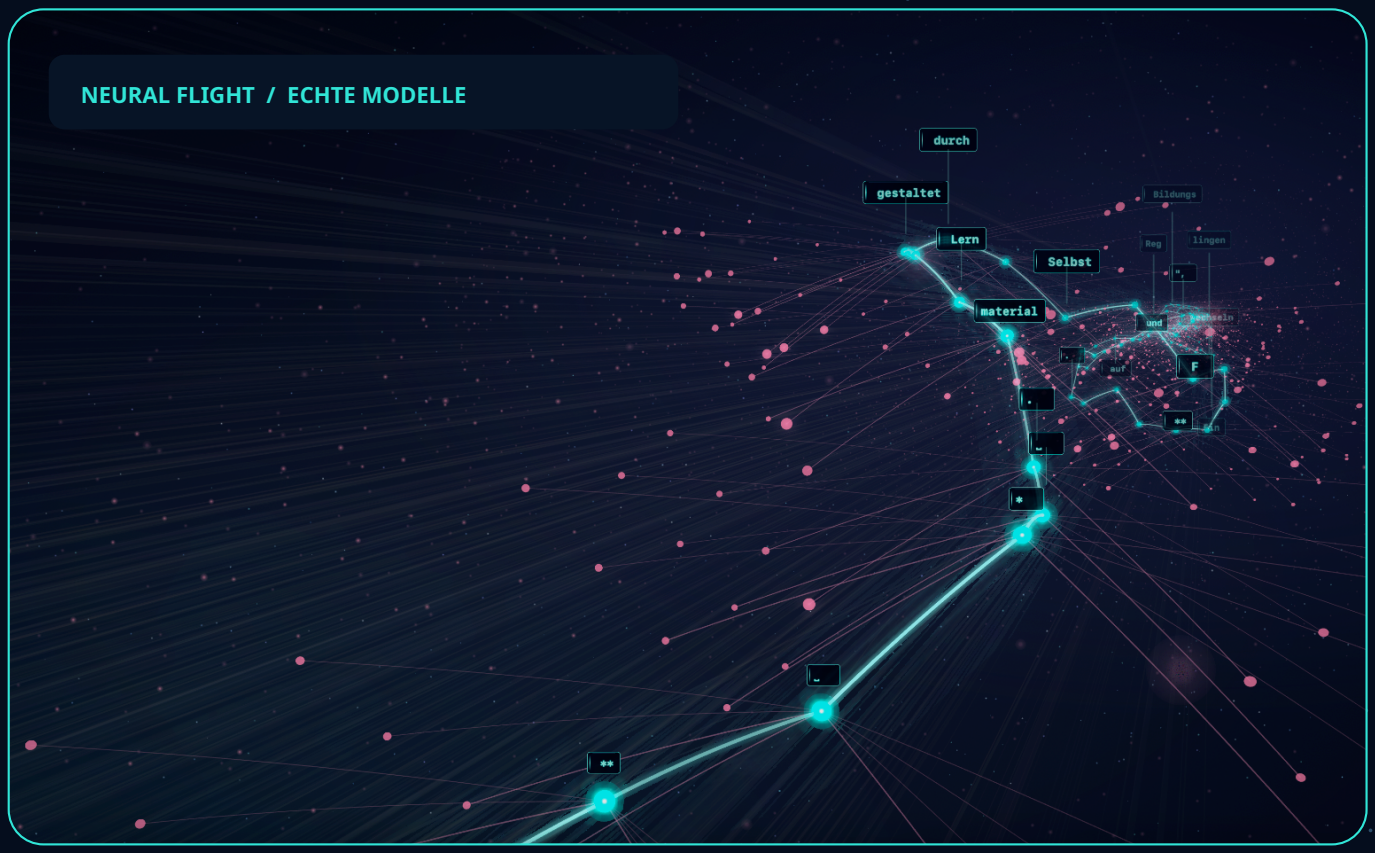
KITA DIGITAL
AKADEMIE

DEIN KOSTENLOSES KI-NACHSCHLAGEWERK

KI VERSTEHEN

DAS KLEINE WÖRTERBUCH

Die wichtigsten Begriffe rund um künstliche Intelligenz
und Sprachmodelle - einfach erklärt.



KI verstehen statt mystifizieren.

34 verständliche Begriffe · 6 Lernmodule · ohne Vorwissen

Ein Begleitmaterial der KITA DIGITAL Akademie
ki-insel.de

NEURAL FLIGHT · AUSGABE 1.0

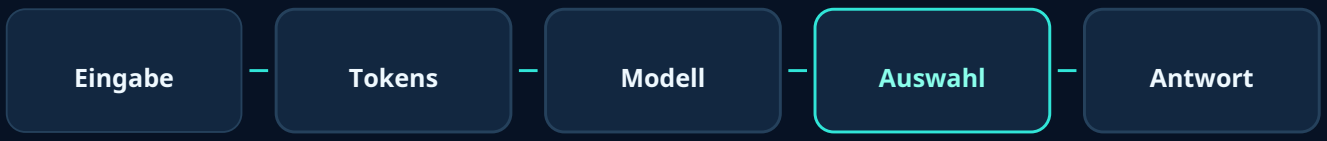
WILLKOMMEN

Weniger Mythos.
Mehr Verständnis.

Künstliche Intelligenz begegnet uns im Alltag. Aber warum erzeugt ein Sprachmodell aus einer Frage einen verständlichen Text? Was sind Tokens, Modellschichten und Wahrscheinlichkeiten? Dieses Glossar hilft dir dabei, die wichtigsten Begriffe jederzeit nachzuschlagen.

Die Erklärungen sind absichtlich einfach. Manche komplexen Zusammenhänge werden verkürzt. Die Begriffe beziehen sich besonders auf generative Sprachmodelle - andere KI-Systeme können anders funktionieren.

SO ENTSTEHT EINE TEXTANTWORT



DREI EBENEN BEI NEURAL FLIGHT

Messung
Tokens, Wahrscheinlichkeiten, Zwischenzustände

Projektion
Mathematische Aufbereitung für den 3D-Raum

Gestaltung
Licht, Nebel, Farben und Kamerafahrten

Du kannst das Glossar von A bis Z lesen oder einzelne Begriffe nachschlagen.

Viel Freude beim Entdecken!

BEGRIFFE A-Z

Das KI-Wörterbuch

3

3D-Visualisierung

DARSTELLUNG

Eine räumliche Darstellung ausgewählter Daten. Neural Flight nutzt sie, um Berechnungen eines Sprachmodells verständlicher zu machen. Sie ist keine wörtliche Karte der Gedanken einer KI.

A

Alternative Route

EXPERIMENT

Eine neu berechnete Fortsetzung nach einer bewusst veränderten Token-Auswahl. Der vorherige Verlauf kann daneben zum Vergleich erhalten bleiben.

Attention

MODELL

Ein Mechanismus, mit dem bestimmte neuronale Netze Informationen aus verschiedenen Positionen im Kontext gewichten und zusammenführen. Attention-Werte sind nicht automatisch eine vollständige Erklärung der Modellentscheidung.

Autoregressive Textgenerierung

GENERIERUNG

Text wird Schritt für Schritt erzeugt. Bei jedem neuen Schritt berücksichtigt das Sprachmodell die bisherige Tokenfolge und bestimmt eine weitere Einheit.

B

Blackbox

EINORDNUNG

Bezeichnung für ein System, dessen innere Abläufe oder einzelne Ergebnisse schwer vollständig nachzuvollziehen sind. Das bedeutet nicht, dass sich gar nichts darin messen lässt.

E

Erzwungenes Token

EXPERIMENT

Ein Token, das bei einem Experiment absichtlich vorgegeben wird, statt es durch das übliche Auswahlverfahren bestimmen zu lassen. Auf dieser geänderten Folge baut eine neue Generierung auf.

G

Generative KI

GRUNDLAGEN

Künstliche Intelligenz, die neue Inhalte erzeugen kann, etwa Texte, Bilder, Musik oder Programmcode.

H

Halluzination

EINORDNUNG

Bezeichnung für falsche, erfundene oder nicht ausreichend gestützte Inhalte eines KI-Systems. Der Begriff ist bildhaft: Er bezeichnet keine menschliche Wahrnehmung der KI.

Hidden State

MODELL

Ein interner, numerischer Zwischenzustand während der Modellverarbeitung. Bei geeigneten Modellen können solche Werte ausgelesen und verglichen werden.

I

Inferenz

GRUNDLAGEN

Die Anwendung eines bereits trainierten Modells auf eine neue Eingabe - beispielsweise die Berechnung einer Antwort auf eine Frage.

K

Kontext

GENERIERUNG

Die Informationen, die das Sprachmodell bei einem Berechnungsschritt berücksichtigt, zum Beispiel die Eingabe und die bisher erzeugten Tokens.

Kosinusähnlichkeit

MESSWERT

Eine Kennzahl für die Ähnlichkeit der Richtungen zweier Vektoren. Sie allein sagt nicht, ob eine KI-Antwort korrekt oder sinnvoll ist.

Künstliche Intelligenz (KI)

GRUNDLAGEN

Gebiet der Informatik mit Verfahren für Aufgaben wie Mustererkennung, Vorhersage, Sprachverarbeitung oder Planung. Nicht jedes KI-System ist ein Chatbot.

L

Layer (Modellschicht)

MODELL

Eine Verarbeitungsschicht eines neuronalen Netzes. In vielen Sprachmodellen verändern aufeinanderfolgende Layer die internen Zahlenmuster.

L

Logit Lens

DIAGNOSTIK

Eine Untersuchungsmethode, die Zwischenzustände näherungsweise in mögliche Token-Vorhersagen übersetzt. Sie zeigt keine tatsächlich getroffene Entscheidung jedes einzelnen Layers; besonders frühe Ergebnisse können irreführend sein.

Logits

GENERIERUNG

Nicht normalisierte Zahlenwerte, die ein Sprachmodell für mögliche nächste Tokens berechnet. Aus ihnen können Wahrscheinlichkeiten abgeleitet werden.

M

Maschinelles Lernen

GRUNDLAGEN

Teilgebiet der KI, in dem Modellparameter mithilfe von Daten und Trainingsverfahren angepasst werden.

Modellparameter

MODELL

Numerische Einstellungen eines Modells, die während des Trainings angepasst werden. Sie beeinflussen die späteren Berechnungen.

N

Neuronales Netz

GRUNDLAGEN

Mathematisches System aus verbundenen Rechenoperationen. Der Name wurde teilweise von biologischen Nervensystemen inspiriert; es besteht jedoch nicht aus lebenden Nervenzellen.

P

PCA (Hauptkomponentenanalyse)

DARSTELLUNG

Ein mathematisches Verfahren, das Daten auf weniger Dimensionen projiziert, damit Muster leichter sichtbar werden. Dabei gehen Informationen verloren.

Prompt

GRUNDLAGEN

Eine Eingabe oder Anweisung an ein KI-System - etwa eine Frage, eine Aufgabe oder ein zu ergänzender Text.

S

Sampling

GENERIERUNG

Zufallsbasierte Auswahl eines Tokens aus einer Wahrscheinlichkeitsverteilung. Wahrscheinlichere Tokens werden meist bevorzugt, aber andere Möglichkeiten bleiben möglich.

Seed

GENERIERUNG

Ausgangswert für einen Zufallszahlengenerator. Er kann Abläufe reproduzierbarer machen, garantiert aber nicht unter allen technischen Bedingungen identische Ergebnisse.

Sprachmodell

GRUNDLAGEN

Ein KI-Modell zur Verarbeitung von Sprache oder Text. Generative Sprachmodelle können Texte beispielsweise fortsetzen und Antworten formulieren.

T

Temperature

GENERIERUNG

Eine Einstellung, die die Verteilung der Token-Wahrscheinlichkeiten bei der Auswahl verändert. Niedrigere Werte konzentrieren stärker auf wahrscheinliche Möglichkeiten; höhere können ungewöhnlichere Fortsetzungen begünstigen.

Token

GENERIERUNG

Eine Einheit, in die Text für die Verarbeitung zerlegt wird. Tokens können ganze Wörter, Wortteile, Satzzeichen oder andere Texteinheiten sein.

Tokenizer

GENERIERUNG

Software, die Text in Tokens zerlegt und zugehörige Kennungen erzeugt. Die genaue Zerlegung ist vom verwendeten Tokenizer abhängig.

Top-K

GENERIERUNG

Filtert die Auswahl auf höchstens K der am höchsten bewerteten Tokens. Bei Top-K = 5 bleiben zum Beispiel höchstens fünf Kandidaten zugelassen.

Top-P

GENERIERUNG

Berücksichtigt die kleinste Gruppe der wahrscheinlichsten Tokens, deren aufsummierte Wahrscheinlichkeit den gewählten Grenzwert erreicht oder überschreitet.

Training

GRUNDLAGEN

Vorgang, bei dem Modellparameter anhand von Trainingsdaten angepasst werden. Training ist nicht dasselbe wie die spätere Generierung einer Antwort.

T

Trainingsdaten

GRUNDLAGEN

Daten, mit denen ein Modell trainiert wird. Deren Auswahl und Qualität beeinflussen auch die Fähigkeiten und Grenzen des Modells.

V

Vektor

MESSWERT

Eine geordnete Folge von Zahlen. Sprachmodelle verwenden Vektoren als Teil ihrer internen Repräsentationen.

Vektornorm

MESSWERT

Eine mathematische Größe eines Vektors, etwa seine Länge nach einer festgelegten Berechnung. Eine größere Norm bedeutet nicht automatisch mehr Sicherheit oder Intelligenz.

W

Wahrscheinlichkeit

GENERIERUNG

Zahl für die modellierte Chance einer Möglichkeit. Bei Sprachmodellen lassen sich Wahrscheinlichkeiten für nächste Tokens aus berechneten Werten ableiten.

ZUM SCHLUSS

KI verstehen heißt, gute Fragen zu stellen.

Eine beeindruckende KI-Antwort muss nicht richtig sein. Eine leuchtende 3D-Grafik muss nicht alle Details der Berechnung zeigen. Aber wir können beides aufmerksam untersuchen, vergleichen und hinterfragen.

Das ist die wichtigste Haltung aus dem Kurs: Neugierig bleiben, Messungen ernst nehmen und die Grenzen unserer Erkenntnisse benennen.

> **Neural Flight macht Berechnungen anschaulich.**
>
> Es zeigt ausgewählte Daten eines Sprachmodells und ihre Darstellung - keine Gedanken, keinen vollständigen dreidimensionalen Gehirnraum und keine Garantie für die Richtigkeit einer Antwort.

ZUM VERTIEFEN

- KITA DIGITAL Akademie
- Hugging Face: Tokenizers
- Hugging Face: Model Outputs
- scikit-learn: PCA
- Belrose et al.: Tuned Lens (Fachpublikation)

KITA DIGITAL Akademie

Künstliche Intelligenz verständlich machen.

ki-insel.de