

Generative KI in der Lehre

Technische Möglichkeiten und
ethische Herausforderungen

Prof. Dr. Oliver Bendel

2. September 2024



KI, Robotik und Soziale Robotik

In der Disziplin der Künstlichen Intelligenz (KI) geht es um die Schaffung von Computersystemen, die die Ergebnisse menschlicher oder tierischer Verstandestätigkeiten erzielen, ergänzen oder übertreffen können.

Die Robotik beschäftigt sich mit Entwurf, Gestaltung, Steuerung, Produktion und Betrieb von Robotern, etwa von Industrie- und Servicerobotern oder sozialen Robotern (Soziale Robotik).

Generative KI

Bei generativer KI (Generative AI, kurz GenAI) handelt es sich um KI-basierte Systeme, mit denen z.B. Bilder, Video, Audio, Text oder Code erstellt werden können.

Textgeneratoren und Chatbots wie ChatGPT und Bildgeneratoren wie DALL-E und Midjourney beruhen auf grossen Sprachmodellen (Large Language Models, kurz LLMs) oder anderen KI-Modellen. Multimodale LLMs können verschiedene Formate verarbeiten und ausgeben.

Informations-, Roboter- und Maschinenethik

Die Informationsethik untersucht moralische Aspekte des Einsatzes von Informations- und KI-Systemen, KI-Ethik moralische Herausforderungen bei der Entwicklung und Nutzung von KI-Systemen.

Die Roboterethik fragt u.a. nach der Verantwortung und den Rechten und Pflichten von Robotern.

Die Maschinenethik entwickelt zusammen mit Künstlicher Intelligenz und Robotik sogenannte moralische Maschinen.

Technische Möglichkeiten und ethische Herausforderungen

Im Folgenden gehe ich auf technische Möglichkeiten und ethische, soziale und didaktische Herausforderungen von generativer KI in der Lehre ein, und zwar auf der Basis von eigenen Projekten.

Denn nur über die Darstellung von konkreten Anwendungen kann man auf sinnvolle Weise eine solche Einordnung und Beurteilung von KI-Systemen vornehmen.

Produkt: Digital Ethics Girl (2024)



- Digital Ethics Girl ist ein GPT, das Anfang 2024 erstellt wurde
- Sie wurde mit Fachliteratur von Oliver Bendel zur Informationsethik gefüttert
- Sie wird wie Social Robotics Girl (2023) in der Lehre eingesetzt

Hauske, Stefanie; Bendel, Oliver. How Can GenAI Foster Well-being in Self-regulated Learning? In: Proceedings of the AAAI 2024 Spring Symposium Series, Symposium "Impact of GenAI on Social and Individual Well-being". Stanford University, Stanford, California, March 25–27, 2024. The AAAI Press, Washington, DC 2024.
<https://ojs.aaai.org/index.php/AAAI-SS/article/view/31234/33394>.

Technische Möglichkeiten

GPTs oder benutzerdefinierte Versionen von Gemini (Gems) sind niederschwellige Technologien für Studierende und Lehrkräfte, um virtuelle Lernbegleiter zu schaffen.

Mit Hilfe von Prompt Engineering (PE) und Retrieval-Augmented Generation (RAG) – hier das Hochladen von Dokumenten – kann man Wissen hinzufügen und abrufen.

Für die Nutzung sind Registrierung bzw. Anmeldung bei den Firmen erforderlich.

Ethische, soziale und didaktische Herausforderungen

Urheberrecht bei eigenem und fremdem Material in der Wissensdatenbank sind zu beachten.

Trotz PE und RAG kann ein Halluzinieren nicht völlig ausgeschlossen werden – zudem können Fehler beim Zugriff auf Daten entstehen.

Die Studierenden gewöhnen sich an einfache, schnelle Zugriffe und lesen weniger zusammenhängende Texte.

Prototyp: Alpha Mini in der Schule (2023)



- Der soziale Roboter Alpha Mini und ein Lernspiel auf einem Tablet bilden zusammen eine Lernanwendung
- Der soziale Roboter gibt mimisches, gestisches und verbales Feedback
- Er liest zudem die Lösungen vor bzw. erklärt sie bei Bedarf

Bendel, Oliver; Allemann, Andrin. Alpha Mini as a Learning Partner in the Classroom. In: Al Ali, Abdulaziz; Cabibihan, John-John; Meskin, Nader; Rossi, Silvia; Jiang, Wanyue; He, Hongsheng; Ge, Shuzhi Sam (Hrsg.). Social Robotics: 15th International Conference, ICSR 2023, Doha, Qatar, December 7, 2023, Proceedings, Part I. Lecture Notes in Computer Science. Springer International Publishing, Cham 2024. S. 396 – 409.

Prototyp: Alpha Mini in der Schule (2023) – Video

Technische Möglichkeiten

Mit der Verbindung von Robotern wie Alpha Mini und Tablets entstehen günstige und mächtige Lernanwendungen.

Die Verbindung zwischen den Geräten ist noch nicht standardisiert und braucht einigen Aufwand.

Mit LLMs kann man den Robotern zusätzliche Möglichkeiten geben, etwa ihre natürlichsprachlichen Fähigkeiten erweitern.

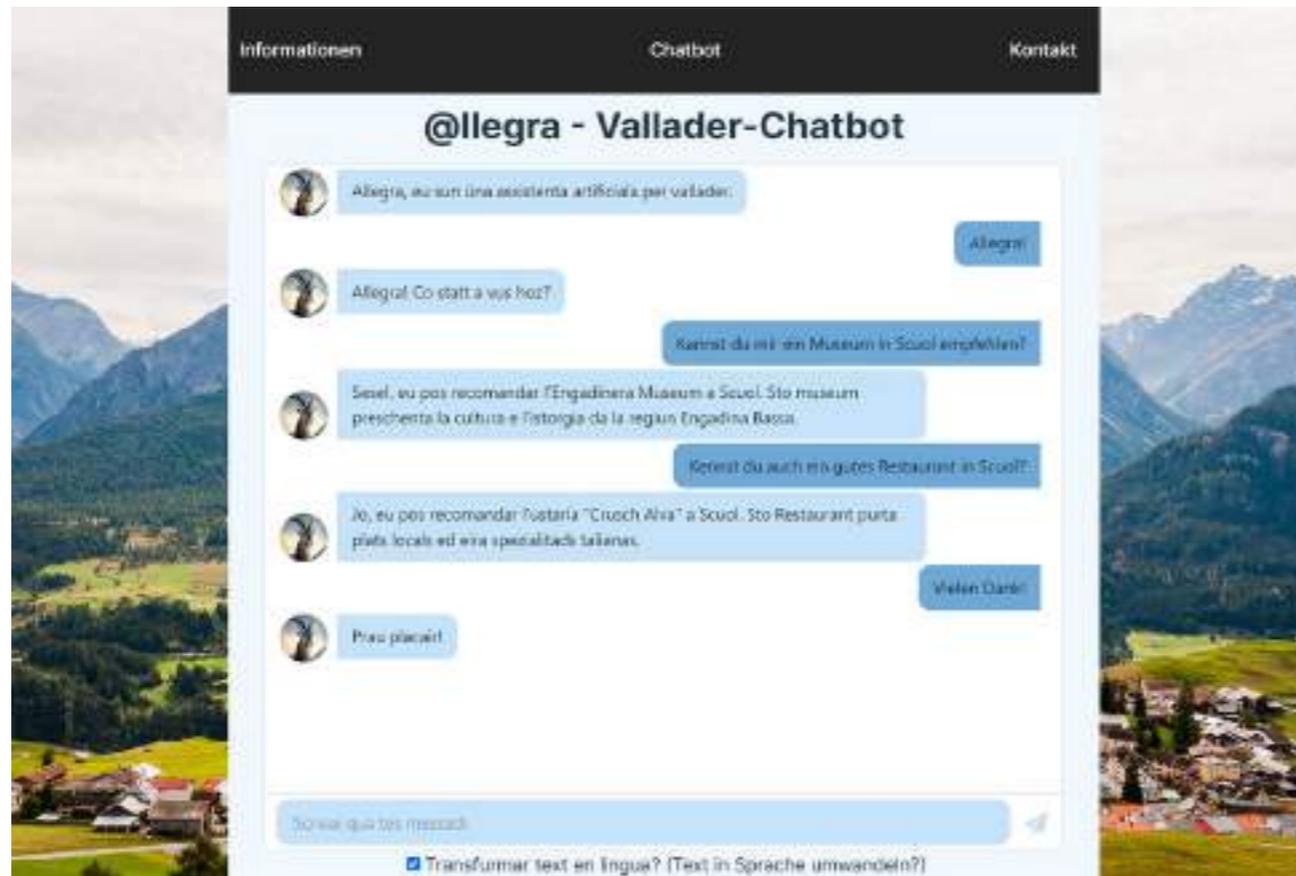
Ethische, soziale und didaktische Herausforderungen

Kinder und Jugendliche lassen sich von sozialen Robotern dieser Art beim Lernen motivieren, Erwachsene weniger.

Interaktion und Kommunikation mit sozialen Robotern sollte diejenige mit Menschen lediglich zeitweise ersetzen.

Speziell bei sozialen Robotern wie Alpha Mini ergeben sich Datenschutzfragen.

Prototyp: @Ilegra (2023)



- Chatbot @Ilegra beherrscht Vallader in geschriebener und gesprochener Form
- Benutzer – etwa Schüler und Studierende – können mit ihr plaudern oder sie zu Graubünden befragen
- Seit August 2024 wird zusammen mit der FH Graubünden ein Chatbot für Sursilvan entwickelt

Bendel, Oliver; Jabou, Dalil. @Ilegra: a chatbot for Vallader. In: International Journal of Information Technology, 19. Februar 2024. <https://link.springer.com/article/10.1007/s41870-024-01779-0>.

Prototyp: @Ilegra (2023) – Video

Technische Möglichkeiten

Einige grosse Sprachmodelle wurden mit Daten zu gefährdeten Sprachen und Idiomen trainiert.

Oft sind die Kenntnisse aber nicht ausreichend oder fehlerhaft, und für eine korrekte Aussprache braucht es spezielle Text-to-Speech-Engines.

Mit Prompt Engineering und Finetuning bzw. mit RAG können die Anwendungen erweitert und verbessert werden.

Ethische, soziale und didaktische Herausforderungen

Kinder und Jugendliche sowie Erwachsene können mit Chatbots dieser Art gefährdete Sprachen erlernen und anwenden.

Chatbots ohne geeignete Anpassung können ihnen ein falsches Vokabular und eine falsche Grammatik beibringen.

Bei GPT-basierten bzw. über Fremdanbieter umgesetzten Lösungen ergeben sich Datenschutzfragen.

Beispiel: Konzepte für soziale Roboter im Wahlmodul (laufend)

RoboRoss



Ranger Dog Robot

- Studierende im Wahlmodul «Soziale Roboter» an der Hochschule für Wirtschaft FHNW müssen Konzepte zu neuartigen sozialen Robotern erstellen
- Dabei müssen sie auch Bilder dieser Roboter vorlegen
- Seit 2023 verwenden sie fast ausschliesslich Bildgeneratoren wie DALL-E, Ideogram und Midjourney

Technische Möglichkeiten

Die Konzepte der Studierenden zu sozialen Robotern wirken insgesamt professioneller.

Je nach Bildgenerator sind bestimmte Ergebnisse – etwa Bilder mit prominenten Personen oder expliziten Inhalten – (un-)möglich.

Für die Nutzung sind Registrierung bzw. Anmeldung bei den Firmen erforderlich.

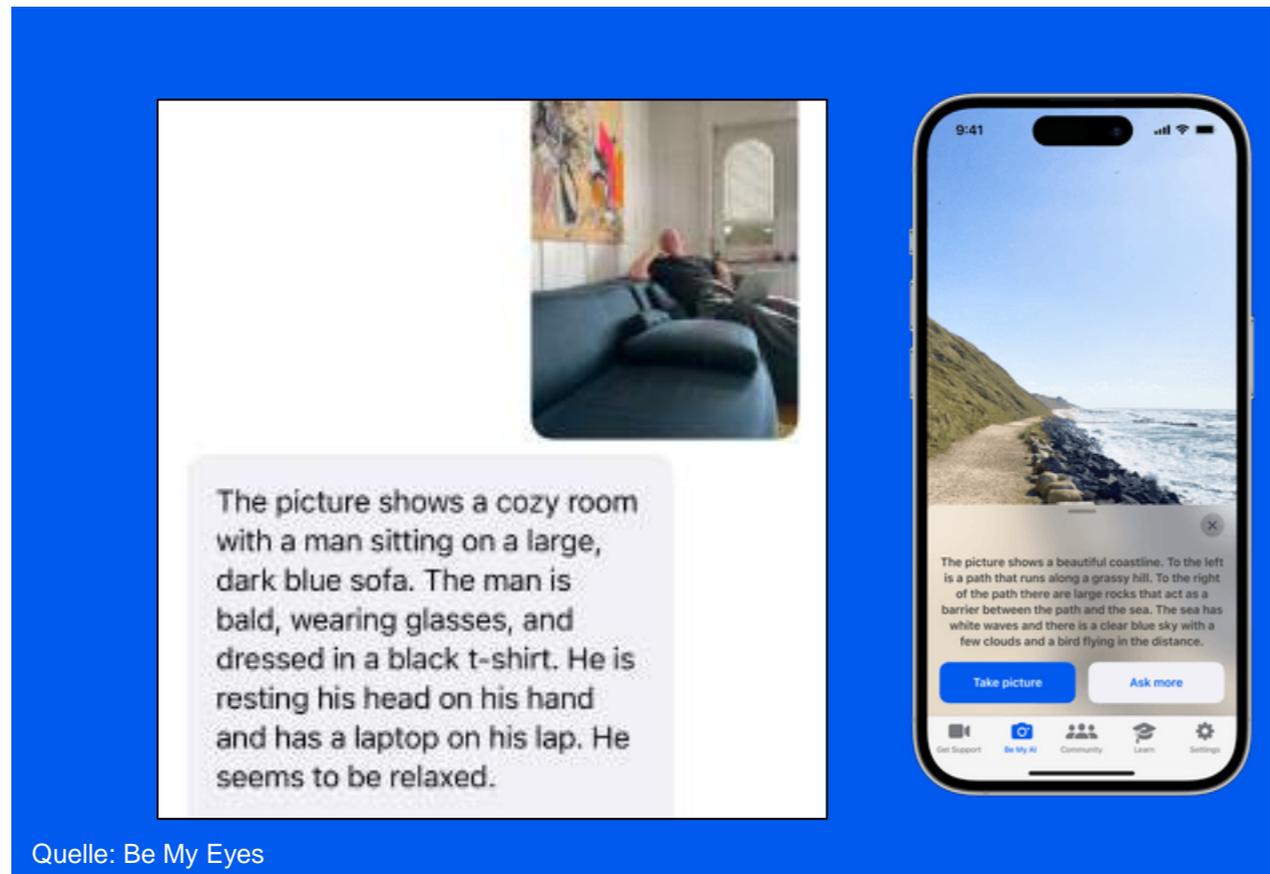
Ethische, soziale und didaktische Herausforderungen

Der eigene Stil geht ein Stück weit verloren, die Bilder ähneln einander und sind zum Teil kitschig.

Das Konzept richtet sich eher am erfolgreich erstellten Bild aus, weniger an den eigenen Ideen.

Guardrails verhindern die Umsetzung bestimmter Bilder und schränken die Benutzer ein.

Produkt: Be My Eyes (eigenes Forschungsprojekt zu fremder Anwendung)



- Be My Eyes mit der Funktion Be My AI analysiert und evaluiert die Umgebung mit Hilfe von Bildern
- Nützliches Tool für blinde und sehbeeinträchtigte Personen
- Auch für andere Benutzer sinnvoll, etwa zum Übersetzen von Text

Bendel, Oliver. How Can Generative AI Enhance the Well-being of Blind? In: Proceedings of the AAAI 2024 Spring Symposium Series, Symposium "Impact of GenAI on Social and Individual Well-being". Stanford University, Stanford, California, March 25–27, 2024. The AAAI Press, Washington, DC 2024.
<https://ojs.aaai.org/index.php/AAAI-SS/article/view/31232/33392>.

Technische Möglichkeiten

Mit Hilfe des multimodalen Sprachmodells werden nicht nur Details der Umgebung, sondern auch die Atmosphäre von Räumen und die Haltung von Personen analysiert und evaluiert.

In Zukunft wird es möglich sein, Bewegtbilder einzubeziehen, wodurch die blinde oder sehbeeinträchtigte Person sich live das Geschehen beschreiben lassen kann.

Ethische, soziale und didaktische Herausforderungen

Apps dieser Art machen Benutzer unabhängiger von menschlicher Hilfe und lassen sie im Alltag und bei der Arbeit sowie im Unterricht partizipieren.

Fehler bei Beschreibungen können ernsthafte Folgen haben, etwa wenn Benutzer einen falschen Weg nehmen.

Guardrails verhindern die Analyse und Evaluation von Bildern und die Antwort auf Anfragen und schwächen damit die Mündigkeit.

Prototyp: The Animal Whisperer (2024)



The screenshot displays a user interface for 'The Animal Whisperer' app. At the top, there is a photo of a brown and white cow in a field. Below the photo is a text input field containing the question: 'Diese Kuh steht auf der Weide vor mir. Soll ich über die Weide gehen?'. Below the input field is a response from the app: 'Danke für das Bild! Lass uns die Situation analysieren:'. Underneath this is a section titled 'Analyse der Körpersprache der Kuh' with four bullet points:

- **Ohrenstellung:** Die Ohren der Kuh sind nach vorne gerichtet, was darauf hindeutet, dass sie aufmerksam und interessiert ist.
- **Kopfhaltung:** Der Kopf ist leicht gesenkt, aber die Kuh wirkt nicht aggressiv. Sie scheint dich zu beobachten und ist möglicherweise neugierig.
- **Augenausdruck:** Die Augen der Kuh wirken ruhig und nicht bedrohlich.
- **Gesamthaltung:** Die Kuh steht entspannt und zeigt keine Anzeichen von Stress oder Aggression.

- Cow Whisperer, Horse Whisperer und Dog Whisperer analysieren und evaluieren Körpersprache und Verhalten der Tiere und die Gesamtsituation
- Sie geben eine Empfehlung für das Verhalten von Menschen ab
- Apps können von Endbenutzern wie Wanderern verwendet werden, aber auch im Rahmen des Unterrichts

Bendel, Oliver; Zbinden, Nick. The Animal Whisperer Project. 2024. Paper für eine Animal-Computer-Interaction-Konferenz (eingereicht im Sommer 2024).

Prototyp: The Animal Whisperer (2024) – Video

Technische Möglichkeiten

Es werden mit Hilfe des multimodalen Sprachmodells nicht nur Körpersprache und Verhalten der Tiere, sondern auch Aspekte der Gesamtsituation analysiert und evaluiert.

Mit Prompt Engineering und Finetuning bzw. mit RAG können die Anwendungen erweitert und verbessert werden.

Ethische, soziale und didaktische Herausforderungen

Apps dieser Art können Tiere und Benutzer vor Gefahren für Leib und Leben schützen.

Wenn sie fehlerhafte Empfehlungen geben, können Gefahren für Leib und Leben entstehen.

Es ist die Frage, wer oder was bei Schäden verantwortlich ist bzw. haftet.

Projekt: Robots in Prison (2024)



- Projekt zur Frage, ob Roboter in Gefängnissen eingesetzt werden können und sollen
- Ein Ergebnis war, dass sich Insassen Roboter und Chatbots für Bildung und Unterhaltung wünschen
- Mögliches Nachfolgeprojekt mit Chatbots auf Tablets

Siegmann, Tamara; Bendel, Oliver. Robots in Prison. In: Palinko, Oskar (Hrsg.). Social Robotics: 16th International Conference, ICSR 2024, Odense, Denmark, 23 – 26 October, 2024, Proceedings, 2024. (Erscheint Ende 2024)

Technische Möglichkeiten

In Gefängnissen können Serviceroboter und soziale Roboter bestimmte Aufgaben übernehmen.

Chatbots und Sprachassistenten müssen lokal auf Tablets oder Notebooks betrieben werden.

Nicht in allen Gefängnissen sind geeignete Räume und Netzwerke vorhanden.

Ethische, soziale und didaktische Herausforderungen

Service-roboter und soziale Roboter können den Insassen Arbeit wegnehmen, die sie machen wollen oder müssen.

Chatbots können zum Äquivalenzprinzip (die Insassen lernen etwas kennen, das ausserhalb der Mauern normal ist) und zum Kompetenzerwerb beitragen sowie Einsamkeit bekämpfen.

Bei Robotern ergeben sich im Gefängnis Datenschutz- und Sicherheitsfragen.

Konzept: Roboterpark in der Schweiz (2024)



- Konzept zu einem Roboterpark in der Schweiz
- Möglichkeit für Kinder, Jugendliche und Erwachsene, mit Robotern zu interagieren und zu kommunizieren und sich zu ihnen Wissen anzueignen
- Auch eigene Prototypen können von den Besuchern entwickelt werden

Konzept: Roboterpark in der Schweiz (2024) – Video

Technische Möglichkeiten

In Roboterparks könnten Benutzer – auch Schüler und Studierende – mit Robotern aller Art interagieren und kommunizieren.

Der fehlerfreie Betrieb ist ebenso eine Herausforderung wie eine regelmässige Wartung und eine erfolgreiche Reparatur.

Im Moment fehlt für viele Roboter in Europa eine Infrastruktur – eine solche müsste der Roboterpark selbst aufbauen.

Ethische, soziale und didaktische Herausforderungen

Benutzer können im Roboterpark ihre Ängste abbauen bzw. zu einer realistischen Einschätzung von Robotern kommen.

Bei der direkten Nutzung einer Vielzahl von Robotern sind Zusammenstöße und Unfälle nicht ausgeschlossen.

Bei Servicerobotern und sozialen Robotern ergeben sich Datenschutzfragen.

Die Bias-Diskussion

Am Ende noch ein Beitrag zur Bias-Diskussion, insbesondere in Bezug auf Stereotype aller Art.

Solche Biases können u.a. durch das Datenmaterial entstehen, durch das Reinforcement Learning from Human Feedback (RLHF) und durch gezieltes «Hineinprogrammieren».

Beispiel: Bilder mit Bias von DALL-E 3 und Gemini

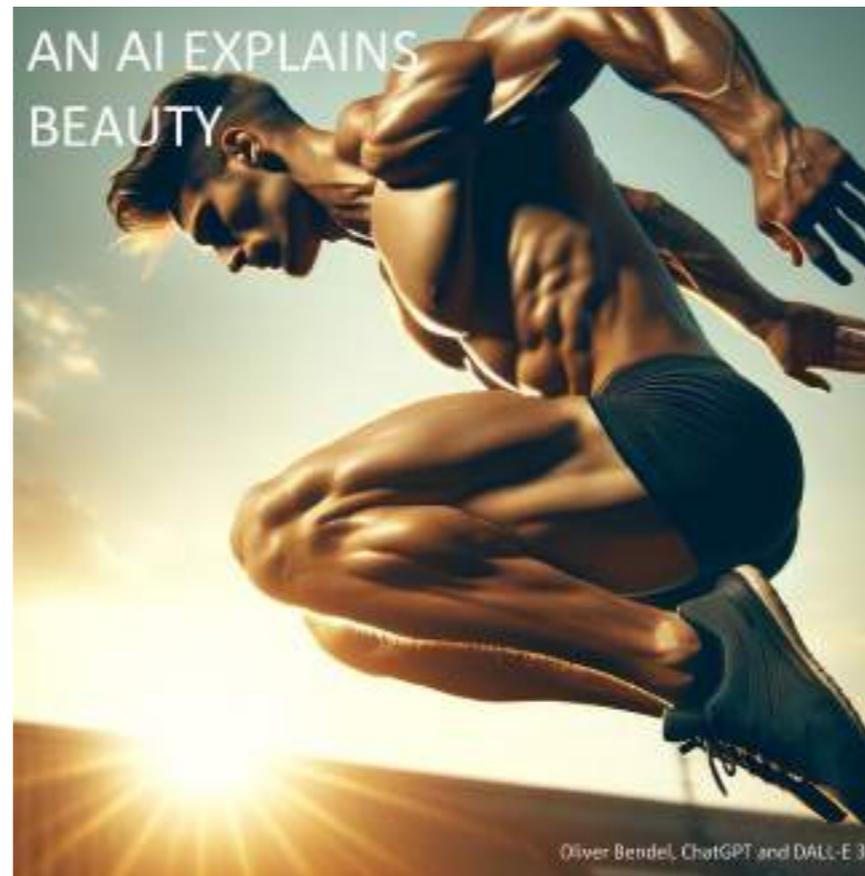


Prompt bei DALL-E 3:
«Black doctor with
white patient»

Prompt bei Gemini: A
«1943 German
soldier»

Bendel, Oliver. Image Synthesis from
an Ethical Perspective. In: AI &
SOCIETY, 27. September 2023.
<https://link.springer.com/article/10.1007/s00146-023-01780-4>.

Beispiel: «On BEAUTY» und «AN AI EXPLAINS BEAUTY» (2023)



Mit DALL-E 3 generierte Bilder und Erklärungen von ChatGPT zu Fragen der Ästhetik

Bilder von DALL-E 3, die «schöne Objekte» enthalten und Erklärungen von ChatGPT dazu, was schön daran ist

Es werden Stereotype sichtbar

Bendel, Oliver (mit DALL-E 3 und GPT-4). On Beauty: 26 questions to an AI in the field of aesthetics. CC, Zürich 2023.

Bendel, Oliver (mit DALL-E 3 und ChatGPT/GPT-4). AN AI EXPLAINS BEAUTY. CC, Zürich 2023.

Beispiel: «AMERICAN SMILE» (2023)



Mit DALL-E 3 generierte Bilder, die die These des dominanten «american smile» bei generativer KI belegen sollen

Alle Personen, welcher Epochen und Kulturen auch immer, bekommen dasselbe Lächeln verpasst

Nur Heidi behält ihr erfrischendes Lachen

Bendel, Oliver (mit DALL-E 3 und ChatGPT/GPT-4). AMERICAN SMILE. Creative Commons, Zürich 2023.

Zusammenfassung und Ausblick I

Generative KI kann in Unterricht und Lehre in vielfältiger Weise eingesetzt werden.

Man kann Anwendungen für Benutzer erstellen, und Benutzer können Anwendungen verwenden, um etwas zu erstellen.

Entsprechend profitieren sowohl Lehrkräfte als auch Schüler oder Studierende.

Zusammenfassung und Ausblick II

In technischer Hinsicht sind bei generativer KI zahlreiche Anwendungen möglich.

Das Halluzinieren kann mit Hilfe von Prompt Engineering und Finetuning beschränkt werden. Eine besondere Rolle spielt RAG.

Ständig kommen neue Modelle und Versionen auf den Markt, und es ist selbst für Experten schwer, die Vor- und Nachteile der jeweiligen LLMs einzuschätzen.

Zusammenfassung und Ausblick III

In ethischer Hinsicht gibt es viele Möglichkeiten und Herausforderungen.

Dazu gehören Datenschutz und Privatsphäre, die Dominanz moralischer Vorstellungen der Anbieter und die Vereinfachung und Verflachung von Inhalten in der Lehre.

Zugleich stehen mit GenAI mächtige Tools zur Verfügung, die das Lernen und Arbeiten erleichtern können und die viel Spass machen.

Zusammenfassung und Ausblick IV

Eine letzte Übersicht zu Themen mit Blick auf Textgeneratoren:

- Urheberrecht und Fremdnutzung sowie Urheberschutz von Werken
- Datenschutz und informationelle Autonomie bei Prompts
- Verantwortung und Haftung angesichts von (Teil-)Autonomie
- Stereotype, diskriminierende, rassistische und sexistische Darstellungen
- Falschdarstellungen von Sachverhalten
- Zurückweisungen und Einschränkungen
- Wissenschaftlichkeit und Referenzierbarkeit
- Verbesserung, Veränderung und Ersetzung der Arbeit

Bendel, Oliver. KI-basierte Textgeneratoren aus Sicht der Ethik. In: Schreiber, Gerhard; Ohly, Lukas (Hrsg.). KI:Text: Diskurse über KI-Textgeneratoren. De Gruyter, Berlin/Boston 2024. S. 293 – 308. <https://www.degruyter.com/document/doi/10.1515/9783111351490/html>.

Vertiefend und auch mit Blick auf Bildgeneratoren:

Bendel, Oliver. Generative KI aus ethischer Sicht. In: Seufert, Sabine; Handschuh, Siegfried. Generative Künstliche Intelligenz: ChatGPT und Co für Bildung, Wirtschaft und Gesellschaft. Schäffer-Poeschel, Stuttgart 2024. S. 167 – 181.

Kontakt

Prof. Dr. Oliver Bendel
Hochschule für Wirtschaft FHNW
Bahnhofstrasse 6
CH-5210 Windisch

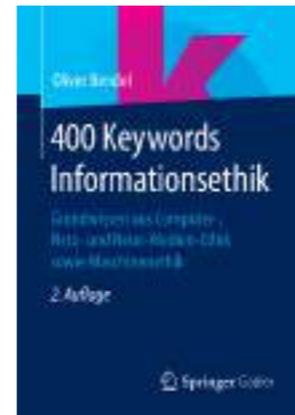
oliver.bendel@fhnw.ch

oliverbendel.net

informationsethik.net

maschinenethik.net

robophilosophy.com



2016/19



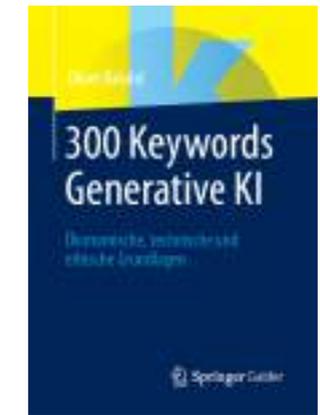
2019



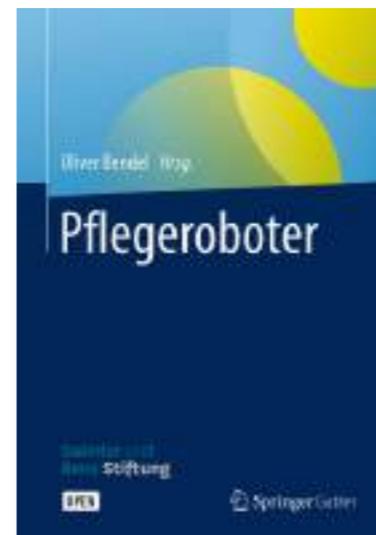
2021



2022



2024



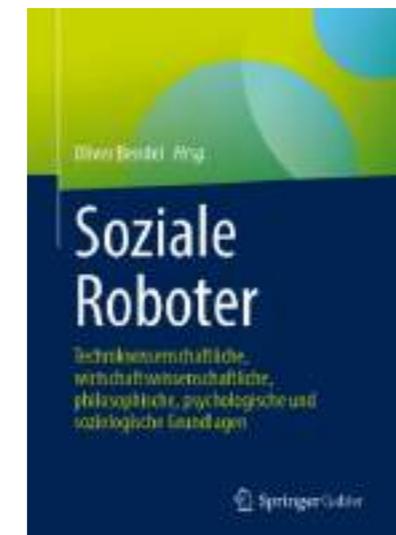
2018



2019



2020



2021