

# Über Generative Modelle sprechen

Erarbeitet von  
Dr. Jacqueline Klusik-Eckert

|                                             |                                           |
|---------------------------------------------|-------------------------------------------|
| <b>Lernziele</b> .....                      | 1                                         |
| <b>Inhalt</b> .....                         | 2                                         |
| Einstieg .....                              | 2                                         |
| Sprechen über Generative KI .....           | <b>Fehler! Textmarke nicht definiert.</b> |
| Technische Begriffe in der GenAI-Welt ..... | 3                                         |
| Menschelnde Begriffe .....                  | 4                                         |
| Take-Home Message .....                     | 5                                         |
| <b>Quellen</b> .....                        | 6                                         |
| <b>Disclaimer</b> .....                     | 6                                         |

## Lernziele

- Du kannst die wichtigsten technischen Begriffe erklären
- Du kannst die Verwendung von anthropomorphisierenden Begriffen kritisch hinterfragen

## Inhalt

### Einstieg

Na, hast du auch schon mal was mit ChatGPT gemacht oder geschrieben, entworfen oder generiert? Oder hast du schon mal gesagt, das Bild hier hat eine KI ‚gemalt‘?

Parallel zur Weiterentwicklung der Technik und gerade seit der Auskopplung großer Generativer Modelle in Form von einfach zu verwendenden Anwendungen, suchen wir nach der richtigen Sprache, um darüber zu sprechen.

Eine präzise Sprache und Benennung der Modelle und Verfahren sind jedoch notwendig, um über die Technik zu sprechen. Wie sagen wir zu den Ergebnissen von Generativen Modellen? Und wann sagt man generative KI und wann generative Modelle?

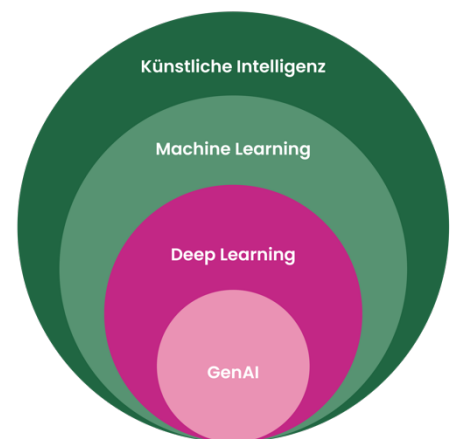
### Sprache und wichtige Begriffe

Im Sprechen über Generative KI ist es nicht immer einfach, genau und präzise zu bleiben. Das kennen wir ja auch, wenn wir generell über Künstliche Intelligenz sprechen. Da ist man auch meistens verallgemeinernd und nennt nicht das konkrete Modell oder Verfahren.

Was ist also nun mit den Begriffen, mit denen wir es zu tun haben, wenn wir über Generative KI sprechen und bei welchen Formulierungen ist Vorsicht geboten?

Generative Künstliche Intelligenz, GenAI, Generative Modelle, ganz gleich welchen Begriff man verwendet, es bleibt ein verallgemeinernder Überbegriff. Dröseln wir das ganze doch mal auf.

Als kleiner Spezialbereich gehört Generative KI zum Forschungsfeld der Künstlichen Intelligenz. Mit diesem Oberbegriff werden damit unterschiedliche Arten von Architekturen und Modellen beschrieben, mit denen neue Inhalte wie Text, Bild, Musik, Audio oder Video erstellt werden, generiert werden. Soweit, soweit. Das ist nichts Neues.



Die GenAI Systeme lassen sich wiederum in verschiedenen Kategorien unterscheiden, die je nach Zielvorstellung sehr unterschiedlich sein können.

Man kann zum Beispiel zwischen unimodalen Modellen und multimodalen Modellen unterscheiden. Also solche, die in ihrer Gattung bleiben, Text-zu-Text oder Bild-zu-Bild oder solche, die auch zwischen den Gattungen wechseln können. Text-zu-Bild, Text-zu-Video, Text-zu-Audio oder auch zurück.

### Quelle [1, 2]

- Unimodale Modelle vs. Multimodale Modelle z. B. Bing-Chat
- Text-zu-Audio z. B. Meta Audio Craft

Ein weiteres Unterscheidungsmerkmal kann sein, wie zugänglich und überprüfbar die Modelle sind, also welcher Grad an Transparenz gegeben ist, also open source oder closed source LLM (GPT4, Midjourney).

„Open“ sind zum Anwendungen, die frei zugänglich sind, Zugriff oder Einblick auf Trainingsdaten sowie Code geben.

„Closed source“ sind Anwendungen, die keinen Zugriff, keine Einsicht auf Trainingsdaten, Codes oder dahinterliegende Modelle und Architektur zulassen. Demnach minimal bis gar nicht transparent sind. Der Übergang kann dabei sehr fließend sein.

Es ist wichtig zu bedenken, dass alle diese Modelle unterschiedliche Eigenschaften haben, teilweise unterschiedliche Trainingsverfahren hinter sich haben, ganz andere Modellarchitekturen verwenden oder für unterschiedliche Zwecke entwickelt wurden. Da die GenAI-Modelle so unterschiedliche sind, ist es wichtig sich bewusst zu machen, von welchem oder von was man eigentlich spricht.

Verwendet man hingegen den Überbegriff „Generative KI“, wirft man alle Ansätze in einen Sack.

## Quelle [2, 3]

### Technische Begriffe in der GenAI-Welt

Wenn wir über generative KI sprechen, ist es wichtig, präzise zu bleiben. Erst das Hilft die spezifischen Konzepte zu verstehen und daran anknüpfend darüber zu debattieren. Lass uns daher einen Blick auf einige dieser Begriffe werfen und dabei auch darauf achten, welche Formulierungen Vorsicht erfordern.

Zunächst einmal müssen wir den Begriff „Auswerten“ verstehen. In Bezug auf Generative KI bezieht sich das auf den Prozess der Analyse von Daten oder Informationen, um relevante Muster oder Informationen zu extrahieren.

## Quelle [3]

Ein wichtiges Konzept in der Welt der generativen KI sind die „Tokens“. Diese sind Teile von Text und dienen als Eingabe für Modelle wie GPT-3.5, GPT-4, BERT jetzt Gemini, und LLAMA 2. Es ist wichtig zu beachten, dass ein Token in der Regel etwa vier Zeichen Text entspricht. Diese Tokens werden dann in einem multi-dimensionalen Vektorraum abgebildet, um ihre Bedeutung zu repräsentieren, ein Prozess, der als „embedding“ bezeichnet wird.

Ein weiteres Konzept ist der „latent space“. Dieser bezieht sich auf den abstrakten Raum, in dem generative Modelle operieren und in dem sie neue Daten erzeugen können, basierend auf den Mustern, die sie während des Trainings gelernt haben.

#### Quelle [4, 5]

Dann ist es ebenso wichtig zwischen Anwendung und Modell zu unterscheiden. Weil es das Bekannteste ist, machen wir das mal bei ChatGPT.

OpenAI ist die Firma.

Chat GPT ist eine von OpenAI entwickelte Anwendung, die uns mittels einer chatähnlichen Benutzeroberfläche, dem Interface, eine menschenähnliche textbasierte Kommunikation mit einem KI-basierten Sprachmodell ermöglicht. GPT3 oder GPT4 ist ein multimodales Large Language Modell, das auf einem generativen, vortrainierten Transformermodell basiert. Auf Englisch dann Generative Pre-trained Transformer. Die Zahl gibt die Version an.

Innhalb eines Interfaces verwenden wir für den Anschlag dieses KI-Systems keine Befehle in einer Programmiersprache, sondern verwenden Prompts.

Ein **Prompt** ist ein Text in natürlicher Sprache, der die Aufgabe beschreibt, die eine KI ausführen soll. Der Prompt ist also der Befehl, der Input für das Modell.

Der wissende Umgang mit Prompts wird dann Prompt **Engineering** genannt. Das ist der Prozess des Entwerfens, Verfeinerns und Optimierens von Prompts, um die Absicht der User\*innen effektiv an ein LLM zu kommunizieren.

#### Quelle [6]

##### Menschelnde Begriffe

Neben diesen technischen Begriffen gibt es auch einige menschliche Begriffe, die oft im Zusammenhang mit generativer KI verwendet werden, wie „halluzinieren“, „plagiierten“, „träumen“, „schöpfen“ und „kreativ sein“. Es ist wichtig zu verstehen, dass diese Begriffe anthropomorphisieren. Sie suggerieren menschliche Fähigkeiten. Fähigkeiten, die diese mathematischen Systeme nicht haben. Wir überschätzen ‚die KI‘ und vermuten etwas wie ein Bewusstsein.

Insgesamt ist es wichtig, sich bewusst zu sein, dass Generative KI-Modelle letztendlich auf Algorithmen und Daten basieren. Sie sind nur für uns nicht sichtbar und werden daher nicht als solche wahrgenommen.

Also: Auch wenn ihre Ergebnisse menschenähnlich erscheinen, ist es entscheidend, den technischen Hintergrund zu verstehen und nicht zu sehr in die Vermenschlichungsfalle zu tappen.

H5P Element mit den Begriffen

**Halluzinieren:**

- Generative KI-Modelle erzeugen Texte oder andere Daten auf der Grundlage von Mustern in den Trainingsdaten und Algorithmen. Sie haben keine bewusste Wahrnehmung oder Fähigkeit zur Halluzination wie ein menschliches Gehirn.
- Der Begriff "halluzinieren" impliziert eine bewusste Wahrnehmung von Dingen, die nicht real sind, was nicht auf generative KI-Modelle zutrifft.

#### **Plagiierten:**

- Generative KI-Modelle kopieren oder übernehmen nicht bewusst die Werke anderer. Sie erstellen neue Inhalte auf der Grundlage der Muster, die sie während des Trainings gelernt haben.
- Plagiierten bezieht sich auf das bewusste unbefugte Kopieren oder Diebstahl von Ideen oder Inhalten anderer Personen, was bei KI-Modellen nicht zutrifft.

#### **Träumen:**

- Träumen bezieht sich auf einen Zustand des Bewusstseins während des Schlafs, in dem das Gehirn verschiedene Gedanken, Bilder und Empfindungen erzeugt.
- Generative KI-Modelle haben keinen Bewusstseinszustand und können daher nicht „träumen“ wie ein menschliches Gehirn.

#### **Schöpfen:**

- Der Begriff „schöpfen“ impliziert eine kreative Aktivität, bei der neue Ideen oder Inhalte entwickelt werden.
- Generative KI-Modelle können zwar neue Inhalte erstellen, basierend auf den Mustern in den Trainingsdaten, aber sie sind nicht in der Lage, bewusst zu schöpfen oder kreativ zu sein.

#### **Kreativ sein:**

- Obwohl Generative KI-Modelle dazu verwendet werden können, neue und einzigartige Inhalte zu erstellen, geschieht dies auf der Grundlage von statistischen Mustern und Algorithmen.
- Sie haben keine Emotionen, Vorstellungen oder bewusste Absicht, daher ist der Begriff „kreativ sein“ bei der Beschreibung ihres Verhaltens irreführend.

#### **Quelle [4]**

##### [Take-Home Message](#)

Bereits zu Beginn der Entwicklungen von Künstlicher Intelligenz haben Forscher\*innen und die Weltöffentlichkeit in die Rechenmaschinen vermenschlichte Eigenschaften hineininterpretiert. Je besser, oder besser gesagt überzeugender, die Ergebnisse der

Generativen Modelle werden, sei es beim Erstellen von Text, Sprache, Bildern oder Musik, ist man nochmal mehr dazu verleitet, es unseren Vorfahren gleichzutun. Das Hilfsmittel dagegen ist, sich zu jedem Zeitpunkt klarzumachen, was im Hintergrund einer Eingabemaske geschieht.

Weiterführendes Material

<https://huggingface.co/>

## Quellen

- Quelle [1] Bommasani, R., et.al. (2021). On the Opportunities and Risks of Foundation Models. <https://doi.org/10.48550/ARXIV.2108.07258>
- Quelle [2] Bommasani, R., Klyman, K., Longpre, S., Kapoor, S., Maslej, N., Xiong, B., Zhang, D., & Liang, P. (2023). The Foundation Model Transparency Index (arXiv:2310.12941). arXiv. <https://doi.org/10.48550/arXiv.2310.12941>
- Quelle [3] Du, X., Kolkin, N., Shakhnarovich, G., & Bhattad, A. (2023). Generative Models: What do they know? Do they know things? Let's find out! (arXiv:2311.17137). arXiv. <http://arxiv.org/abs/2311.17137>
- Quelle [4] Pollin, C. (2023). Workshopreihe „Angewandte Generative KI in den (digitalen) Geisteswissenschaften“ (v1.0.0) [Software]. Zenodo. <https://doi.org/10.5281/ZENODO.10065626>
- Quelle [5] „OpenAI Platform“. Zugriffen 29. April 2024. <https://platform.openai.com>.
- Quelle [6] Schulhoff, Sander, Michael Ilie, Nishant Balepur, Konstantine Kahadze, Amanda Liu, Chenglei Si, Yinheng Li, u. a. „The Prompt Report: A Systematic Survey of Prompting Techniques“, 2024. [https://trigaten.github.io/Prompt\\_Survey\\_Site/](https://trigaten.github.io/Prompt_Survey_Site/).

## Disclaimer

Transkript zu dem Video „Generative Modelle: Über Generative Modelle sprechen“, Dr. Jacqueline Klusik-Eckert.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.