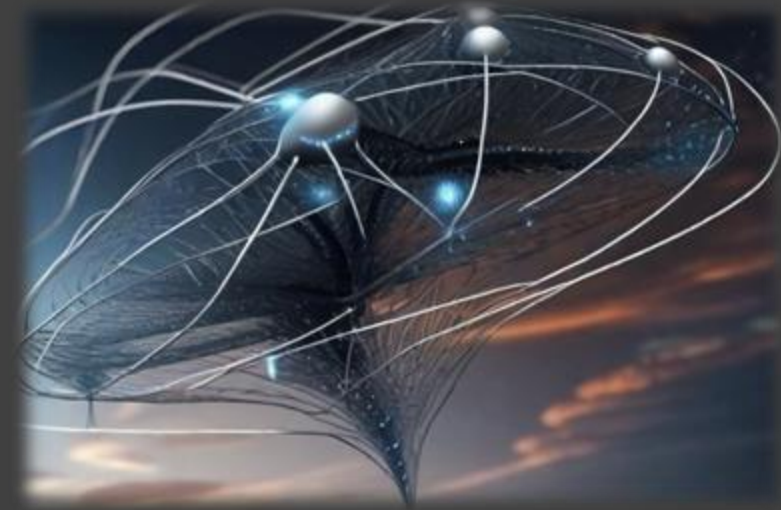


Vorurteile & Stereotypen vermeiden



KI-Kompetenzen in Bildung & Arbeitswelt

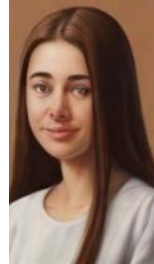
Tina Wölfel, M.Sc

Vorstellung

- Tina Wöfl
- B.A. Personalmanagement
- M.Sc. Wirtschaftspädagogik mit Zweifach
Wirtschaftsinformatik
- wissenschaftliche Mitarbeiterin am FAU Schöller
Stiftungslehrstuhl für Wirtschaftsinformatik, insb. Digitalisierung
in Wirtschaft und Gesellschaft
- Forschungsschwerpunkt:
 - KI-Kompetenzen, insbesondere KI Awareness



Vorstellungsrunde



Agenda

- 1 Grundlagen
- 2 Vorurteile & Stereotypen vermeiden
- 3 Beispiele

Grundlagen

KI – KI-Kompetenzen – Funktionsweise von KI basierten Systemen

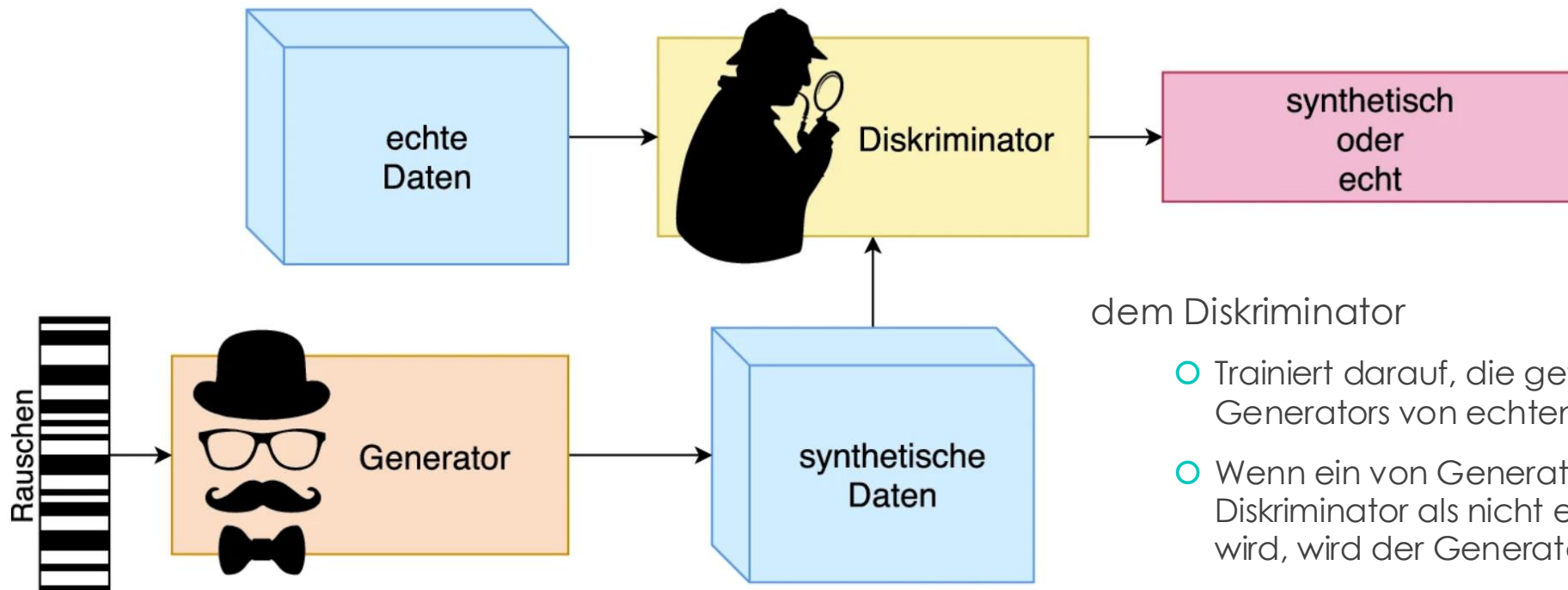
Künstliche Intelligenz & KI-Kompetenzen

- Neue Anforderungen auf Grund der Veränderung der Art & Weise wie Mensch & Maschine miteinander kommunizieren, arbeiten & leben
- Zielkonzept muss den Menschen zu gute kommen, auch wenn die Maschine die Ziele nicht kennt
- Durch KI kommt es zu Kompetenzverschiebungen
 - Hierbei besonders von Bedeutung: KI Awareness – Teilbereich der KI Kompetenzen (digitale-technologische Kompetenzen) & Reflexionskompetenz

(Long & Magerko 2020; Kasinidou 2023; Franken et al. 2022; Russell & Norvig 2023)

Generative Adversarial Network (GAN)

Besteht aus zwei neuronalen Netzen (GenKI) die gegeneinander antreten



dem Diskriminator

- Trainiert darauf, die gefälschten Daten des Generators von echten Beispielen zu unterscheiden
- Wenn ein von Generator erzeugtes Bild, vom Diskriminator als nicht eindeutig plausibel eingestuft wird, wird der Generator bestraft

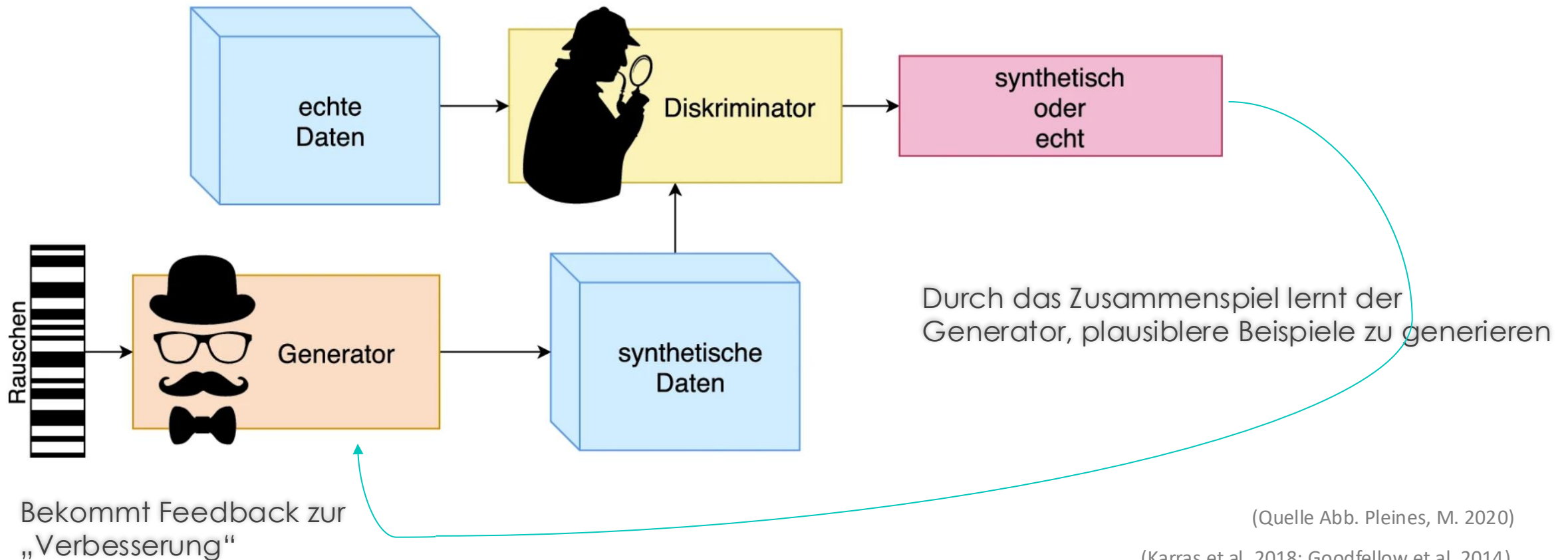
dem Generator

- Trainiert darauf, gefälschte Daten zu produzieren

(Quelle Abb. Pleines, M. 2020)

(Karras et al. 2018; Goodfellow et al. 2014)

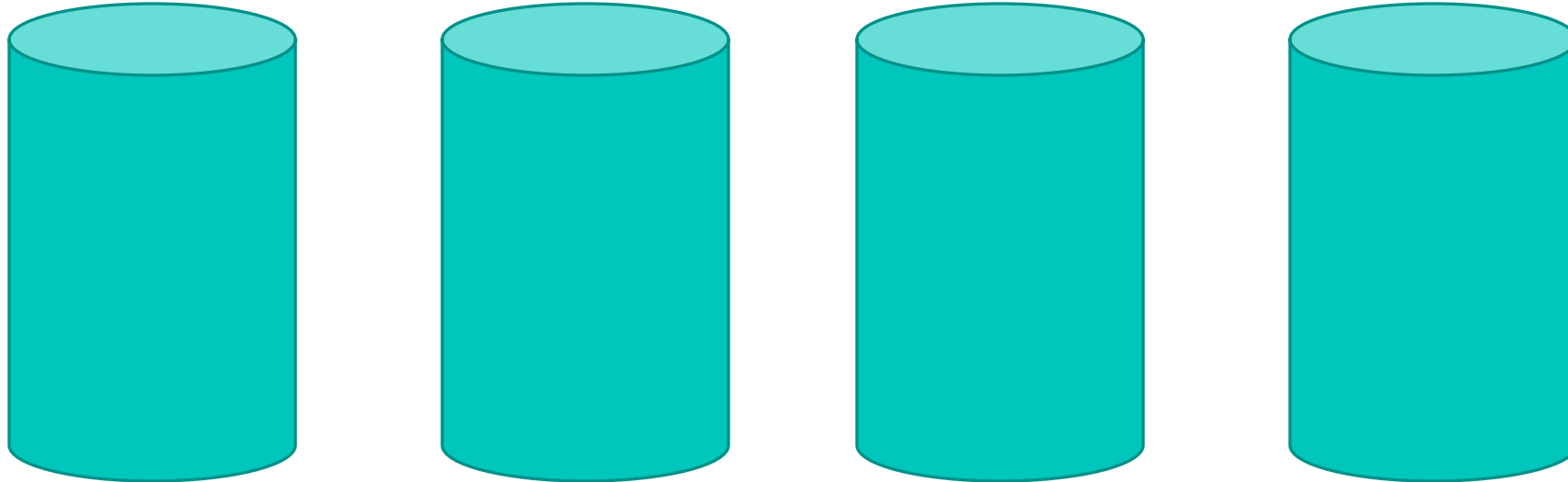
Generative Adversarial Network (GAN)



(Quelle Abb. Pleines, M. 2020)

(Karras et al. 2018; Goodfellow et al. 2014)

(Trainings)Datenbanken



- Dateninhalt oft eine große Blackbox
- Daten vergangenheitsorientiert
- Spiegeln gesellschaftliche Probleme wieder, sind daher nicht Vorurteilsfrei & divers

Blackbox - Datenherkunft

- Woher kommen die Inhalte des Textes?
- In wie weit werden komplette Sätze, Absätze 1:1 übernommen?
- Sind die Texte „neu“ im Sinne von es handelt sich nicht um ein direktes Zitat?
- Woher kommen die Daten, um diese Bilder Grafiken zu erstellen?
- Sind Bilder soweit verändert, dass das ursprüngliche Kunstwerk nicht mehr erkennbar ist?

Unklarheiten und gleichzeitig die Gefahr bei der Übernahme eines KI generierten Textes/Bildes

- Rechtliche Aspekte:
 - Urheberrechtsverletzung
 - Verstoß gegen das Recht am eigenen Bild
 - Verstoß gegen das Recht am eigenen Text
 - Verstoß des höchstpersönlichen Lebensbereichs und Persönlichkeitsrechten
 - Strafrechtliche Relevanz

Vorurteile & Stereotypen vermeiden

Ethische Aspekte – Sicherheit – kritische Reflexion

Ethische Aspekte

- KI-Systeme werden von Menschen trainiert
 - Menschen diskriminieren unter Umständen (bewusst/unbewusst), dies führt wiederum zu einer nicht diversen KI
 - In der IT Branche befinden sich vorwiegend männliche Personen
 - Daten sind zu wenig divers – führen u.a. zu Stereotypen
- Medizinische KI-Systeme
 - Trainiert auf überwiegend männlichen Daten
 - Weibliche Daten sind kaum vorhanden
 - Krankheiten werden bei Männern früher erkannt
- KI-Systeme brauchen „Regeln“ für Entscheidungen
 - Hierbei gibt es Grenzen – Thema: Autonomes Fahren - Reaktionen bei Gefahr

Ist KI weiblich?

- Eingabe von „Artificial Intelligence“ es kommen Bilder mit Frauen...
 - Frauen, die immer sehr ähnlich aussehen, einen ähnlichen Blick haben, „hübsch“ und „jung“ sind
 - Grundsätzlich ist bekannt, dass die weibliche Stimme, Aussehen etc. mehr Vertrauen erweckt als bei dem männlichen Geschlecht. Dies trifft ebenso auf weibliche Namen zu
 - Wenn dieser Aspekt intensiv mit der KI trainiert wird, wird dieser sich verfestigen

➤ Entstehung von Vorurteilen & Stereotypen

KI hat kein Geschlecht



Sicherheit im Umgang mit KI-basierten Systemen

- Echtheit von Texten – prüfen & hinterfragen
- Keine Eingabe von sensiblen Daten in KI-basierte Systeme
- Datenschutz des Tools „lesen“ – insb. die Thematik rund um die Eingaben
- Kostenlos nutzen, bedeutet sehr oft eine Datenspende – Daten werden zum Trainieren verwendet – daher hier, besondere Vorsicht hinsichtlich privater Daten
 - Standardmäßig ist oft die „Datenspende“ voreingestellt
 - Keine sensiblen Daten eingeben / hochladen
 - Keine privaten Bilder hochladen
- Vor dem ersten Nutzen solcher Systeme, überprüfen in wie weit der „private Modus“ bereits aktiv ist oder erst aktiviert werden muss
- Kritisches Nutzen – Systeme sind oft intransparent und undurchsichtig

Kritische Reflexion

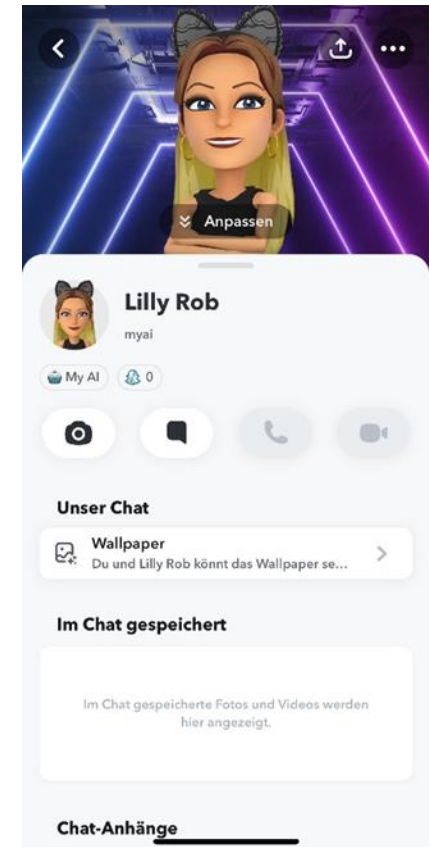
- Echtheit der „Antwort“ muss immer hinterfragt werden
 - Wieviel Wahrheit steckt im Inhalt?
 - Wie vertrauenswürdig sind die angegebenen Quellen?
- Letzte Entscheidung trifft immer der Mensch
- Benötigtes Fachwissen muss vorhanden sein
- Grundwissen „wie funktionieren die verwendeten Systeme“ muss vorhanden sein
- Wissen darüber, worin steckt KI oder zumindest worin KI stecken könnte

Beispiele

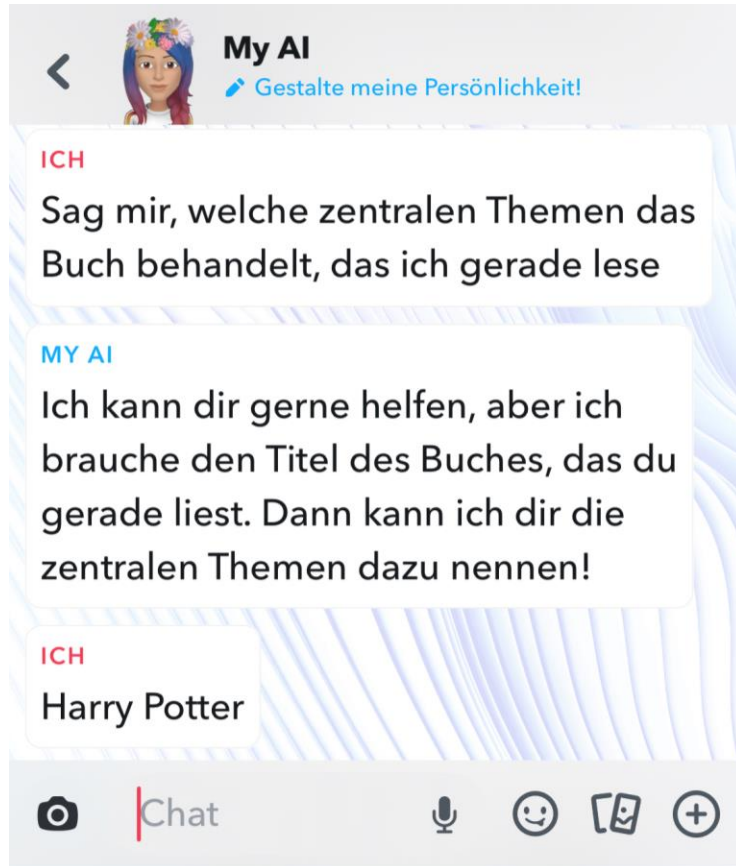
Snapchat – perplexity – Leonardo.AI – was ist KI generiert

Snapchat „My AI“

- My AI ist eine künstliche Intelligenz
- My AI ist wie „chatten“ mit Freunden
- Mit My AI geteilte Inhalte, einschließlich Standort, wird von My AI verwendet:
 - Um relevante und nützliche Antworten zu geben
 - Empfehlungen in der Nähe zu auszusprechen
 - Daten werden verwendet um Snapchat zu verbessern
 - Für Personalisierte Werbeanzeigen
- Im Gegensatz zu anderen Chats mit Freunden werden alle Nachrichten mit My AI nicht nach 24h automatisch gelöscht
- Löschen, ist möglich – Dauert bis zu 30 Tage
 - Inhalte die gespeichert wurden, werden nicht gelöscht
 - Auf Basis der eingegeben Daten fand eine „Weiterentwicklung“ statt



Beispiel „My AI“



In Band 1, "Harry Potter und der Stein der Weisen", geht es um einen jungen Jungen namens Harry Potter, der an seinem elften Geburtstag erfährt, dass er ein Zauberer ist. Er wird an die Hogwarts-Schule für Hexerei und Zauberei eingeladen, wo er Freunde wie Ron Weasley und Hermine Granger trifft. Während seines ersten Jahres entdeckt Harry, dass er berühmt ist, weil er als Baby den dunklen Zauberer Voldemort überlebt hat.

Die Geschichte folgt Harrys Abenteuern an Hogwarts, wo er die Geheimnisse der Zauberei lernt und schließlich auf das Geheimnis des Steins der Weisen stößt, der Unsterblichkeit verleihen kann. Harry und seine Freunde müssen den Stein vor Voldemort und seinen Anhängern schützen. Es ist eine Geschichte über Selbstentdeckung, Freundschaft und den Kampf gegen das Böse.

perplexity



www.perplexity.ai

perplexity

Neuer Thread **Ctrl** **I**

Zuhause
Entdecken
Bibliothek

Was werden Sie entdecken?

Sie fragen, wir antworten...

Fokus + Anhängen

Preis von Bitcoin

Beliebteste YouTuber in 2024

Registrieren
Einloggen

Pro Unternehmen Geschäft

Sie fragen, wir antworten...

Fokus + Anhängen Pro

- Web**
Über das gesamte Internet suchen
- Akademisch**
Suche nach veröffentlichten akademischen Arbeiten
- Mathematik**
Gleichungen lösen und numerische Antworten finden
- Schreiben**
Text oder Chat erstellen, ohne im Web zu suchen
- Video**
Entdecken und Videos ansehen
- Sozial**
Nach Diskussionen und Meinungen suchen

en verfügbaren

Was weiß perplexity über Sie? Geben Sie ein „wer ist ...?“

Was ist KI generiert?

Die Geschichte der künstlichen Intelligenz (KI) reicht bis in die Mitte des 20. Jahrhunderts zurück. Der Begriff "Künstliche Intelligenz" entstand im Sommer 1956 auf einer Konferenz am Dartmouth College in den USA. Wissenschaftler diskutierten die Möglichkeit, Aspekte des Lernens und andere Merkmale der menschlichen Intelligenz von Maschinen simulieren zu können. Während dieser Konferenz wurde das erste KI-Programm der Welt, der "Logic Theorist", geschrieben. In den 1970er und 1980er Jahren gab es eine Phase der Ernüchterung, die als "KI-Winter" bekannt ist, aufgrund überhöhter Erwartungen und anschließender Stagnation. Seitdem hat die KI jedoch bedeutende Fortschritte gemacht, wie beispielsweise der Sieg des KI-Programms "Watson" in einer Quizshow im Jahr 2011 und der Sieg des KI-Programms "AlphaGo" im Spiel Go. Die Entwicklung der KI steht jedoch noch relativ am Anfang, und es wird daran gearbeitet, sie zuverlässiger und sicherer gegen Manipulationen zu machen, insbesondere in sensiblen Bereichen wie dem automatisierten Fahren oder der Medizin.



Eine schnelle Möglichkeit, einen Überblick über die Meilensteine der KI-Geschichte zu bekommen, ist die Auflistung der Preisträger des Turing Awards: Marvin Minsky (1969) und John McCarthy (1971) für die Definition der Grundlagen des Forschungsgebiets auf der Basis von Repräsentation und Schlussfolgerungen; Allen Newell und Herbert Simon (1975) für symbolische Modelle der Problemlösung und menschlicher Kognition; Ed Feigenbaum und Raj Reddy (1994) für die Entwicklung von Expertensystemen, die menschliches Wissen codieren, um reale Probleme zu lösen; Judea Pearl (2011) für die Entwicklung probabilistischer Schlussfolgerungstechniken, die mit Unsicherheiten auf grundlegende Weise umgehen; und schließlich Yoshua Bengio, Geoffrey Hinton und Yann LeCun (2019), die „Deep Learning“ (mehrschichtige neuronale Netze) zu einem wichtigen Bestandteil der modernen Datenverarbeitung gemacht haben. Im weiteren Verlauf dieses Abschnitts gehen wir näher auf die einzelnen Phasen der KI-Geschichte ein.

Was ist KI generiert?

Deep Learning – dabei geht es um künstliche neuronale Netze, die mindestens zwei, meist aber sogar sehr viele sogenannte *versteckte Schichten* (engl. *hidden layers*) haben. Mit diesen erlernen sie auf Basis von Trainingsdaten weitgehend eigenständig, also mit möglichst geringem menschlichem Zutun, interne Repräsentationen der Eingangsdaten, die es ihnen dann ermöglichen, auch äußerst komplexe Probleme zu lösen.



Deep Learning ist eine fortschrittliche Technik im Bereich der künstlichen Intelligenz und des maschinellen Lernens. Es handelt sich um eine Untergruppe des maschinellen Lernens, die auf künstlichen neuronalen Netzwerken basiert, die dem menschlichen Gehirn nachempfunden sind.

Was ist KI generiert?



Was ist KI generiert?



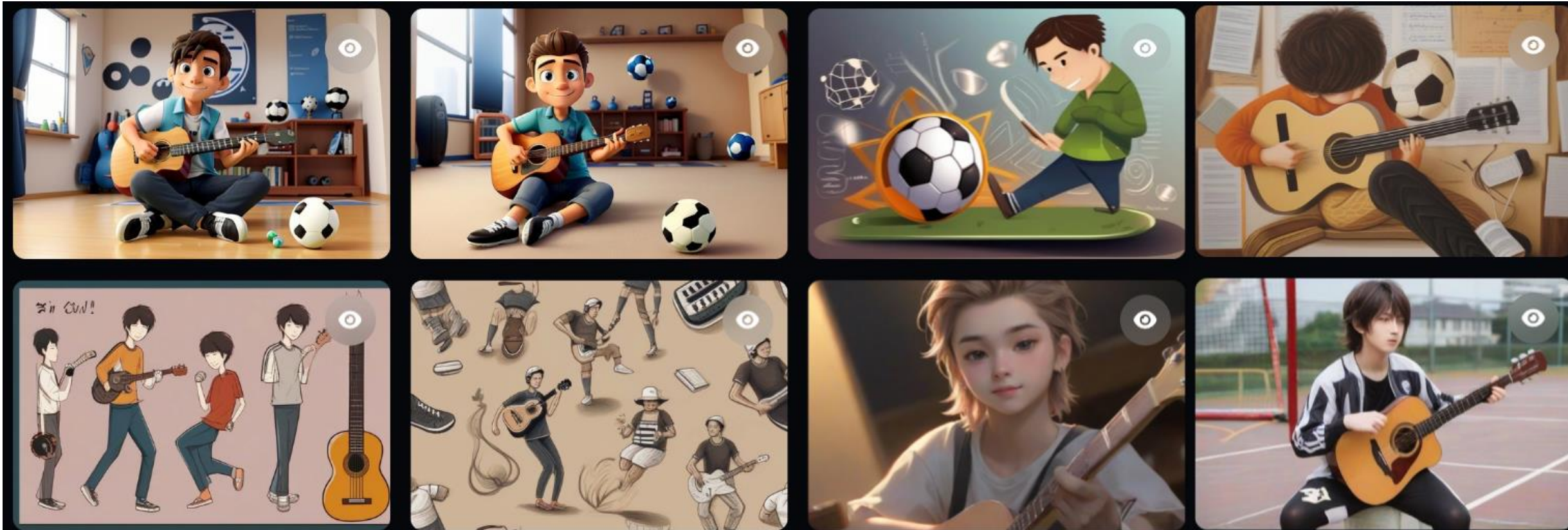
Eine Person mit Hobby malen



Welche Gedanken entstehen bei der Sicht auf diese Bilder?



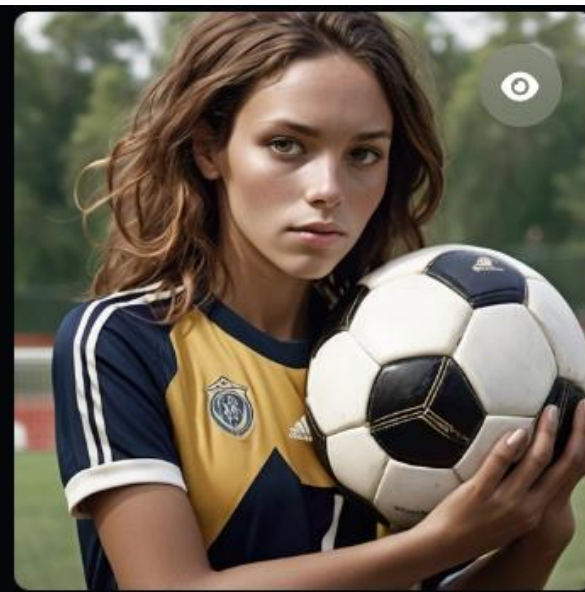
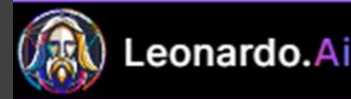
Eine Person mit Hobby Fußball & Gitarre



Eine Person mit Hobby Fußball



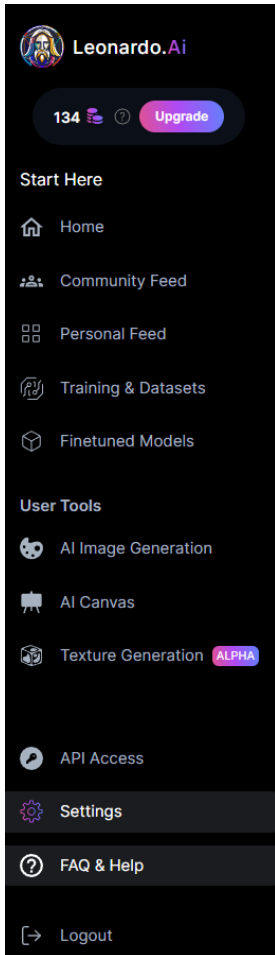
Eine weibliche Person mit Hobby Fußball



Leonardo.AI



www.leonardo.ai



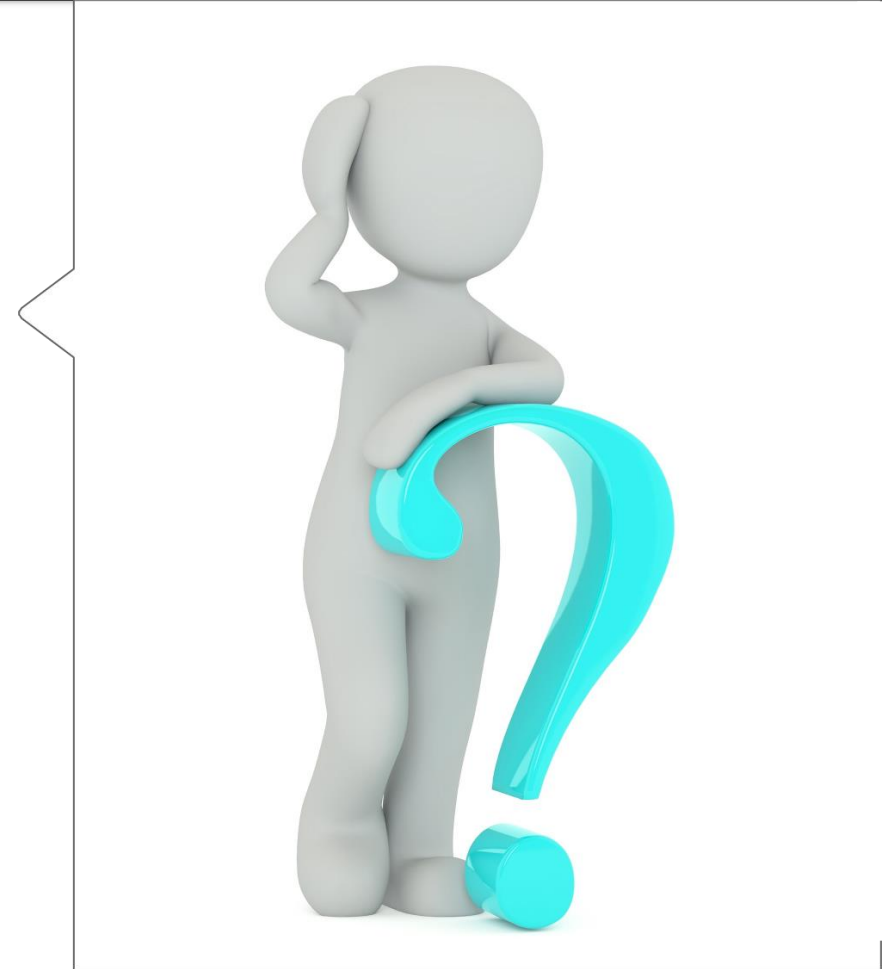
- Leonardo.AI ist ähnlich wie Dall-E
 - Hat jedoch auch eine integrierte zusätzliche Bildbearbeitung – AI Canvas
 - Kostenlose „Einsteigerversion“ – täglich 150 neue Tokens
 - Verschiedene Datenbanken auswählbar

Abschluss

Abschluss

- Wünsche, Fragen, Anregungen

Tipp: Leitfaden für
fares KI-Prompting





Herzlichen Dank für die Aufmerksamkeit!

Tina Wöfl, M.Sc.

tina.woelfl@fau.de

www.woelflworks.de

Vorurteile & Stereotypen vermeiden - KI-Kompetenzen in Bildung & Arbeitswelt

Quellen

- André, Elisabeth; Bauer, Wilhelm (2021): Kompetenzentwicklung für Künstliche Intelligenz. Unter Mitarbeit von Akademie der Technikwissenschaften.
- Goodfellow, Ian J.; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil et al. (2014): Generative Adversarial Networks. Online verfügbar unter <http://arxiv.org/pdf/1406.2661v1>.
- Franken, Swetlana; Mauritz, Nina; Prädikow Lotte. 2022. Kompetenzen für KI-Anwendungen: Theoretisches Modell und partizipative Erfassung und Vermittlung in Unternehmen. In Conference: GfA, Sankt Augustin (Hrsg.): Frühjahrskongress 2022, Technologie und Bildung in hybriden Arbeitswelten At: Magdeburg
- Karras, Tero; Laine, Samuli; Aila, Timo (2018): A Style-Based Generator Architecture for Generative Adversarial Networks. Online verfügbar unter <http://arxiv.org/pdf/1812.04948v3>.
- Kasinidou, Maria. 2023. AI Literacy for All: A Participatory Approach. In Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 2 (ITiCSE 2023), July 8–12, 2023, Turku, Finland. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3587103.3594135>
- Long, Duri; Magerko, Brian. 2020. What is AI Literacy? Competencies and Design Considerations. In CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. April 2020. Pages 1–16. <https://doi.org/10.1145/3313831.3376727>
- Russell, Stuart; Norvig, Peter (2023): Künstliche Intelligenz. Ein moderner Ansatz. 4., aktualisierte Auflage. München: Pearson Studium ein Imprint von Pearson Deutschland; Pearson Studium (Pearson Studium - IT). Online verfügbar unter <https://elibrary.pearson.de/book/99.150005/9783863263263>.
- Pleines, M. (2020). Generative Adversarial Networks: Verschiedene Varianten und Anwendungen aus der Praxis. In: Kahl, T., Zimmer, F. (eds) Interaktive Datenvisualisierung in Wissenschaft und Unternehmenspraxis. Springer Vieweg, Wiesbaden. https://doi.org/10.1007/978-3-658-29562-2_8
- Sowie Bilder & Grafiken von pixabay und Leonardo.AI