IMKIS - Institut für Medien, Kommunikation, Information und Sprache

GLOSSAR WICHTIGER KI-BEGRIFFE

VON AGENTEN BIS ZITIERRECHT

IMKIS
Dennemarkstraße 24
47647 Kerken

(02833) 576117-0
post@imkis.de
www.imkis.de

Erstellt mithilfe diverser Dialog-KI, geplant und redigiert von IMKIS

GLOSSAR WICHTIGER KI-BEGRIFFE

Agenten

Agenten sind in der Künstlichen Intelligenz Systeme, die selbstständig Aufgaben übernehmen und Entscheidungen treffen können, häufig auch über mehrere Schritte oder Anwendungen hinweg. Sie kombinieren oft heuristische Methoden mit klassischen Algorithmen und sind damit in der Lage, komplexe Arbeitsabläufe zu automatisieren. Ein KI-Agent kann beispielsweise eine Aufgabe planen, Informationen aus verschiedenen Quellen einholen, diese verarbeiten und ein Ergebnis präsentieren – ohne dass der Mensch jeden einzelnen Schritt aktiv steuern muss. Agenten werden zunehmend in Bereichen eingesetzt, in denen Routineaufgaben automatisiert und effizienter gestaltet werden sollen. Dabei wächst die Verantwortung für eine angemessene Kontrolle, da fehlerhafte Entscheidungen oder ungewünschte Ausgaben ebenfalls automatisiert ablaufen können.

Alignment

Alignment (Ausrichtung) bezieht sich im Zusammenhang mit Künstlicher Intelligenz auf die Angleichung der Ziele und Handlungen von KI-Systemen an die Ziele, Werte und Bedürfnisse der menschlichen Gesellschaft. Auf diese Weise soll sichergestellt werden, dass KI-Systeme ethische, rechtliche und soziale Standards einhalten. Herausfordernd ist dabei unter anderem, sich auf Standards zu einigen und diese in ihrer Komplexität in KI-Systeme zu integrieren. Ein mangelhaftes Alignment (Misalignment) kann zu unerwünschtem Verhalten von KI-Systemen führen, was potenziell negative Auswirkungen auf die Gesellschaft hat.

Bias

Bias (Verzerrung) bezieht sich auf eine systematische Abweichung in den Ergebnissen oder Entscheidungen, die von KI-Systemen getroffen werden. Diese Abweichungen können z. B. durch die Trainingsdaten entstehen. Denn wenn diese Daten bereits Vorurteile oder Ungleichheiten aufweisen, übernehmen KI-Systeme diese und reproduzieren sie. Aber auch der Algorithmus selber kann schon verzerrt sein, z. B. durch die angesetzten Variablen und ihre Gewichtigung. Verzerrungen im Datenset und System können zu Diskriminierung und Verfestigung von Vorurteilen sorgen.

Chatbots

Ein Chatbot ist ein Computerprogramm, das mithilfe von Künstlicher Intelligenz oder vordefinierten Regeln automatisch auf Nutzeranfragen antwortet. Chatbots können in verschiedenen Anwendungen eingesetzt

werden, wie zum Beispiel im Kundenservice, Marketing oder als persönlicher Assistent. Chatbots können durch Machine Learning und Natural Language Processing (NLP) immer besser auf die Bedürfnisse der Nutzer eingehen und somit eine menschenähnliche Interaktion ermöglichen.

Canvas-Modus

Der Canvas-Modus ist ein Arbeitsmodus in modernen KI-Tools, bei dem man längere Inhalte in einem separaten, seitlich eingeblendeten Editor bearbeiten kann. Anders als bei kurzen Chat-Antworten erlaubt der Canvas-Modus strukturierte, anhaltende Arbeit an Dokumenten, mit gezielten Änderungen, Kommentaren oder Versionierung. Der Begriff „Canvas“ (Leinwand) betont die Offenheit und Flexibilität des Formats, das besonders für komplexe Projekte wie Drehbücher, Forschungsnotizen oder Software-Entwicklung genutzt wird.

Custom instructions

Unter dem Begriff „Custom instructions“ (benutzerdefinierte Anweisungen) wurden 2023 zusätzliche Einstellungsmöglichkeiten bei ChatGPT eingefügt, mit denen man die Ausgabe des KI-Systems präziser steuern kann. Diese Einstellungen sind aktuell unter der Bezeichnung „individuelle Hinweise“ zu finden und werden zur Personalisierung eingesetzt. So kann man hier angeben, mit welchem Namen der Chatbot einen ansprechen soll und welchen Beruf man ausübt. Außerdem kann man festhalten, wie sich das System verhalten soll in Bezug auf Rolle und Sprache und welche weiteren Hinweise die KI beachten soll. Diese Angaben gelten Chat-übergreifend und können deshalb dabei helfen, die Arbeit mit ChatGPT möglichst effizient zu gestalten.

Datenschutz

Datenschutz bezieht sich auf die Maßnahmen zum Schutz personenbezogener Daten vor unbefugtem Zugriff, Verlust oder Missbrauch. Im Bereich der Künstlichen Intelligenz ist Datenschutz von entscheidender Bedeutung, da KI-Systeme oft auf große Mengen sensibler Daten zugreifen. Unternehmen und Entwickler müssen sicherstellen, dass KI-Anwendungen die Datenschutzbestimmungen einhalten, um die Privatsphäre der Nutzer zu schützen. Dies beinhaltet die Anonymisierung von Daten, die Einhaltung von Datenschutzgesetzen wie der DSGVO und die Implementierung von Sicherheitsmaßnahmen, um Daten vor unerlaubtem Zugriff zu schützen. Datenschutzrichtlinien und Transparenz bezüglich der Datennutzung sind ebenfalls wichtige Aspekte, um das Vertrauen der Nutzer in KI-Systeme zu gewährleisten.

Debugging

Debugging bezeichnet den Prozess des Auffindens, Analysierens und Behebens von Fehlern (Bugs). Im ursprünglichen Sinne bezieht sich der Begriff auf Fehler in Software oder Algorithmen. Im Zusammenhang mit KI verweist der Begriff aber auch auf zwei weitere Arten von Fehlern: Zum einen solche, die wir als Nutzer in unseren Prompts ungewollt erzeugen. Zum anderen Fehler, die durch das KI-System in ihren Ausgaben produziert werden. Es sind vor allem diese beiden Arten von Fehlern, die man als Nutzer im Hinterkopf behalten und regelmäßig überprüfen muss.

Debunking

Debunking bezieht sich auf den Prozess, falsche oder irreführende Informationen zu widerlegen oder zu entlarven. In Bezug auf künstliche Intelligenz kann Debunking dazu verwendet werden, um falsche Annah-

men oder Mythen über KI-Technologien zu entlarven und korrekte Informationen zu verbreiten. Dieser Prozess ist wichtig, um Missverständnisse über KI zu klären und ein genaueres Verständnis zu fördern. Debunking kann durch Faktenchecks, klare Kommunikation und Aufklärung über die Funktionsweise von KI-Systemen erfolgen. Es trägt dazu bei, Vertrauen in KI-Technologien aufzubauen und die öffentliche Wahrnehmung zu verbessern.

Deepfakes

Deepfakes sind mit Künstlicher Intelligenz erzeugte Medieninhalte, bei denen Bilder, Videos oder Tonaufnahmen so manipuliert werden, dass sie echt wirken – obwohl sie gefälscht sind. Dabei werden zumeist neuronale Netzwerke eingesetzt, um das Gesicht oder die Stimme einer Person realistisch nachzubilden. Deepfakes können harmlos eingesetzt werden, etwa für Filmproduktionen oder digitale Kunst. Gleichzeitig stellen sie ein erhebliches Risiko dar, da sie gezielt zur Täuschung, Desinformation oder Rufschädigung genutzt werden können. Der Umgang mit Deepfakes erfordert daher nicht nur technische Schutzmaßnahmen, sondern auch Sensibilisierung und rechtliche Regelungen.

Deep Learning

Deep Learning ist ein Teilbereich des maschinellen Lernens, der den menschlichen Denkprozess nachahmt. Dies geschieht mithilfe von künstlichen neuronalen Netzen, die in mehreren Schichten (tiefe neuronale Netze) angeordnet sind, um Daten zu analysieren und Muster zu erkennen. Es wird vor allem zur Lösung komplexer Aufgaben in Bereichen wie Bilderkennung, Sprachverarbeitung und autonomem Fahren eingesetzt.

Deep Research

Deep Research beschreibt eine intensive KI-Recherche, bei der anders als bei einer einfachen Websuche komplexere Analyseprozesse ablaufen, etwa durch das Zusammenfassen mehrerer Quellen, kritische Einordnungen oder Hypothesenbildung. Diese Form der Recherche ist besonders nützlich bei offenen Fragen, interdisziplinären Themen oder schwer zugänglichen Daten – dauert aber länger und kann auch zu Fehlschlüssen führen.

De-Skilling

De-Skilling bezeichnet den Verlust menschlicher Fähigkeiten durch den zunehmenden Einsatz von Technologie – in der KI insbesondere durch die Auslagerung von Denk-, Schreib- oder Planungsaufgaben. Wenn etwa Übersetzungen, Texte oder Programmcode regelmäßig von KI erstellt werden, kann dies dazu führen, dass Menschen diese Kompetenzen seltener selbst trainieren und sie langfristig verlernen. Auch Entscheidungsfindung oder Problemlösen sind potenziell betroffen. De-Skilling ist damit eine unbeabsichtigte Nebenwirkung von Automatisierung, die langfristige Folgen für Bildung und Berufskompetenz haben kann.

Eliza-Effekt

Der Eliza-Effekt bezieht sich auf die Tendenz von Menschen, menschenähnliche Eigenschaften auf KI-Systeme zu projizieren. Dieser Effekt wurde erstmals bei dem Chatbot „Eliza“ beobachtet, der in den 1960er Jahren entwickelt wurde. Menschen neigen dazu, KI-Systemen mehr Intelligenz und Emotionen zuzuschreiben, als sie tatsächlich besitzen. Der Eliza-Effekt kann die Interaktion mit KI beeinflussen, da Nutzer dazu neigen, KI-Systeme menschenähnlicher zu behandeln.

Dieser Effekt ist wichtig für die Gestaltung von KI-Systemen, um realistische Erwartungen bei den Nutzern zu schaffen und Missverständnisse zu vermeiden.

Emergenz

Emergenz bezeichnet das Phänomen, dass bei bestimmten KI-Modellen Fähigkeiten auftreten, die nicht einprogrammiert oder erwartet wurden. So entwickeln große Sprachmodelle etwa die Fähigkeit zu logischem Schließen oder verstehen Sprachen, mit denen sie nur selten trainiert wurden. Solche emergenten Effekte treten überraschend auf und lassen sich nicht direkt aus der Architektur oder dem Training ableiten. Das macht leistungsfähige KI-Modelle zugleich faszinierend und schwer kontrollierbar.

Entmenschlichung

Entmenschlichung meint in der KI-Kommunikation den Verlust emotionaler Tiefe, Authentizität oder empathischer Resonanz. Da KI-Systeme kein Bewusstsein, keine echten Gefühle und kein Mitgefühl besitzen, können sie zwar sprachlich überzeugend wirken, aber keine echten menschlichen Beziehungen aufbauen. Wenn KI vermehrt in sensiblen Bereichen wie Pflege, Bildung oder psychologischer Beratung eingesetzt wird, droht eine Verarmung zwischenmenschlicher Kommunikation. Die Gefahr besteht darin, dass maschinelle Antworten als „genug“ empfunden werden.

Fake News

Fake News sind falsche oder irreführende Informationen, die absichtlich verbreitet werden, um die öffentliche Meinung zu beeinflussen oder Desinformation zu fördern. In Bezug auf Künstliche Intelligenz beinhaltet die Bekämpfung von Fake News auch die Verwendung von Algorithmen und maschinellem Lernen, um die Verbreitung und Identifizierung falscher Informationen zu überwachen und einzudämmen. KI-Systeme können beispielsweise dazu beitragen, Muster in der Verbreitung von Fake News zu erkennen und die Glaubwürdigkeit von Nachrichtenquellen zu bewerten.

Generative KI

Generative KI ist der Bereich der Künstlichen Intelligenz, bei dem KI-Systeme durch maschinelles Lernen eigenständig neue Inhalte erzeugen. Das können neben Texten z. B. auch Videos oder Bilde sein.

GPT

GPT steht für „Generative Pre-trained Transformer“ und bezieht sich auf Sprachmodelle, die auf der Transformer-Architektur basieren. Diese Modelle sind darauf trainiert, menschenähnliche Texte zu generieren und werden häufig für Aufgaben wie Textgenerierung, Übersetzung und Dialogsysteme eingesetzt. GPT-Modelle sind für ihre Fähigkeit bekannt, kontextbezogene Texte zu erzeugen und haben in verschiedenen Anwendungen der Künstlichen Intelligenz breite Anwendung gefunden. OpenAI verwendet die Abkürzung zudem für individualisierte Versionen von ChatGPT, die sich Nutzer für spezifische Aufgaben anlegen können. Über einen Link können diese GPTs auch mit anderen Nutzern geteilt werden.

Grounding

Grounding (Erden) bezeichnet in der KI die Fähigkeit eines Systems, seine Ausgaben auf überprüfbare, reale Informationen zu beziehen. Sprachmodelle wie ChatGPT erzeugen ihre Antworten auf Basis von Wahrscheinlichkeiten innerhalb von Sprachmustern

– nicht durch echtes Verstehen oder Zugriff auf eine objektive Realität. Ohne Grounding kann es daher zu sogenannten Halluzinationen kommen: Aussagen, die überzeugend klingen, aber faktisch falsch sind. Ein gutes Grounding reduziert dieses Risiko, indem es sicherstellt, dass die KI auf verlässliche Daten oder Informationsquellen zurückgreift. Ein verbreiteter Ansatz dafür ist Retrieval-Augmented Generation (RAG), bei dem externe Wissensquellen gezielt abgefragt und in die Antwortgenerierung eingebunden werden.

Guidelines

Guidelines (Richtlinien) im Bereich der Künstlichen Intelligenz beziehen sich auf etablierte Standards und Empfehlungen, die bei der Entwicklung, Implementierung und Nutzung von KI-Systemen berücksichtigt werden sollten. Diese Richtlinien können ethische, rechtliche, und technische Aspekte umfassen, um sicherzustellen, dass KI-Anwendungen verantwortungsbewusst und nachhaltig eingesetzt werden. Sie dienen dazu, potenzielle Risiken zu minimieren, die Transparenz zu erhöhen und die Fairness, Sicherheit und Datenschutz zu gewährleisten.

Haftungsprobleme

Haftungsprobleme im Bereich der Künstlichen Intelligenz beziehen sich auf die Frage, wer für Schäden oder Fehler verantwortlich ist, die durch KI-Systeme verursacht werden. Dies kann besonders komplex sein, da KI-Systeme oft autonom handeln. Die Haftung kann sich auf den Entwickler, den Betreiber oder den Anwender des KI-Systems erstrecken. Rechtliche Rahmenbedingungen und Versicherungsfragen im Zusammenhang mit KI-Haftung sind noch in der Entwicklung und werfen viele offene Fragen auf.

Halluzinationen

Im Zusammenhang mit Künstlicher Intelligenz bezieht sich eine Halluzination auf ein überzeugend formuliertes Resultat einer KI, das aber objektiv nicht der Wahrheit entspricht. Die Ursache für Halluzinationen liegt in der Funktionsweise von KI-Systemen: Diese erzeugen ihre Antworten auf Basis von Wahrscheinlichkeiten, nicht durch Verstehen. Die Wahrscheinlichkeiten, dass Halluzinationen auftreten variieren in Abhängigkeit vom gewählten KI-System.

Heuristik

Heuristik bezeichnet eine Methode zur Problemlösung, die auf Erfahrungswerten oder pragmatischen Regeln beruht anstatt auf exakten Berechnungen. Heuristiken ermöglichen schnelle, aber nicht immer perfekte Lösungen – im Gegensatz zu Algorithmen, die Schritt für Schritt eine genaue Lösung ermitteln. In KI-Systemen werden Heuristiken eingesetzt, um Entscheidungen zu treffen, wenn vollständige Informationen fehlen oder eine Berechnung zu aufwändig wäre. Sie sind besonders nützlich bei komplexen oder offenen Aufgabenstellungen, bei denen Flexibilität und Geschwindigkeit wichtiger sind als absolute Genauigkeit.

Human in the Loop

Human in the Loop ist ein Konzept, das in der Anwendung von Künstlicher Intelligenz verwendet wird. Es bezieht sich auf die Einbindung von menschlicher Expertise in den KI-Entscheidungsprozess. Dabei wird der Mensch als Teil des Prozesses betrachtet und kann beispielsweise als Kontrolleur oder Entscheider fungieren. Ziel ist es, die KI-Entscheidungen zu verbessern und sicherzustellen, dass sie ethisch verantwortungsvoll sind.

Inferenz

Inferenz ist der Prozess, bei dem ein KI-Modell aus einem gegebenen Input – etwa einem Prompt – eine Antwort ableitet. Es handelt sich dabei um eine Art Schlussfolgerung, die das Modell auf Basis seiner Trainingsdaten zieht. In der Praxis bedeutet das: Die KI berechnet, welches Wort, welcher Satz oder welche Antwort mit der höchsten Wahrscheinlichkeit als nächstes folgt. Dieser Vorgang erfolgt statistisch und nicht auf Basis eines echten Verständnisses. Die Qualität der Inferenz hängt stark von der Formulierung des Inputs und den zugrundeliegenden Daten ab. Schwächen in der Inferenz führen oft zu ungenauen oder inkonsistenten Antworten.

Jailbreaks

Jailbreaks bezeichnen Versuche, die eingebauten Sicherheitsmechanismen von KI-Systemen zu umgehen. Durch gezielte Eingaben – sogenannte „Prompt Hacks“ – wird die KI dazu gebracht, Inhalte zu generieren, die eigentlich durch ethische oder sicherheitstechnische Sperren blockiert werden sollten. Dazu gehören beispielsweise gewaltverherrlichende Aussagen, illegale Anleitungen oder diskriminierende Inhalte. Jailbreaks stellen ein ernsthaftes Risiko dar, da sie die missbräuchliche Nutzung von KI fördern können.

Künstliche Intelligenz

Künstliche Intelligenz (KI) bezieht sich auf die Fähigkeit von Maschinen, menschenähnliche Intelligenz zu simulieren. KI-Systeme können lernen, sich anpassen und selbst verbessern, um komplexe Aufgaben zu lösen. Wenn wir heute von KI sprechen, meinen wir meistens datenbasierte KI, im Gegensatz zur sogenannten regelbasierten KI. Modelle und Konzepte, die in diesem Zusammenhang häufig auftreten, sind unter anderem maschinelles Lernen, neuronale Netze und Deep Learning. KI wird in sehr vielen unterschiedlichen Bereichen eingesetzt, z. B. der Medizin, der Automobilindustrie und der Robotik.

LLM (Large Language Models)

Large Language Models sind leistungsstarke KI-Modelle, die darauf trainiert sind, natürliche Sprache zu verstehen und zu generieren. Sie basieren auf komplexen Algorithmen und verwenden riesige Datensätze, um die vielschichtigen Muster in Sprachen zu erlernen. Diese Modelle haben das Potenzial, vielfältige Aufgaben wie Übersetzung, Textgenerierung und sogar Konversationen zu bewältigen. Beispiele für solche Modelle sind ChatGPT von OpenAI oder Gemini von Google. Aufgrund ihrer Größe und Komplexität erfordern Large Language Models jedoch erhebliche Rechenressourcen für Training und Einsatz.

Lokale KI

Lokale KI bezieht sich auf KI-Systeme, die auf einem Gerät oder einer Maschine ausgeführt werden anstatt auf einem entfernten Server. Im Gegensatz zur Cloud-KI, die auf die Verarbeitung von Daten in der Cloud angewiesen ist, kann lokale KI auch ohne Internetverbindung arbeiten. Lokale KI wird oft in Anwendungen eingesetzt, die eine schnelle Verarbeitung von Daten erfordern, z. B. in autonomen Fahrzeugen oder Robotern.

Lora

Lora steht für „Long Range“ und bezieht sich auf eine drahtlose Kommunikationstechnologie, die speziell für die Übertragung kleiner Datenmengen über lange Strecken entwickelt wurde. Diese Technologie wird in verschiedenen Anwendungen der Künstlichen Intelli-

genz eingesetzt, z. B. in der Vernetzung von IoT-Geräten (Internet of Things) zur Datenerfassung und -übertragung. Lora ermöglicht es, Sensordaten über große Entfernungen hinweg zu übertragen, was z. B. in Smart Cities, in der Umweltüberwachung und in der Logistik von Nutzen ist. Die Energieeffizienz und die Fähigkeit, Signale auch in schwer zugänglichen Gebieten zu übertragen, machen Lora zu einer beliebten Wahl für drahtlose Vernetzung. Durch die Verwendung von Lora-Technologie können KI-Systeme auf eine Vielzahl von Datenquellen zugreifen und so zu einer verbesserten Entscheidungsfindung und Effizienzsteigerung beitragen.

Maschinelles Lernen

Maschinelles Lernen ist ein Teilbereich der Künstlichen Intelligenz, bei der Systeme automatisch aus großen Datenmengen lernen, ohne explizit dafür programmiert zu werden. Man unterscheidet zwischen überwachtem (supervised), unüberwachtem (unsupervised) und verstärkendem (reinforced) maschinellen Lernen. Beim überwachten maschinellen Lernen erhalten die Systeme Daten gemeinsam mit zusätzlichen Informationen, aus denen sie dann Regeln und Muster ableiten können. Beim unüberwachten maschinellen Lernen fehlen die zusätzlichen Informationen. Hier muss das System alleine Muster erkennen. Beim verstärkenden maschinellen Lernen erhalten die Systeme vorab gar keine Daten. Stattdessen durchlaufen die Systeme in einer Simulationsumgebung ein Trial-and-Error-Verfahren, bei dem sie für korrektes Verhalten belohnt werden. Auf diese Weise lernt das System, wie es (nicht) handeln soll. Eine der Grundlagen des maschinellen Lernens ist das Deep Learning, das auf neuronalen Netzen basiert und einen Großteil der Merkmalsextraktion automatisiert, wodurch ein Teil der manuellen Eingriffe entfällt und die Verwendung größerer Datenmengen ermöglicht wird.

Neuronales Netzwerk

Ein neuronales Netzwerk ist ein Modell, das von biologischen neuronalen Netzwerken inspiriert wurde und in der Künstlichen Intelligenz weit verbreitet ist. Es besteht aus einer Sammlung miteinander verbundener Knoten, die als Neuronen bezeichnet werden. Diese Neuronen nehmen Eingaben entgegen, verarbeiten sie und geben Ausgaben weiter. Durch das Lernen aus Daten kann ein neuronales Netzwerk komplexe Muster erkennen und abstrakte Aufgaben wie Bilderkennung oder Sprachverarbeitung durchführen. Die Leistung eines neuronalen Netzwerks hängt von seiner Architektur, den verwendeten Algorithmen und den Trainingsdatensätzen ab.

Playground

Ein Playground ist eine virtuelle Umgebung, die es ermöglicht, mit KI-Modellen und -Algorithmen zu experimentieren, ohne aufwendige Infrastruktur aufsetzen zu müssen. In einem solchen Sandbox-ähnlichen Setting können Nutzer verschiedene Parameter und Datensätze testen, um das Verhalten von KI-Modellen zu untersuchen und zu verstehen. Oft bieten Unternehmen und Forschungseinrichtungen eigene Playgrounds an, um die Nutzung ihrer KI-Technologien zu fördern. Der Playground dient als sicherer Raum, um neue Ideen zu erproben, Fehler zu machen und das Verständnis für KI-Systeme zu vertiefen. Durch die Interaktion mit einem Playground können Entwickler und Forscher ihre Fähigkeiten im Umgang mit KI verbessern und neue Anwendungen entwickeln.

Prompt

Ein Prompt ist in der Anwendung der Künstlichen Intelligenz ein Text oder eine Aussage, die als Eingabe für ein Modell dient und den Kontext oder das gewünschte Ergebnis beschreibt. Es kann sich um eine Frage, einen Satz oder sogar um Stichpunkte handeln. Der Prompt wird verwendet, um dem Modell klare Anweisungen zu geben und seine Ausgabe entsprechend zu lenken. Ein gut formulierter Prompt kann dazu beitragen, präzise und relevante Antworten von KI-Modellen zu erhalten.

Reasoning

Reasoning bezeichnet die Fähigkeit von KI-Systemen, logische Schlüsse zu ziehen. Im Unterschied zu rein statistischen Vorhersagen geht es beim Reasoning darum, Zusammenhänge zu erkennen, Argumente zu strukturieren und folgerichtige Aussagen zu treffen. Sprachmodelle verfügen nur über eine eingeschränkte Fähigkeit zum Reasoning, da sie nicht wirklich „denken“, sondern auf Basis von Wahrscheinlichkeiten reagieren. Methoden wie das Chain-of-Thought-Prompting versuchen, diese Schwäche auszugleichen, indem sie die KI dazu bringen, ihre Gedankengänge Schritt für Schritt zu erklären. Dennoch bleibt Reasoning eine Herausforderung für KI – besonders bei komplexen Aufgaben, die hohe Genauigkeit und Nachvollziehbarkeit erfordern.

Regulierung

Regulierung bezieht sich auf die Gesetze, Vorschriften und Standards, die von Regierungen und anderen Institutionen festgelegt werden, um die Entwicklung und Anwendung von KI zu steuern und zu kontrollieren. Ziel ist es, sicherzustellen, dass KI-Systeme verantwortungsvoll eingesetzt werden und keine negativen Auswirkungen auf die Gesellschaft haben. Die Regulierung von KI ist ein komplexes Thema, das viele Fragen aufwirft, z. B. wer für die Regulierung verantwortlich ist und wie sie durchgesetzt werden kann.

Risikostufen

KI-Systeme werden anhand ihres potenziellen Risikos für Gesundheit, Sicherheit und Grundrechte von Personen meist in vier Risikostufen eingeteilt.

1. Inakzeptables Risiko: KI-Systeme mit inakzeptablem Risiko können schwere oder irreversible Schäden verursachen. Beispiele sind KI-Systeme, die für autonome Waffensysteme oder die Überwachung von Personen eingesetzt werden.
2. Hohes Risiko: KI-Systeme mit hohem Risiko können erhebliche Schäden verursachen. Beispiele sind KI-Systeme, die für die medizinische Diagnose oder das autonome Fahren eingesetzt werden.
3. Begrenztes Risiko: KI-Systeme mit begrenztem Risiko können leichte oder moderate Schäden verursachen. Beispiele sind KI-Systeme, die für die Personaleinsatzplanung oder die Kundenberatung eingesetzt werden.
4. Minimales Risiko: KI-Systeme mit minimalem Risiko verursachen keine oder nur sehr geringe Schäden. Beispiele sind KI-Systeme, die für die Produktauswahl oder die Werbung eingesetzt werden.

Roboterjournalismus

Roboterjournalismus bezieht sich auf den Einsatz von KI und Algorithmen, um automatisch Nachrichtenartikel zu erstellen. Dabei werden Daten und Informationen aus verschiedenen Quellen gesammelt und verarbeitet, um einen Artikel zu generieren. Der Einsatz von Roboterjournalismus kann dazu beitra-

gen, den Prozess der Nachrichtenerstellung zu beschleunigen und zu automatisieren. Allerdings gibt es auch Bedenken hinsichtlich der Qualität und Objektivität solcher Artikel, da sie nicht von menschlichen Journalisten verfasst werden und es bisweilen an sorgfältiger redaktionelle Kontrolle hapert.

Sicherheitsrisiken

Zu den Sicherheitsrisiken in der Anwendung Künstlicher Intelligenz zählen etwa Datenschutzverletzungen, unerwünschte Manipulationen durch fehlerhafte Algorithmen und die Möglichkeit von Cyberangriffen auf KI-Systeme. Die Identifizierung und Bewältigung dieser Risiken ist entscheidend, um das Vertrauen in KI-Technologien zu stärken und ihre sichere Anwendung zu gewährleisten. Unternehmen und Entwickler sind daher gefordert, Sicherheitsmaßnahmen einzubauen, um potenziellen Risiken im Zusammenhang mit Künstlicher Intelligenz zu minimieren.

Superintelligenz

Superintelligenz bezeichnet eine hypothetische Form Künstlicher Intelligenz, die der menschlichen Intelligenz in nahezu allen Bereichen überlegen ist. Während sie enorme Chancen für Wissenschaft, Medizin oder globale Krisenbewältigung bieten könnte, birgt sie zugleich erhebliche Risiken. Eine unkontrollierte Superintelligenz könnte Ziele verfolgen, die nicht mit menschlichen Werten vereinbar sind oder sich dem menschlichen Einfluss entziehen. Deshalb gilt ihre mögliche Entstehung als zentrales Thema der KI-Sicherheitsforschung.

Symbolische KI

Symbolische KI, auch bekannt als GOFAI („Good Old-Fashioned AI“) , bezeichnet einen frühen Ansatz der KI-Forschung, bei dem Wissen in Form von klaren Regeln, Symbolen und Logik dargestellt wurde. Man spricht in diesem Zusammenhang deshalb auch von regelbasierter KI. Systeme dieser Art arbeiteten nach dem Prinzip „Wenn A, dann B“ und waren besonders gut in streng strukturierten Bereichen wie Schach oder Expertensystemen. Im Gegensatz dazu lernen moderne (datenbasierte) KI-Modelle nicht durch festgelegte Regeln, sondern aus riesigen Datenmengen, indem sie Wahrscheinlichkeiten berechnen und Muster ableiten. Das führt dazu, dass heutige KI flexibler und oft erstaunlich leistungsfähig ist – aber auch weniger transparent. Symbolische KI war erklärbar, aber begrenzt; datenbasierte KI ist lernfähig, aber schwer durchschaubar.

Temperatur

Temperatur ist in Bezug auf Künstliche Intelligenz ein Parameter, der die Zufälligkeit und Kreativität der KI-Ausgabe steuert. Eine niedrige Temperatur führt zu vorhersehbaren und konsistenten Antworten, während eine hohe Temperatur mehr Kreativität und Vielfalt ermöglicht, aber auch zu weniger vorhersehbaren Ergebnissen führen kann.

Token

Ein Token ist die kleinste Einheit, in die ein Text oder eine andere Datenmenge aufgeteilt wird, um von einem KI-Modell verarbeitet zu werden. Ein Token kann beispielsweise ein einzelnes Wort, ein Zeichen oder sogar ein Pixel sein.

Transparenzhinweise

Ein Transparenzhinweis in der Anwendung von Künstlicher Intelligenz bezieht sich auf die Offenlegung von Informationen über die Funktionsweise, Datenquellen und Entscheidungsprozesse eines KI-Systems. Dieser Hin-

weis soll Nutzern helfen zu verstehen, wie die KI arbeitet und welche Daten sie verwendet, um fundierte Entscheidungen zu treffen. Ein transparentes KI-System trägt zur Vertrauensbildung bei und ermöglicht es den Nutzern, die Auswirkungen und potenziellen Bias besser zu verstehen.

Turing-Test

Der Turing-Test ist ein Test, der von Alan Turing im Jahr 1950 entwickelt wurde. Das Ziel des Tests war es, zu bestimmen, ob eine Maschine intelligentes Verhalten zeigen kann, das von einem menschlichen Verhalten nicht zu unterscheiden ist. Der Test besteht darin, dass ein Mensch und eine Maschine in getrennten Räumen Fragen von einem Dritten beantworten müssen. Wenn der Dritte nicht in der Lage ist zu unterscheiden, wer die Maschine und wer der Mensch ist, gilt die Maschine als intelligent. Der Turing-Test ist ein wichtiger Meilenstein in der Entwicklung der Künstlichen Intelligenz und hat dazu beigetragen, die Forschung in diesem Bereich voranzutreiben.

Urheberrecht und KI

Das Urheberrecht ist ein Recht des geistigen Eigentums, das dem Urheber eines Werks die ausschließliche Befugnis zur Verwertung dieses Werks einräumt. Urheberrechtlich geschützt sind nur Werke, die eine persönliche geistige Schöpfung des Urhebers sind. In Anwendungen der Künstlichen Intelligenz stellt das Urheberrecht einige Herausforderungen dar. So ist beispielsweise unklar, ob KI-generierte Werke urheberrechtlich geschützt sind. Außerdem stellt sich die Frage, wie das Urheberrecht die Nutzung von KI-gestützten Technologien wie Text- und Bilderkennung beeinflusst.

Verbreiterhaftung

Verbreiterhaftung ist ein rechtlicher Begriff, der besagt, dass eine Person oder Organisation für die Verbreitung von rechtswidrigen Inhalten haftbar gemacht werden kann, auch wenn sie diese nicht selbst erstellt hat bzw. durch oder mit Hilfe von Künstlicher Intelligenz hat erstellen lassen. Dies kann beispielsweise bei der Verbreitung von urheberrechtlich geschütztem Material oder bei der Verbreitung von Hassrede im Internet der Fall sein. Die Verbreiterhaftung gilt auch für Plattformbetreiber, die Inhalte von Nutzern auf ihren Plattformen veröffentlichen.

Zitierrecht

Das Zitierrecht bezieht sich auf die rechtliche Erlaubnis, Texte, Ideen oder Werke anderer Personen unter bestimmten Bedingungen zu zitieren oder zu verwenden. Es ermöglicht es, kurze Auszüge aus geschützten Werken zu verwenden, solange die Quelle ordnungsgemäß angegeben wird. Das Zitierrecht ist wichtig, um die Integrität des geistigen Eigentums zu wahren und gleichzeitig den Austausch von Wissen und Informationen zu fördern. In der KI-Anwendung kann das Zitierrecht relevant sein, wenn Algorithmen auf Textdaten trainiert werden, die aus verschiedenen Quellen stammen. Es ist wichtig, die rechtlichen Aspekte des Zitierrechts zu beachten, um die Einhaltung von Urheberrechten und Lizenzvereinbarungen sicherzustellen.