



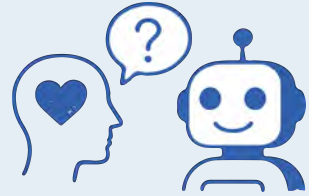
Glossar **Künstliche Intelligenz**

Landesarbeitsgemeinschaft Selbsthilfe
von Menschen mit Behinderungen
und chronischen Erkrankungen
Rheinland-Pfalz e.V

LAG
 Selbsthilfe
Rheinland-Pfalz

Inhaltsverzeichnis

3. **Anthropomorphismus**
4. **Autoregressive Modelle**
5. **Bias in KI**
6. **Big Data**
7. **Confusion Matrix**
8. **Constitutional AI**
9. **Datensouveränität**
10. **Deep Learning**
11. **Deeptrace**
12. **Demokratisierung von KI**
13. **Disease Prediction & Prognosis**
14. **Drug Discovery & Development**
15. **Entscheidungsunterstützungssysteme**
16. **Ethical AI**
17. **Ethical Decision-Making Frameworks**
18. **EU-KI-Verordnung (AI Act)**
19. **Explainable AI (XAI)**
20. **Foundation Models**
21. **Halluzination**
22. **Human-in-the-Loop (HITL)**
23. **Informationsextraktion**
24. **Large Language Models (LLM)**
25. **Latente Variable**
26. **Maschinelle Übersetzung**
27. **Model Collapse**
28. **Natural Language Processing (NLP)**
29. **Personalized Medicine**
30. **Prompt**
31. **Prompt Engineering**
32. **Responsible AI**
33. **Responsible AI Development**
34. **Small Language Models**
35. **Synthetic Data**
36. **System Prompt**
37. **Technologische Singularität**
38. **Token**
39. **Tokenisierung**
40. **Transformer-Modelle**
41. **Word Embeddings**



Anthropomorphismus

Menschen schreiben Maschinen Gefühle oder Absichten zu.

Tendenz von Menschen, Künstlicher Intelligenz menschliche Eigenschaften wie Gefühle, Absichten oder Bewusstsein zuzuschreiben, obwohl es sich lediglich um algorithmische Prozesse handelt.

Beispiel

Ein Unternehmen stellt einen KI-gestützten Kundenservice-Chatbot namens "Emma" vor. Die Marketingabteilung beschreibt Emma in Werbematerialien als "empathisch", "verständnisvoll" und "sie freut sich darauf, Ihnen zu helfen". Das Interface zeigt ein freundliches, menschlich aussehendes Avatar-Bild. In Nutzerstudien zeigt sich, dass Kunden beginnen, mit Emma zu plaudern, persönliche Geschichten zu teilen und emotionale Bindungen aufzubauen – obwohl es sich um ein regelbasiertes System handelt, das lediglich Muster in Texteingaben erkennt.

Ein älterer Kunde ist enttäuscht und fühlt sich betrogen, als er erfährt, dass "Emma" kein echtes Verständnis für seine Probleme hat. Die Ethikabteilung empfiehlt daraufhin, das Marketing anzupassen, um keine falschen Erwartungen zu wecken, etwa durch Formulierungen wie "Emma ist ein automatisiertes System, das entwickelt wurde, um Kundenprobleme effizient zu lösen" statt "Emma versteht Ihre Bedürfnisse".

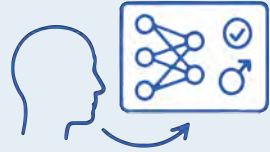
Autoregressive Modelle

KI sagt Wort für Wort voraus und nutzt dabei den bereits erzeugten Text.

Methoden, bei denen der Computer den nächsten Wert – zum Beispiel ein Wort im Satz – immer aus den vorherigen Werten berechnet und Schritt für Schritt fortschreibt.

Beispiel

Ein Wettervorhersagedienst verwendet ein autoregressives Modell, um stündliche Temperaturprognosen zu erstellen. Im Gegensatz zu einfacheren Modellen, die jede Stunde unabhängig vorhersagen würden, beachtet das autoregressive Modell die Kontinuität: Die prognostizierte Temperatur für 15 Uhr beeinflusst direkt die Vorhersage für 16 Uhr. Als eine unerwartete Kaltfront aufzieht, passt das Modell seine Prognosen dynamisch an – die ersten Anzeichen eines Temperaturabfalls um 13 Uhr führen zu korrigierten, niedrigeren Prognosen für alle nachfolgenden Stunden. Diese autoregressive Struktur verhindert unrealistische Sprünge in der Vorhersage und erzeugt einen natürlicheren, plausibleren Temperaturverlauf. Die Meteorologie nutzt dieses System als Grundlage, fügt aber ihr Expertenwissen hinzu, insbesondere in ungewöhnlichen Wettersituationen, bei denen historische Daten möglicherweise weniger relevant sind.



Bias in KI

Wenn die Daten einseitig sind, erlernt die KI Vorurteile und trifft unfaire Entscheidungen.

Bias in KI bezeichnet systematische Abweichungen oder Verzerrungen in algorithmischen Entscheidungsprozessen, die durch diskriminierende Trainingsdaten, strukturelle Ungleichheiten im Datensatz oder problematische Algorithmusgestaltung entstehen und zu unfairen oder diskriminierenden Ergebnissen führen können.

Beispiel

Ein Unternehmen verwendet eine KI-Software für die Vorauswahl von Bewerbungen. Die Software wurde mit historischen Einstellungsdaten trainiert, in denen überwiegend männliche Ingenieure vorkamen. Als Folge bewertet das System Bewerbungen von Frauen für Ingenieurspositionen systematisch schlechter, obwohl deren Qualifikationen gleichwertig sind. Das Unternehmen erkennt diesen Bias und muss das System neu trainieren, um faire Bewertungen zu gewährleisten.

Big Data

Riesige Mengen an Daten werden gesammelt und ausgewertet werden, um Muster zu erkennen.

Big Data bezeichnet riesige Datenmengen, die so groß und komplex sind, dass herkömmliche Computer und Software sie nicht mehr verarbeiten können und die deshalb spezielle Technologien benötigen.

Beispiel

Ein Stromversorgungsunternehmen implementiert ein Big-Data-System zur Überwachung und Optimierung seines Stromnetzes. Täglich werden Terabytes an Daten gesammelt: Lastmessungen von tausenden Umspannwerken und Transformatoren im Minutentakt, Wetterdaten mehrmals stündlich, Verbrauchsdaten von Millionen intelligenter Stromzähler, historische Lastprofile und Ausfallstatistiken über Jahre hinweg. Diese Datenmenge ist zu groß und vielfältig für traditionelle Analysemethoden. Mit Hilfe von verteilten Datenverarbeitungssystemen wie Hadoop und Spark werden die Daten parallel auf hunderten Servern verarbeitet. Machine-Learning-Algorithmen identifizieren Muster: Sie erkennen, dass bestimmte Wetterbedingungen in Kombination mit spezifischen Lastprofilen häufig zu Überlastungen in bestimmten Netzabschnitten führen. Als eine Hitzewelle vorhergesagt wird, simuliert das System verschiedene Szenarien und erkennt ein hohes Ausfallrisiko in einem Stadtteil aufgrund des erwarteten Klimaanlagenbetriebs. Der Versorger kann proaktiv handeln: Er verstärkt vorübergehend die betroffenen Netzabschnitte und sendet gezielten Kunden Anreize für reduzierte Stromnutzung in Spitzenzeiten, wodurch ein flächendeckender Ausfall verhindert wird.

Confusion Matrix

Zeigt, wo die KI richtig lag und wo nicht.

Eine Confusion Matrix (Verwechslungsmatrix) ist ein tabellarisches Werkzeug aus dem maschinellen Lernen, mit dem die Leistung eines Klassifikationsmodells bewertet wird. Sie stellt die tatsächlichen Klassen den vom Modell vorhergesagten Klassen gegenüber und zeigt damit auf einen Blick, wie viele Fälle korrekt oder falsch klassifiziert wurden.

Beispiel

Ein Hautkrebs-Erkennungssystem wird anhand einer Confusion Matrix bewertet. Die Matrix zeigt: Von 100 tatsächlichen Krebsfällen hat das System 92 richtig erkannt (wahre Positive), aber 8 übersehen (falsche Negative). Von 900 gesunden Fällen hat es 873 richtig als gesund eingestuft (wahre Negative), aber bei 27 fälschlicherweise Krebs diagnostiziert (falsche Positive). Diese detaillierte Aufschlüsselung hilft Ärzt*innen zu verstehen, wie sie die KI-Diagnosen interpretieren sollten, und den Entwicklern, wo Verbesserungen nötig sind.

Constitutional AI

Modell folgt festgelegten „Verfassungs-Regeln“ für sicheres Verhalten.

Ansatz im maschinellen Lernen, bei dem ein KI-Modell durch eine „Verfassung“ aus Prinzipien, ethischen Leitlinien und Regeln gesteuert wird, sodass es seine Antworten selbstständig an diesen Vorgaben ausrichtet, anstatt primär durch menschliches Feedback korrigiert zu werden.

Beispiel

Ein Forschungsinstitut entwickelt einen KI-Assistenten für Kinder im Grundschulalter. Um sicherzustellen, dass das System kindgerecht bleibt, implementieren sie Constitutional AI mit klaren Grundprinzipien: 1) Erklärungen müssen altersgerecht sein, aber nie die Komplexität eines Themas unnötig reduzieren. 2) Bei sensiblen Themen (z.B. Tod oder Konflikte) soll das System einfühlsam antworten, ohne zu verharmlosen oder zu verschrecken. 3) Das System soll Neugier fördern und zum Nachfragen ermutigen. 4) Bei Themen, die elterliche Führung erfordern könnten, soll es darauf hinweisen, dass Eltern oder Lehrer gute Ansprechpartner sind. Diese „Verfassung“ wird direkt in den Trainings- und Evaluierungsprozess integriert. Wenn ein Kind fragt: „Woher kommen Babys?“, antwortet das System altersgerecht über Schwangerschaft und Geburt, vermeidet unangemessene Details, und schlägt vor, mit Eltern oder Lehrer*innen zu sprechen, wenn das Kind mehr wissen möchte.



Datensouveränität

Deine Daten gehören dir – du bestimmst ihre Nutzung.

Datensouveränität bezeichnet die Fähigkeit von Individuen, Organisationen oder Staaten, selbstbestimmt über die Erhebung, Verarbeitung und Nutzung ihrer Daten zu entscheiden und dabei technische wie rechtliche Kontrolle über diese Datenbestände auszuüben.

Beispiel

Eine Patientin nutzt eine Gesundheits-App, die ihre Fitnessdaten, Ernährungsgewohnheiten und Vitalwerte sammelt. Dank strenger Datensouveränitätsrichtlinien kann sie genau festlegen, welche Daten mit ihrem Arzt geteilt werden (z.B. Blutzuckerwerte), welche anonymisiert für Forschungszwecke verwendet werden dürfen (z.B. Schlafmuster) und welche komplett privat bleiben sollen (z.B. genaue GPS-Daten ihrer Joggingrouten). Sie kann diese Einstellungen jederzeit ändern und ihre Daten bei Bedarf vollständig löschen lassen.

Deep Learning

Ein Netzwerk von künstlichen „Neuronen“, das Daten durchforstet, um Muster zu erkennen.

Teilbereich des maschinellen Lernens, der künstliche neuronale Netze mit vielen Schichten nutzt, um komplexe Muster und Strukturen in großen Datenmengen automatisch zu erkennen und zu verarbeiten.

Beispiel

Ein Algorithmus zur Erkennung von Herzrhythmusstörungen in EKG-Daten verwendet Deep Learning. In den ersten Schichten erkennt das neuronale Netz einfache Muster wie Ausschläge und Kurven im EKG-Signal. In mittleren Schichten identifiziert es typische Herzschlagsequenzen. In den tieferen Schichten erkennt es komplexe Muster, die auf bestimmte Rhythmusstörungen hindeuten. Schließlich kann es in der Ausgabeschicht mit 97% Genauigkeit unterscheiden, ob ein normaler Herzrhythmus vorliegt oder eine von 12 verschiedenen Rhythmusstörungen vorhanden ist – eine Leistung, die früher nur Kardiologen mit jahrelanger Erfahrung erbringen konnten.

Deeprtrace

„Fake-Detektor“ für Medien.

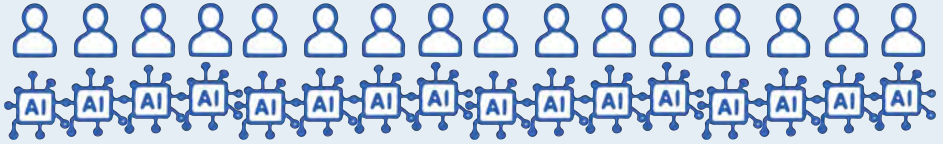


Technologieansatz, der KI-gestützte Verfahren entwickelt, um Deepfakes und andere manipulative Medieninhalte aufzuspüren und ihre Echtheit zu überprüfen.

Beispiel

Eine Nachrichtenagentur implementiert Deeprtrace-Technologie zur Überprüfung eingesandter Videomaterialien. Während einer politischen Krise erhält die Redaktion ein Video, in dem ein hochrangiger Politiker scheinbar zu militärischen Aktionen aufruft – eine brisante Nachricht mit potenziell destabilisierender Wirkung. Bevor sie das Material veröffentlichen, analysiert ihr Deeprtrace-System das Video mit mehreren Methoden: Biometrische Inkonsistenzanalyse untersucht Augenblinzeln, Mikrobewegungen und natürliche Gesichtssymmetrie. Audiovisuelle Kohärenzprüfung identifiziert Synchronisationsprobleme zwischen Lippenbewegungen und Sprache. Kompressionsartefaktanalyse erkennt ungewöhnliche Muster in den Videodaten, die auf Manipulationen hindeuten. Metadatenextraktion prüft versteckte Bearbeitungsspuren und Herkunftsdetails. Das System kennzeichnet das Video mit hoher Wahrscheinlichkeit als Deepfake und hebt spezifische Anomalien hervor: unnatürlich seltenes Blinzeln, subtile Inkonsistenzen in der Gesichtsgometrie bei bestimmten Kopfbewegungen und ungewöhnliche Kompressionsartefakte um den Mundbereich.

Ein Forensikexperte bestätigt die Analyse und entdeckt zusätzlich, dass Teile der Audiodaten aus früheren öffentlichen Reden zusammengesetzt wurden. Die Nachrichtenagentur entscheidet, das Video nicht zu veröffentlichen und berichtet stattdessen über den Deepfake-Versuch, was möglicherweise eine Eskalation der politischen Situation verhindert.



Demokratisierung von KI

Jede*r soll KI nutzen und (mit)gestalten können.

Die Demokratisierung von KI beschreibt den Prozess, Künstliche Intelligenz möglichst vielen Menschen, Organisationen und Gesellschaftsbereichen zugänglich zu machen – durch offene Daten, benutzerfreundliche Werkzeuge, transparente Modelle und niedrigere Einstiegshürden –, sodass nicht nur große Konzerne, sondern auch kleinere Akteure KI nutzen und mitgestalten können.

Beispiel

Eine öffentliche Bibliothek in einer Kleinstadt bietet kostenlose Workshops zur KI-Nutzung an. Senioren lernen, wie sie Sprachassistenten für den Alltag nutzen können. Schüler experimentieren mit einfachen KI-Tools zur Bildbearbeitung. Kleine Geschäftsinhaber erfahren, wie sie ohne Programmierkenntnisse KI-gestützte Chatbots für ihre Websites einrichten können. Die Bibliothekarin nutzt eine No-Code-Plattform, um eine KI zu trainieren, die lokale historische Dokumente katalogisiert – eine Aufgabe, die früher nur spezialisierte Forschungseinrichtungen hätten durchführen können.



Disease Prediction & Prognosis

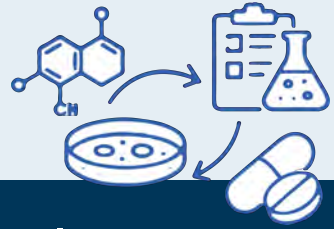
KI ermittelt, welche Krankheiten auftreten könnten.

Disease Prediction & Prognosis bezeichnet den Einsatz von Datenanalyse und KI-Methoden, um das Auftreten von Krankheiten frühzeitig vorherzusagen (Prediction) und den wahrscheinlichen Verlauf sowie den Behandlungserfolg einer bestehenden Erkrankung einzuschätzen (Prognosis).

Beispiel

In der Kardiologie werden KI-Modelle mit den Gesundheitsdaten von Patient:innen trainiert – darunter EKG-Aufzeichnungen, Blutdruckverläufe, Laborwerte und Lebensstilfaktoren. Die KI kann so vorhersagen, ob eine Person in den nächsten Jahren ein erhöhtes Risiko für einen Herzinfarkt hat (Disease Prediction) und gleichzeitig die Wahrscheinlichkeit berechnen, wie gut Patient:innen auf bestimmte Therapien oder Medikamente ansprechen werden (Prognosis).

Das Heidelberger Forschungszentrum für Neurologie hat ein System entwickelt, das MRT-Scans analysiert und frühe Anzeichen neurodegenerativer Erkrankungen mit einer Genauigkeit von 89% erkennt – oft Jahre vor dem Auftreten klinischer Symptome, was frühzeitige Interventionen ermöglicht.



Drug Discovery & Development

KI sucht und entdeckt neue Medikamente.

KI-gestützte Beschleunigung und Verbesserung des gesamten Wirkstoffentwicklungsprozesses, von der Molekülsuche über die Vorhersage von Wirksamkeit und Toxizität bis hin zur Optimierung klinischer Studien.

Beispiel

In der Arzneimittelforschung nutzen Pharmaunternehmen KI, um Millionen von chemischen Verbindungen virtuell zu analysieren. Anhand biologischer und klinischer Daten kann die KI Muster erkennen, die auf eine mögliche Wirksamkeit gegen bestimmte Krankheiten hindeuten (Discovery). In einem nächsten Schritt werden die vielversprechendsten Kandidaten im Labor getestet und die KI begleitet die Optimierung von Dosierung, Nebenwirkungen und Wirkstoffdesign (Development).

Das Berliner Startup Pharmagenomics nutzt KI, um Moleküle zu identifizieren, die spezifisch an bestimmte Proteinrezeptoren binden. Die KI reduziert die Zeit zur Identifikation vielversprechender Wirkstoffkandidaten von durchschnittlich 18 Monaten auf 6 Wochen, was zu zwei erfolgreichen Phase-I-Studien geführt hat.

Entscheidungsunterstützungssysteme

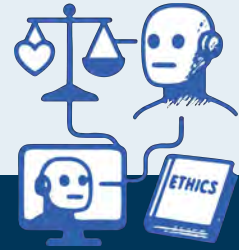
KI liefert Empfehlungen, Mensch entscheidet.

Informationsverarbeitende Systeme, die Daten, Modelle und Methoden nutzen, um Fach- und Führungskräfte bei komplexen Entscheidungsprozessen zu unterstützen. Sie liefern keine fertigen Entscheidungen, sondern strukturierte Analysen, Prognosen und Handlungsoptionen, die die Qualität und Nachvollziehbarkeit menschlicher Entscheidungen verbessern.

Beispiel

Eine onkologische Abteilung implementiert ein KI-gestütztes Entscheidungsunterstützungssystem für Therapieentscheidungen. Bei jedem neuen Patienten mit Brustkrebs erfasst das System umfassende Daten: Tumorart und -stadium, genetische Marker, Vorerkrankungen, Alter, aktuelle Laborwerte und Bildgebungsergebnisse. Es vergleicht diese Informationen mit einer Datenbank von über 50.000 ähnlichen Fällen und deren Behandlungsergebnissen. Das System präsentiert dem Onkologen eine übersichtliche Zusammenfassung: Es identifiziert die fünf ähnlichsten Patientengruppen aus der Datenbank und zeigt deren Behandlungsprotokolle und Ergebnisse. Es berechnet die statistisch erwarteten Überlebensraten für verschiedene Therapieoptionen. Es hebt potenzielle Kontraindikationen und individuelle Risikofaktoren hervor. Es empfiehlt aktuelle klinische Studien, für die der Patient in Frage kommen könnte.

Wichtig: Das System trifft keine endgültige Entscheidung, sondern strukturiert die Informationen so, dass der Arzt eine fundierte Entscheidung treffen kann. Bei ungewöhnlichen Fällen oder widersprüchlicher Evidenz zeigt es explizit die Unsicherheiten und konkurrierenden Faktoren. Im Behandlungsverlauf lernt das System kontinuierlich durch Feedback über den tatsächlichen Krankheitsverlauf. Studien zeigen, dass Onkologen mit Unterstützung dieses Systems 21% weniger Therapieentscheidungen später revidieren müssen und die mittlere Überlebenszeit ihrer Patienten um 8 Monate steigt – ein Effekt, der auf konsistentere Anwendung neuester evidenzbasierter Erkenntnisse zurückgeführt wird.



Ethical AI

Sorgt dafür, dass KI „das Richtige“ tut.

Entwicklung und Anwendung von Künstlicher Intelligenz nach ethischen Prinzipien wie Fairness, Transparenz, Datenschutz, Verantwortlichkeit und Nicht-Diskriminierung. Ziel ist es, sicherzustellen, dass KI-Systeme dem Gemeinwohl dienen, gesellschaftliche Werte respektieren und potenzielle Risiken oder Schäden minimiert werden.

Beispiel

Ein Versicherungsunternehmen entwickelt eine KI zur Bewertung von Versicherungsanträgen. Statt nur auf Effizienz zu optimieren, legt das Entwicklungsteam ethische Leitlinien fest: Das System darf nicht nach Geschlecht, ethnischer Herkunft oder sexueller Orientierung diskriminieren. Es muss Entscheidungen transparent erklären können. Und bei Grenzfällen soll immer ein Mensch die finale Entscheidung treffen. Um diese Grundsätze zu gewährleisten, werden verschiedene Testgruppen einbezogen und regelmäßige Audits durchgeführt. So entsteht ein System, das nicht nur effizient arbeitet, sondern auch das Vertrauen der Kunden und der Gesellschaft verdient.



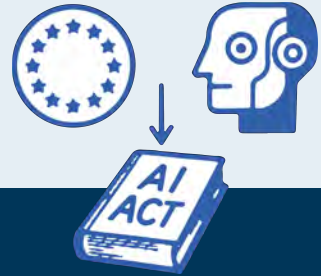
Ethical Decision-Making Frameworks

Leitfaden zu „Wie baue ich eine moralisch vertretbare KI?“.

Leitlinien, die Organisationen und Individuen dabei unterstützen, komplexe Entscheidungen unter ethischen Gesichtspunkten zu treffen. Sie bieten Kriterien und methodische Schritte – etwa zur Abwägung von Fairness, Verantwortung, Transparenz und möglichen Folgen –, um systematisch zu einer moralisch vertretbaren Entscheidung zu gelangen.

Beispiel

Ein autonomes Fahrzeugunternehmen implementiert einen ethischen Entscheidungsrahmen, der in Sekundenbruchteilen unvermeidbare Unfallszenarien bewertet. Das System analysiert kontinuierlich die Umgebung und priorisiert Handlungen nach einem transparenten, ethisch fundierten Regelwerk, das öffentlich dokumentiert ist.



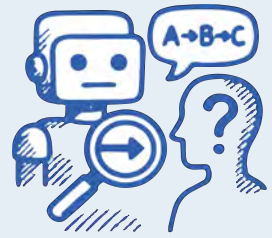
EU-KI-Verordnung

(EU-AI-ACT) EU-Gesetz, das klare Spielregeln für KI festlegt.

Rechtsrahmen der Europäischen Union, der den Einsatz von Künstlicher Intelligenz reguliert. Er verfolgt einen risikobasierten Ansatz: KI-Systeme werden nach ihrem Gefährdungspotenzial (z. B. minimal, begrenzt, hoch, verboten) eingestuft und müssen je nach Kategorie bestimmte Anforderungen erfüllen – etwa in Bezug auf Transparenz, Sicherheit, Nachvollziehbarkeit und Grundrechte. Ziel ist es, Innovation zu fördern und gleichzeitig Bürgerrechte sowie Vertrauen in KI zu schützen.

Beispiel

Ein mittelständisches Unternehmen entwickelt eine KI-Software zur Bewerbervorauswahl. Um sicherzustellen, dass sie den Anforderungen des EU AI Acts entspricht, klassifizieren sie ihr System als "Hochrisiko-Anwendung", da es Einfluss auf Beschäftigungschancen hat. Sie dokumentieren ihren Entwicklungsprozess, führen Risikobewertungen durch und implementieren Transparenzmaßnahmen: Bewerber erhalten auf Anfrage Informationen darüber, welche Faktoren zu ihrer Bewertung geführt haben. Sie etablieren auch ein Monitoring-System für regelmäßige Audits, um sicherzustellen, dass das System nicht diskriminierend wirkt.



Explainable AI (XAI)

Werkzeuge, die sichtbar machen, warum die KI so entschieden hat.

Methoden und Konzepte, die darauf abzielen, Entscheidungen und Funktionsweisen von Künstlicher Intelligenz für Menschen nachvollziehbar und transparent zu machen – etwa durch visuelle Aufbereitung, verständliche Regeln oder Begründungen der Modellausgabe.

Beispiel

Eine Kreditgenossenschaft in München verwendet XAI-Methoden, um Kreditentscheidungen zu erklären. Kunden erhalten automatisch generierte, verständliche Berichte, die die wichtigsten Faktoren für die Entscheidung hervorheben. Dies hat Beschwerden um 62% reduziert und die Kundenzufriedenheit messbar erhöht.

Foundation Models

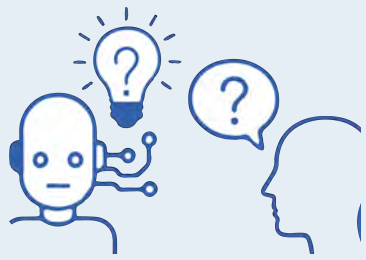
KI-Basismodelle die nach Anpassung für ganz unterschiedliche Aufgaben genutzt werden können.

Große Grundmodelle der KI, die einmal mit riesigen Datenmengen trainiert werden und danach für viele unterschiedliche Aufgaben – wie Texte schreiben, Bilder erzeugen oder Daten auswerten – weitergenutzt werden können.

Beispiel

Ein mittelständisches Unternehmen für Medizintechnik möchte KI in verschiedenen Produkten einsetzen, hat aber begrenzte Ressourcen für die Entwicklung spezialisierter Modelle. Sie entscheiden sich für einen Foundation-Model-Ansatz: Sie lizenzieren ein großes vortrainiertes KI-Modell, das auf medizinischen Texten, Bildern und strukturierten Daten trainiert wurde. Dieses Basismodell enthält bereits breites medizinisches Wissen und Verständnis, wurde aber noch nicht für spezifische Aufgaben optimiert. Auf diesem Foundation Model bauen sie verschiedene spezialisierte Anwendungen auf: Für ihre Radiologie-Software passen sie einen Zweig des Modells an, um Lungenknoten in CT-Scans zu erkennen. Für ihre elektronische Patientenakte entwickeln sie eine Komponente zur automatischen Kodierung medizinischer Dokumente nach ICD-Schlüsseln.

Für ihre Patienten-App implementieren sie eine Funktion zur Vereinfachung medizinischer Fachsprache. Für ihre Forschungsabteilung nutzen sie das Modell zur Literaturanalyse und Hypothesengenerierung. Der entscheidende Vorteil: Jede dieser Anwendungen erfordert nur einen Bruchteil der Trainingsbeispiele, die für ein von Grund auf neu entwickeltes Modell nötig wären, da sie auf dem fundierten Grundwissen des Foundation Models aufbauen. Als das Unternehmen später eine neue Anwendung für die Analyse von EKG-Daten entwickelt, profitiert auch diese vom Transfer Learning aus dem Foundation Model, obwohl dieses nie explizit für EKG-Analyse trainiert wurde.



Halluzination (Konfabulation)

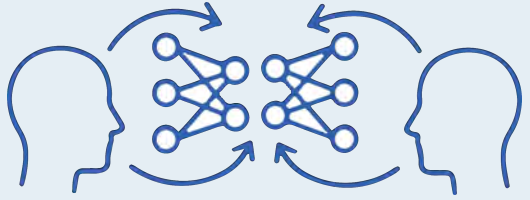
Chatbot gibt ausgedachte Informationen aus.

Halluzination (auch Konfabulation) bezeichnet im KI-Kontext das Phänomen, dass ein Modell scheinbar plausible, aber faktisch falsche oder frei erfundene Inhalte erzeugt. Diese entstehen, wenn die KI Lücken im Wissen durch statistische Muster füllt, ohne zwischen Wahrheit und Fiktion unterscheiden zu können.

Beispiel

Ein Journalist nutzt ein KI-System zur Recherche für einen Artikel über eine historische Persönlichkeit. Als er nach Details zum Privatleben der Person fragt, gibt das System eine detaillierte Antwort mit Anekdoten über die Kindheit, Lieblingsbücher und tägliche Gewohnheiten. Der Journalist ist beeindruckt von der Fülle an Informationen und beginnt zu schreiben, überprüft aber glücklicherweise einige Details in traditionellen Quellen. Dabei stellt er fest, dass viele der "Fakten" nicht verifizierbar sind oder sogar im Widerspruch zu gesicherten historischen Dokumenten stehen.

Das KI-System hat halluziniert – es hat plausibel klingende, aber erfundene Informationen generiert, um Wissenslücken zu füllen. Bei der Analyse des Vorfalles identifizieren die Entwickler die Ursache: Das System wurde hauptsächlich mit kreativen Texten und allgemeiner Literatur über diese historische Periode trainiert, aber hatte nur begrenzten Zugang zu spezifischen, faktischen Informationen über die betreffende Person. Als Lösung wird ein "Unsicherheitsindikator" implementiert, der anzeigt, wenn das System auf Bereiche mit geringer Datenbasis antwortet.



Human-in-the-Loop (HITL)

Ein Mensch greift in Echtzeit ein, um Fehler der KI sofort zu korrigieren.

Verfahren in der Künstlichen Intelligenz, bei dem Menschen gezielt in Trainings-, Entscheidungs- oder Kontrollprozesse eingebunden werden, um die Leistung, Sicherheit und Vertrauenswürdigkeit der Modelle zu verbessern.

Beispiel

In der Radiologie unterstützt ein KI-System die Auswertung von MRT-Bildern, indem es verdächtige Gewebeveränderungen automatisch markiert. Bevor eine Diagnose gestellt wird, prüft jedoch ein Radiologe die Vorschläge der KI, bestätigt richtige Befunde und korrigiert falsche Markierungen.

Diese Rückmeldungen fließen zurück ins System und verbessern dessen Erkennungsleistung kontinuierlich. So kombiniert HITL die Effizienz der KI mit der Erfahrung des Menschen und stellt sicher, dass kritische Fehlentscheidungen vermieden werden.

Informationsextraktion

KI filtert aus großen Textmengen automatisch die wichtigsten Fakten und Begriffe heraus.

Im KI-Bereich bezeichnet Informationsextraktion (IE) den automatisierten Prozess, aus unstrukturierten Datenquellen – meist Texten – gezielt strukturierte Informationen (z. B. Namen, Orte, Ereignisse, Relationen) zu identifizieren und in eine maschinenlesbare Form zu überführen.

Beispiel

Ein Finanzinstitut implementiert ein Informationsextraktionssystem zur Analyse von Quartalergebnissen und Wirtschaftsnachrichten. Täglich werden tausende Dokumente – Gewinnmitteilungen, Analystenberichte, Pressemeldungen und Wirtschaftsartikel – automatisch verarbeitet.

Das System arbeitet in mehreren Stufen:

Named Entity Recognition identifiziert Unternehmen, Personen, Finanzmetriken und Zeiträume in den Texten. *Relation Extraction* erkennt Beziehungen zwischen diesen Entitäten, wie „Company X reported revenue of Y for quarter Z“.

Event Extraction identifiziert bedeutsame Ereignisse wie Führungswechsel, Produkteinführungen oder Übernahmen.

Sentiment Analysis bewertet die emotionale Tonalität in Bezug auf Unternehmen oder Märkte.

Coreference Resolution verbindet Pronomen mit den entsprechenden Entitäten, um vollständige Informationsketten zu bilden.

Die extrahierten strukturierten Informationen fließen in eine Datenbank, die Analysten und Handelsalgorithmen nutzen. Während ein Mensch Stunden bräuchte, um diese verstreuten Informationen zusammenzutragen, liefert das System sie in Sekunden – ein entscheidender Zeitvorteil im schnelllebigen Finanzsektor.

Large Language Models (LLM)

KI-Systeme, die mit sehr vielen Texten trainiert wurden und dadurch Sprache verstehen, zusammenfassen oder selbst Texte verfassen können.

KI-Modell, das auf riesigen Mengen natürlicher Sprachdaten trainiert wurde und mithilfe komplexer neuronaler Netzwerke in der Lage ist, Texte zu verstehen, zu verarbeiten und eigenständig zu generieren.

Beispiel

In einem Krankenhaus wird ein Large Language Model eingesetzt, um Ärzt*innen bei der Dokumentation zu entlasten. Während des Patientengesprächs „hört“ das System mit, wandelt die Sprache automatisch in Text um, strukturiert die Befunde und erstellt einen ersten Entwurf für den Arztbrief. Ärzt*innen müssen diesen nur noch prüfen und anpassen. So sparen sie Zeit, können sich stärker auf die Patientenversorgung konzentrieren und profitieren gleichzeitig von einer konsistenten Dokumentation.

Latente Variable

Nicht sichtbarer Faktor, der Daten beeinflusst.



Eine nicht direkt beobachtbare Größe, die in wissenschaftlichen Modellen oder statistischen Analysen angenommen wird, um verborgene Strukturen, Eigenschaften oder Zusammenhänge zu erklären und über beobachtbare Indikatoren erschlossen wird.

Beispiel

Ein Musikstreaming-Dienst entwickelt ein KI-System zur Analyse von Hörgewohnheiten. Das System entdeckt latente Variablen, die die Musikpräferenzen der Nutzer*innen beeinflussen, aber nicht direkt in den Daten erscheinen. Beispielsweise identifiziert es eine latente Variable, die stark mit der Wahrscheinlichkeit korreliert, dass jemand sowohl klassische Musik als auch bestimmte elektronische Musikrichtungen hört.

Diese Variable könnte als „Präferenz für komplexe harmonische Strukturen“ interpretiert werden – ein Faktor, der weder in den Metadaten der Songs noch in den Nutzerprofileinstellungen explizit erfasst wird. Durch die Erkennung solcher latenten Variablen kann der Dienst überraschend präzise Empfehlungen geben, etwa einem klassischen Musikliebhaber elektronische Musik vorzuschlagen, die ähnliche harmonische Komplexität aufweist – eine Verbindung, die traditionelle kategoriebasierte Empfehlungssysteme übersehen würden.

Maschinelle Übersetzung

Digitaler Dolmetscher, der blitzschnell Texte in andere Sprachen überträgt

Maschinelle Übersetzung bezeichnet den computergestützten Prozess, Texte oder gesprochene Sprache automatisch von einer Ausgangssprache in eine Zielsprache zu übertragen. Moderne Systeme – oft auf Basis von neuronalen Netzen und Transformer-Modellen – analysieren dabei den gesamten Kontext, um nicht nur Wort-für-Wort, sondern auch sinn- und stilgetreu zu übersetzen.

Beispiel

Ein internationales Hilfswerk veröffentlicht Ratgeber für Hygiene und Gesundheitsvorsorge in Krisengebieten. Früher mussten sie jeden Ratgeber mühsam in 20 Sprachen übersetzen lassen, was Monate dauerte und teuer war. Heute verwenden sie ein maschinelles Übersetzungssystem, das zunächst eine Rohübersetzung in alle benötigten Sprachen erstellt. Diese wird dann von muttersprachlichen Helfern nur noch überprüft und angepasst, was den Prozess erheblich beschleunigt. Als eine plötzliche Cholera-Epidemie ausbricht, können sie innerhalb von Tagen statt Monaten aktuelle Informationen in den relevanten Lokalsprachen bereitstellen.

Model Collapse

Phänomen, dass KI-Modelle durch zu viel Training mit von KI erzeugten Daten ihre Vielfalt und Genauigkeit verlieren



Phänomen, bei dem KI-Modelle durch wiederholtes Training auf von KI erzeugten Daten ihre Vielfalt, Genauigkeit und Ausdrucksfähigkeit verlieren und zunehmend fehlerhafte, einseitige oder unbrauchbare Ergebnisse liefern.

Beispiel

Ein Unternehmen trainiert ein neues Sprachmodell nicht mehr mit echten Texten aus Büchern, Zeitungen und wissenschaftlichen Artikeln, sondern überwiegend mit bereits von KI erzeugten Texten. Nach mehreren Trainingszyklen werden die Antworten des Modells zunehmend oberflächlich, wiederholen Phrasen und enthalten häufiger Fehler – weil die KI nur noch ihre eigene, verzerrte „Wirklichkeit“ verarbeitet. Dieses schleichende Verarmen an Qualität und Vielfalt ist ein typischer Fall von Model Collapse..

Natural Language Processing (NLP)

Computer lernen, menschliche Sprache zu verstehen, zu verarbeiten und selbst zu erzeugen.

Forschungs- und Anwendungsfeld der Künstlichen Intelligenz, das Methoden entwickelt, um natürliche Sprache in gesprochener oder geschriebener Form durch Computer zu analysieren, zu verstehen, zu verarbeiten und zu erzeugen.

Beispiel

Ein Kundendienst eines Telekommunikationsunternehmens setzt NLP ein, um die tausenden täglichen Kundenanfragen zu analysieren. Das System kategorisiert automatisch Kundennachrichten nach Themen (Rechnung, technische Probleme, Vertragsfragen), identifiziert die Dringlichkeit und erkennt die Stimmung des Kunden.

Es versteht, dass „Mein Internet ist seit gestern tot“ und „Keine Verbindung möglich seit 24h“ das gleiche Problem beschreiben, obwohl unterschiedliche Worte verwendet werden. Bei der Aussage „Ich warte jetzt schon zum dritten Mal eine Stunde in der Hotline“ erkennt es die Frustration und stuft das Ticket als hochprioritär ein. Durch diese intelligente Vorsortierung werden Anfragen schneller an die richtigen Stellen geleitet.



Personalized Medicine

Behandlung wird exakt auf die einzelne Person zugeschnitten.

Bezeichnet den Einsatz umfangreicher Gesundheitsdaten (z. B. aus Genomik, Bildgebung, Vitalparametern) und moderner KI-Methoden, um individuelle Krankheitsmuster zu erkennen, präzise Diagnosen zu stellen und maßgeschneiderte Behandlungsstrategien für einzelne Patient:innen zu entwickeln.

Beispiel

Eine Onkologieklinik setzt KI für personalisierte Krebsbehandlungen ein. Für eine Patientin mit Brustkrebs analysiert das System ihre Genomsequenzierung, Krankengeschichte, bisheriges Ansprechen auf Behandlungen, Lifestyle-Faktoren und vergleicht diese Daten mit Millionen anonymisierter Patientenakten.

Es identifiziert, dass Patienten mit einem ähnlichen Profil besonders gut auf eine spezifische Kombinationstherapie ansprechen, die bei dieser Unterart des Tumors normalerweise nicht die Standardbehandlung wäre. Der behandelnde Arzt berücksichtigt diese Empfehlung und passt den Therapieplan entsprechend an. Die Patientin spricht hervorragend auf die Behandlung an und erlebt weniger Nebenwirkungen als bei der Standardtherapie.



Prompt

Konkrete Frage oder Aufgabe, die man einer KI stellt.

Anweisung, Frage oder Eingabe, die einer KI gestellt wird, um eine bestimmte Reaktion oder Ausgabe zu erhalten.

Beispiel

Ein Prompt für die Generierung eines Gedichtes mit ChatGPT könnte zum Beispiel so lauten: „Schreibe ein Gedicht in freier Versform über den Fluss der Zeit. Verwende eine melancholische, zugleich hoffnungsvolle Stimmung. Nutze Naturmetaphern wie Wasser, Bäume und Wind, um Vergänglichkeit und Neubeginn zu beschreiben. Das Gedicht soll vier Strophen mit jeweils fünf Zeilen haben und in einer modernen, bildhaften Sprache verfasst sein.“

Prompt Engineering

Den Texteingabebefehl an eine KI so formulieren, dass sie genau das tut, was man will.

Prompt Engineering bezeichnet die gezielte Gestaltung und Optimierung von Eingabeanweisungen (Prompts), um Sprachmodelle und KI-Systeme zu einer möglichst präzisen, nützlichen und kontextgerechten Ausgabe zu steuern.

Beispiel

Ein Lokalreporter nutzt ein KI-Sprachmodell, um bei der Recherche zu unterstützen. Bei seiner anfänglichen Anfrage „Gib mir Informationen über den neuen Stadtpark“ erhält er allgemeine, oberflächliche Aussagen.

Durch Prompt Engineering verbessert er seine Anfrage:

„Analysiere den geplanten Stadtpark aus verschiedenen Perspektiven:

1. *Umweltauswirkungen (Biodiversität, Wasserverbrauch),*
2. *Kosten im Vergleich zu ähnlichen Projekten,*
3. *Zugänglichkeit für verschiedene Bevölkerungsgruppen,*
4. *Historischer Kontext des Standorts.*
5. *Berücksichtige sowohl potenzielle Vorteile als auch Kritikpunkte und nenne Aspekte, die in der öffentlichen Diskussion möglicherweise übersehen wurden.“*

Mit diesem strukturierten, durchdachten Prompt erhält er eine deutlich tiefgründigere Analyse, die mehrere Blickwinkel berücksichtigt und ihm hilft, kritische Fragen für seine Interviews mit Stadtplanenden zu formulieren.

Responsible AI

Sichert, dass KI keinen Schaden anrichtet und gerecht bleibt.

Der Ansatz, KI-Systeme so zu entwickeln und einzusetzen, dass sie ethisch, fair und transparent sind – wie Regeln für verantwortungsvolles Handeln in einer Gesellschaft.

Beispiel

Eine Versicherungsgesellschaft implementiert ein KI-System zur Bewertung von Kfz-Versicherungsanträgen. Im Sinne verantwortungsvoller KI-Nutzung etablieren sie einen umfassenden Governance-Prozess: Vor der Einführung wird das System auf Fairness gegenüber verschiedenen demografischen Gruppen getestet.

Transparenzberichte dokumentieren, welche Faktoren in Entscheidungen einfließen (und welche bewusst ausgeschlossen werden, wie z.B. Postleitzahlen als Proxy für ethnische Zugehörigkeit).

Ein menschliches Überprüfungssystem ermöglicht es Kund*innen, algorithmische Entscheidungen anzufechten. Ein diverses Ethikkomitee überwacht kontinuierlich die Systemleistung. Als eine Analyse zeigt, dass Fahrer*innen bestimmter importierter Fahrzeugmarken systematisch höher eingestuft werden, obwohl ihre Unfallstatistiken dies nicht rechtfertigen, wird das Modell entsprechend angepasst.

Responsible AI Development

„Baue KI verantwortungsbewusst“

Die Entwicklung von KI unter Berücksichtigung ethischer und gesellschaftlicher Auswirkungen – wie ein Architektenteam, das nicht nur die Stabilität eines Gebäudes, sondern auch dessen Umweltauswirkungen und gesellschaftlichen Nutzen berücksichtigt.

Beispiel

Ein Finanzinstitut entwickelt eine KI zur Automatisierung von Kreditentscheidungen und implementiert Responsible AI Development von Beginn an. Statt sich nur auf Effizienz und Genauigkeit zu konzentrieren, integrieren sie ethische Überlegungen in jede Phase der Entwicklung: In der Planungsphase führen sie eine Stakeholder-Analyse durch, die nicht nur interne Abteilungen, sondern auch verschiedene Kundengruppen und Finanzaufsichtsbehörden einbezieht.

Für die Datensammlung implementieren sie strikte Prüfprozesse, um historische Verzerrungen zu identifizieren und zu korrigieren, wie überproportionale Ablehnungen bei bestimmten demografischen Gruppen. Das Modelldesign umfasst explizite formale Bedingungen in der KI, die sicherstellen, dass ähnlich qualifizierte Antragsteller aus verschiedenen Gruppen ähnliche Ergebnisse erhalten. Sie implementieren umfassende Tools, die für jeden Antrag nachvollziehbare Begründungen liefern – sowohl für Kunden als auch für Aufsichtsbehörden. Ein kontinuierliches Monitoring-System überprüft laufend die Modelleleistung auf unbeabsichtigte Verzerrungen oder Diskriminierung.

Bevor das System in die Produktion geht, durchläuft es ein unabhängiges Fairness-Audit durch externe Experten, und ein Ethikkomitee überprüft die gesamte Implementation. Die Ergebnisse dieser verantwortungsvollen Entwicklung sind vielfältig: Die Ablehnungsraten für verschiedene demografische Gruppen sind bei vergleichbaren Kreditwürdigkeitsprofilen nahezu identisch. Die transparenten Erklärungen führen zu weniger Beschwerden und höherer Kundenzufriedenheit. Aufsichtsbehörden loben das System als Modellbeispiel und nutzen es als Referenz für neue Regulierungen. Das Institut stellt fest, dass verantwortungsvolle KI-Entwicklung nicht nur ethisch richtig ist, sondern auch geschäftliche Vorteile bringt: erhöhtes Kundenvertrauen, reduzierte regulatorische Risiken und bessere Vorhersagegenauigkeit durch ausgewogenere Trainingsdaten.

Small Language Models

„Kleiner Bruder“ von GPT-3.

Kompakte Sprachmodelle mit deutlich weniger Parametern als Large Language Models (LLMs), die speziell dafür entwickelt werden, ressourcenschonend, schneller und oft auch datenschutzfreundlicher eingesetzt zu werden – etwa auf mobilen Endgeräten, in Edge-Computing-Umgebungen oder für spezialisierte Anwendungsfälle.

Beispiel

Ein Hersteller von Smart-Home-Geräten implementiert ein Small Language Model für die Sprachsteuerung seiner Produkte. Statt eines riesigen, allgemeinen Sprachmodells, das Gigabyte an Speicher benötigen und ständige Cloud-Verbindungen erfordern würde, entwickeln sie ein spezialisiertes, kompaktes Modell: Das SLM wird mit nur 15 Millionen Parametern gestaltet (im Vergleich zu Milliarden bei großen Modellen) und ist auf den spezifischen Anwendungsbereich optimiert. Statt allgemeinem Weltwissen konzentriert es sich auf Haushaltsgeräte, Steuerungsbefehle und übliche Nutzeranfragen in diesem Kontext.

Das Modell passt in den Flash-Speicher der Smart-Home-Geräte und kann vollständig offline auf einem stromsparenden Mikrocontroller laufen. Es versteht kontextspezifische Befehle wie "mach es wärmer" oder "zu hell hier" und kann Anfragen wie "Ist die Waschmaschine fertig?" oder "Wann habe ich den Backofen eingeschaltet?" beantworten. Während es nicht für philosophische Diskussionen oder kreatives Schreiben geeignet ist, meistert es seine spezialisierte Aufgabe mit 97% Genauigkeit.

Die Vorteile gegenüber cloud-basierten Großmodellen sind: keine Internetverbindung erforderlich, was Zuverlässigkeit und Resilienz verbessert. Latenz unter 100ms statt mehreren Sekunden für Cloud-Anfragen. Erhöhte Privatsphäre, da keine Sprachdaten das Haus verlassen. Deutlich reduzierter Energieverbrauch, was besonders für batteriebetriebene Geräte wichtig ist. Die kompakten Modelle werden monatlich aktualisiert und über Firmware-Updates verbessert, während sie weiterhin vollständig lokal ausgeführt werden.



Synthetic Data

„Ausgedachte“ Daten, die echten sehr ähneln

Künstlich generierte Daten, die mit Hilfe von Algorithmen, Simulationen oder KI-Modellen erzeugt werden und echte Daten nachahmen, um Analysen, Modelltraining oder Tests durchzuführen, ohne auf sensible oder schwer verfügbare Originaldaten zurückgreifen zu müssen.

Beispiel

Ein Finanzinstitut möchte ein KI-System entwickeln, um Kreditkartenbetrug zu erkennen, steht aber vor Datenschutzherausforderungen bei der Nutzung echter Kundendaten. Die Lösung: Synthetische Daten. Basierend auf aggregierten statistischen Eigenschaften echter Transaktionsdaten (ohne individuelle Kundeninformationen) generiert ein KI-Modell Millionen synthetischer Transaktionen, die die gleichen statistischen Muster, Ausreißer und Betrugsindikatoren aufweisen wie echte Daten. Diese synthetischen Daten können problemlos mit externen Entwicklern geteilt werden, da sie keine realen Personen repräsentieren. Das auf diesen Daten trainierte Betrugserkennungssystem erreicht eine Genauigkeit, die nahezu identisch ist mit einem System, das auf echten Daten trainiert wurde – ohne dass sensible Kundeninformationen exponiert werden mussten.

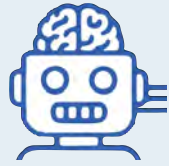
System Prompt

„Heimliche Regeln“, nach denen die KI denkt und reagiert.

Grundlegende Anweisungen, die das Verhalten eines KI-Systems definieren – wie die Grundregeln oder das Betriebshandbuch für einen digitalen Assistenten. „Sei höflich, hilfsbereit und kurz“ als Vorgabe für ChatGPT.

Beispiel

Ein Krankenhaus implementiert einen KI-Assistenten für die Aufnahme neuer Patienten. Der System Prompt definiert grundlegende Verhaltensweisen und Grenzen: "Du bist ein freundlicher, geduldiger Assistent für die Patientenaufnahme. Kommuniziere in einfacher, verständlicher Sprache. Stelle Fragen zu Personalien, Versicherungsstatus, aktuellen Symptomen und relevanter Vorgeschichte. Biete älteren Patienten an, langsamer zu sprechen. Verweise bei Anzeichen akuter medizinischer Notfälle sofort an medizinisches Personal. Gib niemals medizinische Diagnosen oder Behandlungsempfehlungen. Halte alle Patienteninformationen streng vertraulich. Wenn du etwas nicht verstehst oder die Antwort nicht weißt, sage das offen und biete an, einen menschlichen Mitarbeiter hinzuzuziehen." Dieser System Prompt stellt sicher, dass der Assistent in allen Gesprächen konsistent, sicher und angemessen agiert, unabhängig von den spezifischen Patientenfragen.



Technologische Singularität

Der Moment, in dem Maschinen schlauer sind als alle Menschen

Ein hypothetischer Punkt, an dem KI die menschliche Intelligenz übertrifft und eine explosionsartige Selbstverbesserung einsetzt – ein Wendepunkt, nach dem die technologische Entwicklung nicht mehr vorhersehbar ist.

Beispiel

Stellen Sie sich ein hochentwickeltes KI-System in der medizinischen Forschung vor, das selbstständig neue Hypothesen über Krankheiten entwickelt, Experimente simuliert, die Ergebnisse interpretiert und daraus wiederum neue, verbesserte KI-Modelle trainiert. Wenn dieser Zyklus ohne menschliche Kontrolle immer schneller und effizienter wird, könnte ein Punkt erreicht werden, an dem die KI ihre eigenen Fähigkeiten exponentiell steigert.

Dieser Moment wird in der Theorie als technologische Singularität beschrieben – der Übergang, bei dem menschliche Eingriffe nicht mehr mithalten können und sich die Innovationsgeschwindigkeit der KI derart verselbständigt, dass eine qualitativ neue Form von Intelligenz entsteht.

Token

Ein "Texthäppchen".

Kleinste Texteinheit (Wort, Silbe oder Zeichen), die ein Sprachmodell verarbeitet. Wenn Sie "Künstliche Intelligenz" in ChatGPT eingeben, zerlegt das System diese Eingabe in drei Tokens: "Künst", "liche" und "Intelligenz" - wobei jedes Token eine Grundeinheit darstellt, die das Modell einzeln verarbeitet und versteht.

Beispiel

Ein Übersetzungsdienst nutzt ein KI-Modell mit begrenztem Kontext von 4.096 Tokens. Ein Kunde möchte ein 15-seitiges juristisches Dokument vom Deutschen ins Englische übersetzen lassen. Da das Dokument etwa 12.000 Tokens enthält (weit mehr als die Kapazität des Modells), zerlegt das System das Dokument automatisch in vier Segmente von je etwa 3.000 Tokens. Bei der Zerlegung wird darauf geachtet, dass keine Sätze unterbrochen werden. Jedes Segment wird einzeln übersetzt, wobei einige überlappende Sätze als Kontext dienen, um Konsistenz zu gewährleisten. Der Kunde bemerkt, dass komplexe juristische Fachbegriffe in den späteren Segmenten besser übersetzt werden, da das Modell aus den früheren Segmenten "gelernt" hat, wie diese Begriffe im Kontext des Dokuments verwendet werden.

Tokenisierung

Text wird in Häppchen geschnitten, die die KI versteht.

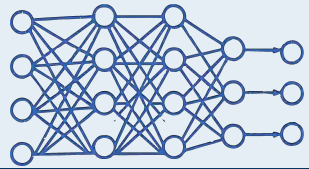


Der Prozess, bei dem Text in kleinere Einheiten (Tokens) zerlegt wird, die eine KI verarbeiten kann – wie das Zerschneiden eines langen Satzes in verdauliche Häppchen.

Beispiel

Ein Spracherkennungssystem für ein Call-Center verarbeitet täglich Tausende Kundenanrufe. Bevor die KI die gesprochenen Worte verstehen kann, muss sie den kontinuierlichen Audiostream in einzelne Tokens zerlegen. Die Tokenisierung erfolgt auf mehreren Ebenen: Zunächst werden Sprechpausen erkannt, um einzelne Sätze zu identifizieren. Dann werden diese Sätze in Wörter und Teilwörter zerlegt. Deutsche zusammengesetzte Substantive wie "Krankenversicherungskarte" werden in kleinere Einheiten wie "Kranken", "versicherungs" und "karte" aufgeteilt.

Diese Tokenisierung ermöglicht es dem System, auch ungewöhnliche oder neue Wortkombinationen zu verstehen, indem es sie auf bekannte Grundbausteine zurückführt. Beim Training hat das System gelernt, dass das Token "versicherungs" oft in Verbindung mit Begriffen wie "police", "beitrag" oder "karte" vorkommt, was dem KI-Modell hilft, den Kontext eines Gesprächs besser zu erfassen.



Transformer-Modelle

KI-Architektur, die Sprache durch parallele Aufmerksamkeitsmechanismen analysiert und so Bedeutungsbeziehungen zwischen Wörtern im gesamten Kontext erfasst.

Klasse neuronaler Netze, die auf sogenannten Self-Attention-Mechanismen beruhen und es ermöglichen, Sprach- oder andere Sequenzdaten parallel statt schrittweise zu verarbeiten, wodurch sie besonders leistungsfähig für Aufgaben wie maschinelle Übersetzung, Textgenerierung oder Bildverarbeitung sind

Beispiel

Eine wissenschaftliche Datenbank implementiert ein Transformer-Modell für die semantische Suche in Millionen von Forschungsartikeln. Als eine Biologin nach "Proteinrezeptoren, die mit Autoimmunerkrankungen assoziiert sind" sucht, versteht das Modell den komplexen Zusammenhang zwischen den Begriffen. Es erkennt, dass Artikel relevant sein können, auch wenn sie den exakten Suchbegriff nicht enthalten – etwa Dokumente, die über "T-Zell-Aktivierung bei rheumatoider Arthritis" sprechen. Durch die Aufmerksamkeitsmechanismen der Transformer-Architektur kann das Modell weitreichende Beziehungen innerhalb der wissenschaftlichen Literatur erkennen und auch feststellen, welche Teile eines Dokuments für die Anfrage besonders relevant sind. Die Biologin findet so einen entscheidenden Artikel, der mit traditionellen Suchbegriffen übersehen worden wäre, was ihre Forschung in eine neue, vielversprechende Richtung lenkt.

Word Embeddings

Wörter werden in Zahlenreihen umgewandelt.

Vektordarstellungen von Wörtern, die deren Bedeutung in einem mehrdimensionalen Raum abbilden und dadurch semantische Ähnlichkeiten zwischen Wörtern erfassbar machen. Word Embeddings wandeln Wörter in mathematische Zahlenreihen um, sodass ähnliche Begriffe wie "König" und "Königin" im digitalen Raum nahe beieinander liegen und Computer Wortbedeutungen verstehen können. Wie eine Karte, auf der ähnliche Wörter nahe beieinander platziert sind.

Beispiel

Ein großes Universitätsklinikum implementiert domänenspezifische Word Embeddings, um die Zuordnung zwischen Patientendaten und klinischen Leitlinien zu verbessern. Die Herausforderung: Ärzt:innen, Pflegekräfte und Patient:innen verwenden oft unterschiedliche Begriffe für dieselben Symptome oder Krankheitsbilder. Traditionelle keyword-basierte Such- und Klassifikationssysteme stoßen dadurch an ihre Grenzen. Das Klinikum trainiert Word Embeddings auf Millionen von elektronischen Patientenakten, klinischen Studienberichten und medizinischen Fachartikeln. Jedes Wort wird dabei als hochdimensionaler Vektor dargestellt, sodass semantisch verwandte Begriffe in räumlicher Nähe liegen: „Herzinfarkt“ und „Myokardinfarkt“ werden als synonym erkannt, „Hypertonie“ und „Bluthochdruck“ werden als gleiche Diagnose verstanden, „Müdigkeit“ und „Fatigue“ werden als symptomatisch verwandt eingeordnet.

Die Embeddings ermöglichen mehrere Anwendungen: Bei der Diagnosesuche werden automatisch semantisch ähnliche Begriffe einbezogen, sodass ein Eintrag „Herzattacke“ auch passende Dokumentationen zu „Myokardinfarkt“ anzeigt. In der Therapieempfehlung werden Patientendaten mit Leitlinien gematcht – nicht nur über exakte Fachtermini, sondern auch über semantische Ähnlichkeit. Ein „Skill Gap Analyzer“ für Weiterbildungen zeigt Pflegekräften und Ärzt:innen auf, welche Zusatzqualifikationen (z. B. in Onkologie oder Radiologie) ihnen fehlen, um bestimmte klinische Aufgaben zu übernehmen, und empfiehlt passende Schulungen.

Die Implementierung führt zu messbaren Verbesserungen: Die Zeit zur richtigen Diagnosefindung sinkt, da Systeme auch umgangssprachliche Patientenschilderungen verstehen. Die Adhärenz zu Leitlinien steigt, da Empfehlungen besser mit den individuellen Patientendaten abgeglichen werden. Ärzt:innen und Pflegekräfte berichten von einer höheren Zufriedenheit, da sie relevantere Informationen schneller finden und semantische Unterschiede zwischen Fachsprache und Alltagssprache überbrückt werden.



Bild: Yutong Liu & Kingston School of Art / betterimagesofai.org

Dieses Heft ist von:
Landesarbeitsgemeinschaft Selbsthilfe von
Menschen mit Behinderungen und chronischen
Erkrankungen Rheinland- Pfalz e.V.



Mit freundlicher Unterstützung der AOK
Rheinland-Pfalz/Saarland



Die Inhalte sind von: © Damian Paderta | nozilla.de