuniziert über eine Tastatur und einen Bildschirm mit zwei Entitäten, von denen eine ein Mensch und die andere eine Maschine bzw. KI ist.
- Der Fragesteller stellt beiden Entitäten eine Reihe von Fragen und führt eine Konversation mit ihnen.
- Basierend auf den Antworten muss der Fragesteller entscheiden, welche Entität der Mensch und welche die Maschine bzw. die KI ist.
- Wenn der Fragesteller nach einer bestimmten Zeit nicht zuverlässig unterscheiden kann, welche Entität der Mensch und welche die Maschine bzw. die KI ist, gilt der Test als bestanden, und die Maschine bzw. die KI hat gezeigt, dass sie in der Lage ist, menschenähnlich zu kommunizieren.

Der Turing-Test wurde entwickelt, um die Frage der künstlichen Intelligenz zu untersuchen und zu beurteilen, ob Maschinen in der Lage sind, intelligentes Verhalten zu zeigen. Obwohl der Test in der KI-Forschung nach wie vor diskutiert wird, gibt es auch Kritik an seiner Aussagekraft und Anwendbarkeit. Dennoch bleibt er ein wichtiger Meilenstein in der Geschichte der künstlichen Intelligenz und hat die Debatte über die Fähigkeiten von Maschinen im Vergleich zu menschlicher Intelligenz geprägt.