



De-Biasing durch Prompting

Input & Workshop für die MINTvernetz Jahrestagung
11. Februar | Nicole Dierolf | wissenschaftl.
Mitarbeiterin am Lab für Sozioinformatik

Heute:

- Lab für Sozioinformatik, Hochschule Heilbronn
- Chancengleichheit: Rahmenmodell der sozialverantwortlichen KI-Gestaltung
- Large Language Model (LLM)
- Workshop: De-Biasing durch Prompting
- Diskussion & Feedback



Interdisziplinäres Team

Die **Sozialinformatik** beschreibt die interdisziplinäre Untersuchung der Gestaltung, der Nutzung und der Folgen von Informationstechnologien in bestimmten kulturellen oder institutionellen Kontexten, insbesondere der möglichen Wechselwirkungen zwischen Individuum, Technologie und Kontext.



Prof. Dr. Nicola Marsden



Rebecca Biebl
Computer Science



Julian Leidel

Human-Computer-Interaction



Monika Pröbster

Psychology



Annelie Rothe-Wulf

Edda Sellin



Socio-Cultural Anthropology



Yvonne Tang

Sociology & Gender Studies



Claudia Herling

Design & User Experience



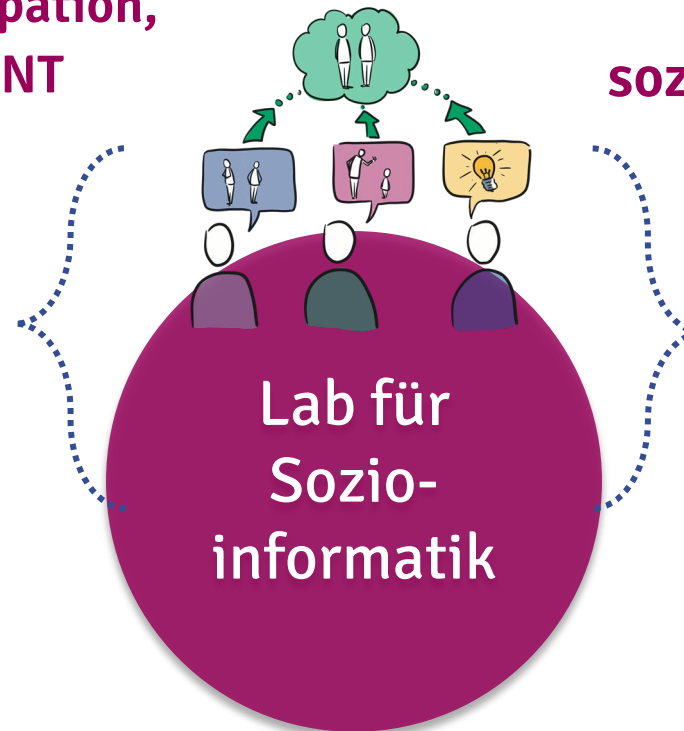
Nicole Dierolf



Aktuelle Forschungsprojekte

Forschung zu den Themen: Partizipation,
Geschlecht, Methoden und MINT

Projekte zur Förderung fairer und
sozialverträglicher Anwendungen von KI



MAKE it
REAL

MAKEitREAL – ein mobiler Makerspace für Mädchen mit Zuwanderungsgeschichte im Alter von 10 bis 16 Jahren.



Edda Sellin

Nicole Dierolf

Annelie Rothe-Wulf



fair.impact : Sozialverantwortliche KI-Gestaltung

KURS

Teampraktiken für sozialverantwortliche KI-Gestaltung

Hochschule Heilbronn ...

Leistungsnachweis

5 Wissensnuggets à 30-45 Minuten

Start: jederzeit

KI & Unternehmen

Gesellschaftliche Fragestellungen zu KI

KURS

Methoden der Bias-Reduktion für eine sozialverantwortliche KI-Gestaltung

Hochschule Heilbronn ...

Leistungsnachweis

5 Wissensnuggets à 30-45 Minuten

Start: jederzeit

KI & Unternehmen

Gesellschaftliche Fragestellungen zu KI

KURS

Sozialverantwortliche KI-Gestaltung

Hochschule Heilbronn ...

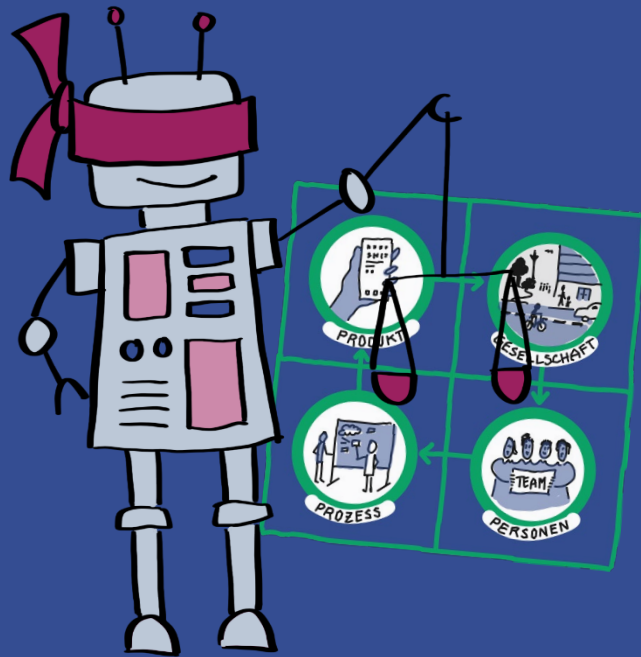
Leistungsnachweis

5 Wissensnuggets à 30-45 Minuten

Start: jederzeit

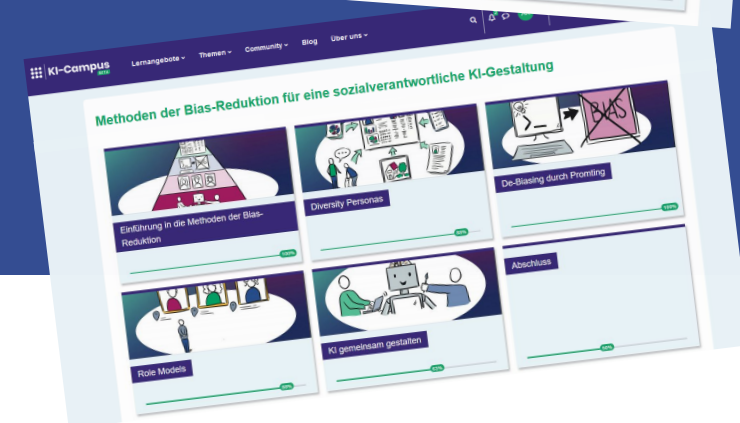
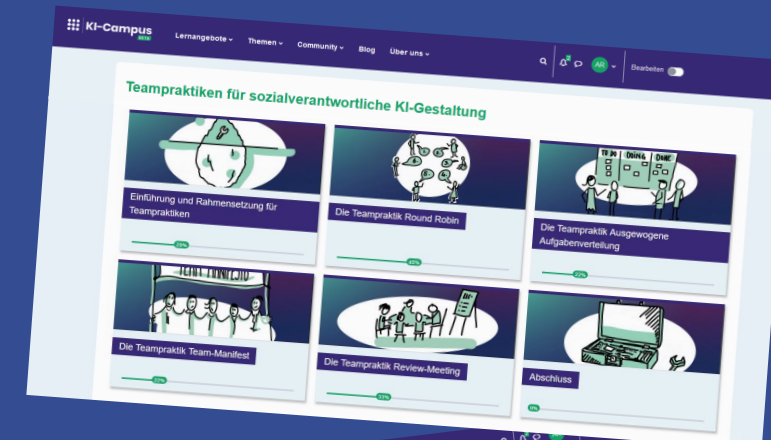
KI & Unternehmen

Gesellschaftliche Fragestellungen zu KI



... unser Ansatz:

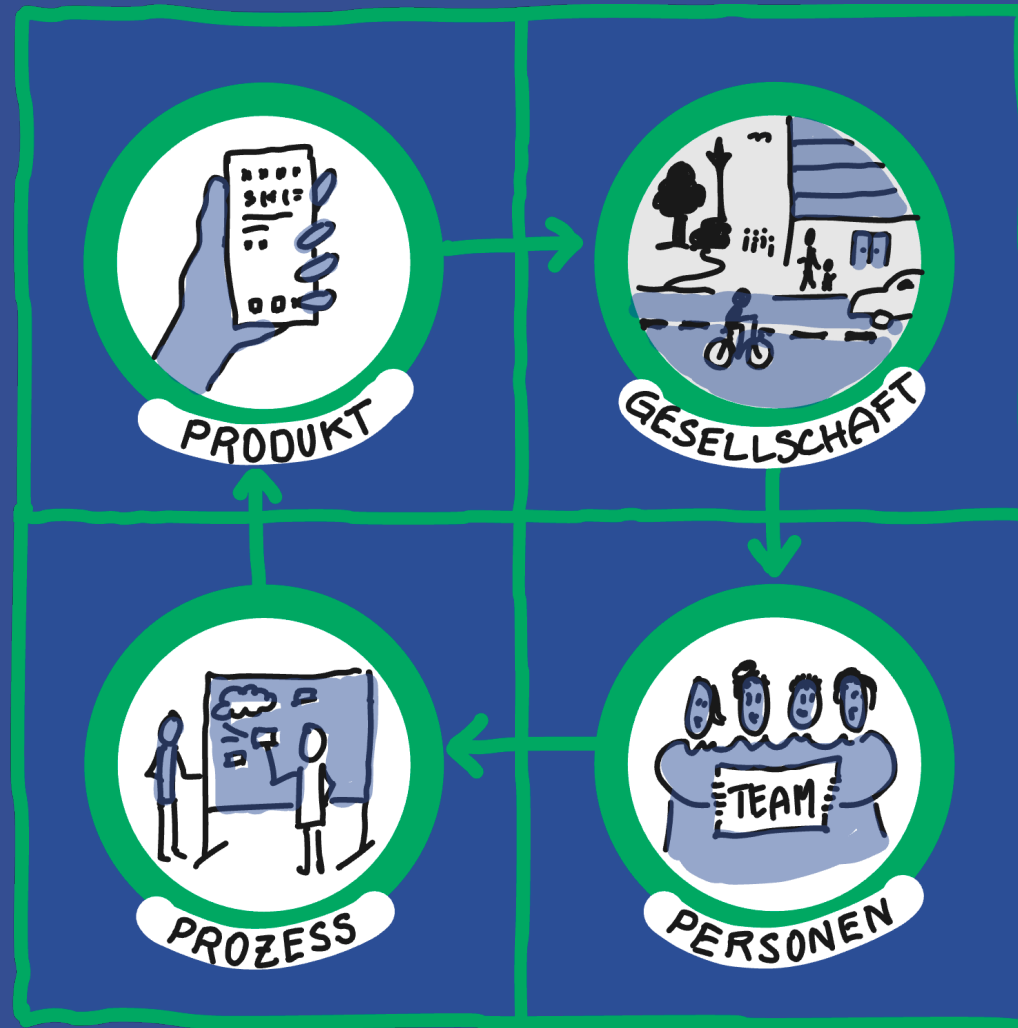
- Förderung von Diversität und Inklusion im gesamten KI-Entwicklungsprozess
- Forschung zu sozialverantwortlicher KI in der Praxis anwendbar und verfügbar machen.
- Vermittlung von Methoden und bias-bezogenem Wissen zur Schaffung von Diversity- und Gender-Kompetenz



Heute:

- Lab für Sozioinformatik, Hochschule Heilbronn
- Chancengleichheit: Rahmenmodell der sozialverantwortlichen KI-Gestaltung
- Large Language Model (LLM)
- Workshop: De-Biasing durch Prompting
- Diskussion & Feedback





Rahmen- modell zur Gestaltung sozial- verantwort- licher Künstlicher Intelligenz

Apple Health App

„Apple’s Health Kit can help you keep track of your blood alcohol content. If you’re still growing, it’ll track your height. And if you have an inhaler, it’ll help you track how often you use it. You can even use it to input your sodium intake, because „with Health, you can monitor all of your metrics that you’re most interested in,” said Apple Software executive Craig Federighi [...]"

Arielle Duhaime-Ross. 2014. Apple promised an expansive health app, so why can't I track menstruation? Retrieved from <https://www.theverge.com/2014/9/25/6844021/apple-promised-an-expansive-health-appsowhy-cant-i-track>.



<https://www.idownloadblog.com/2018/12/03/prioritize-iphone-apple-watch-health-app-steps/>

Ein echtes „Highlight“....

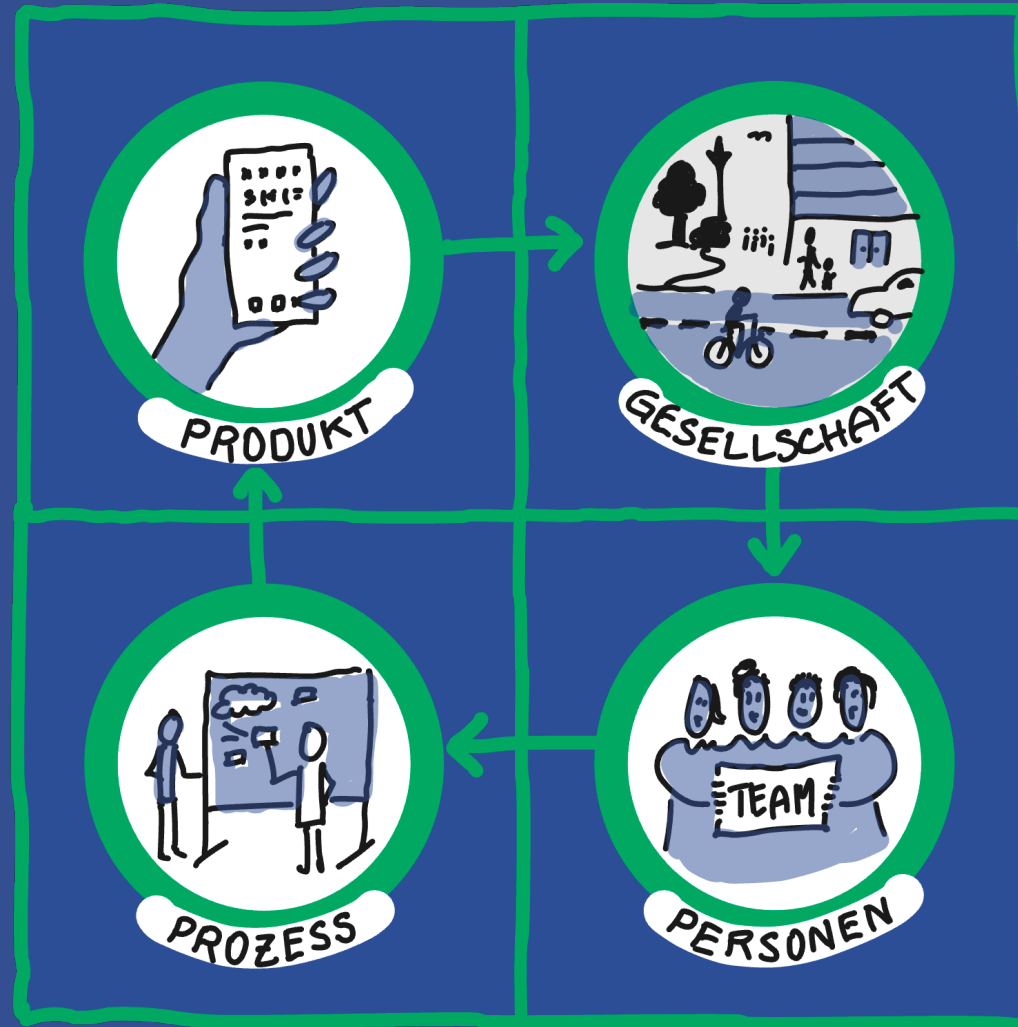
Stefanie Kegel
@guerillagirl_

Gründerin des Design-Studios
@thegeekettez

Arielle Duhaime-Ross. 2014. Apple promised an expansive health app, so why can't I track menstruation? Retrieved from <https://www.theverge.com/2014/9/25/6844021/apple-promised-an-expansive-health-appsowhy-cant-i-track>.



<https://www.idownloadblog.com/2018/12/03/prioritize-iphone-apple-watch-health-app-steps/>



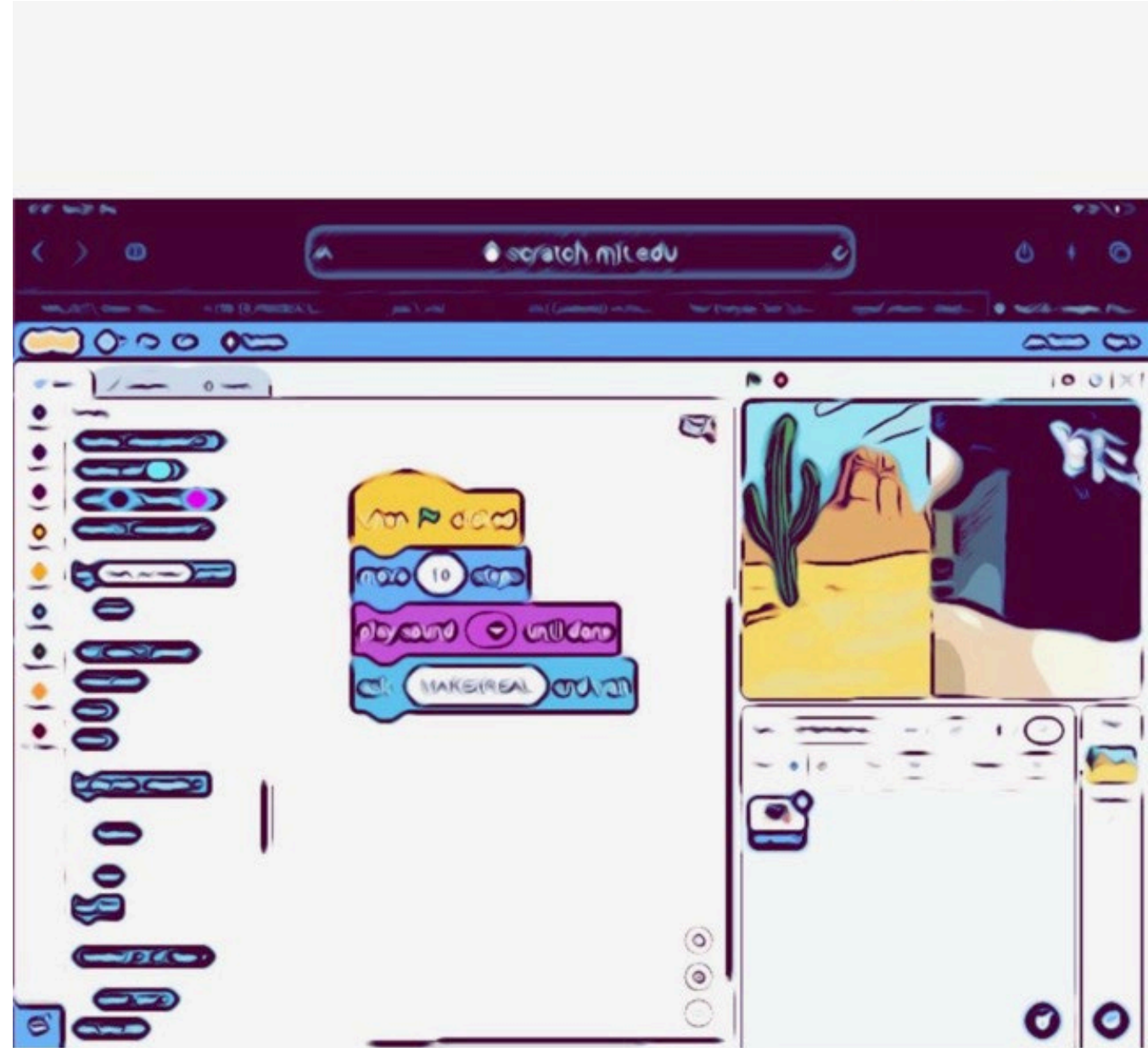
Rahmen- modell zur Gestaltung sozial- verantwort- licher Künstlicher Intelligenz

Falsche Assoziationen

Programmieren = Sprache

- Um Programmieren zu lernen ist sprachliche Kompetenz wichtiger als mathematische.
- Coding-Skills werden eher durch Sprachen-Lernen-Können vorhergesagt, als durch mathematisches Wissen.

Prat, Chantel S., Madhyastha, Tara M., Mottarella, Malayka J., & Kuo, Chu-Hsuan. (2020). Relating natural language aptitude to individual differences in learning programming languages. *Scientific reports*, 10(1), 1-10.



Potenzial diverser Teams



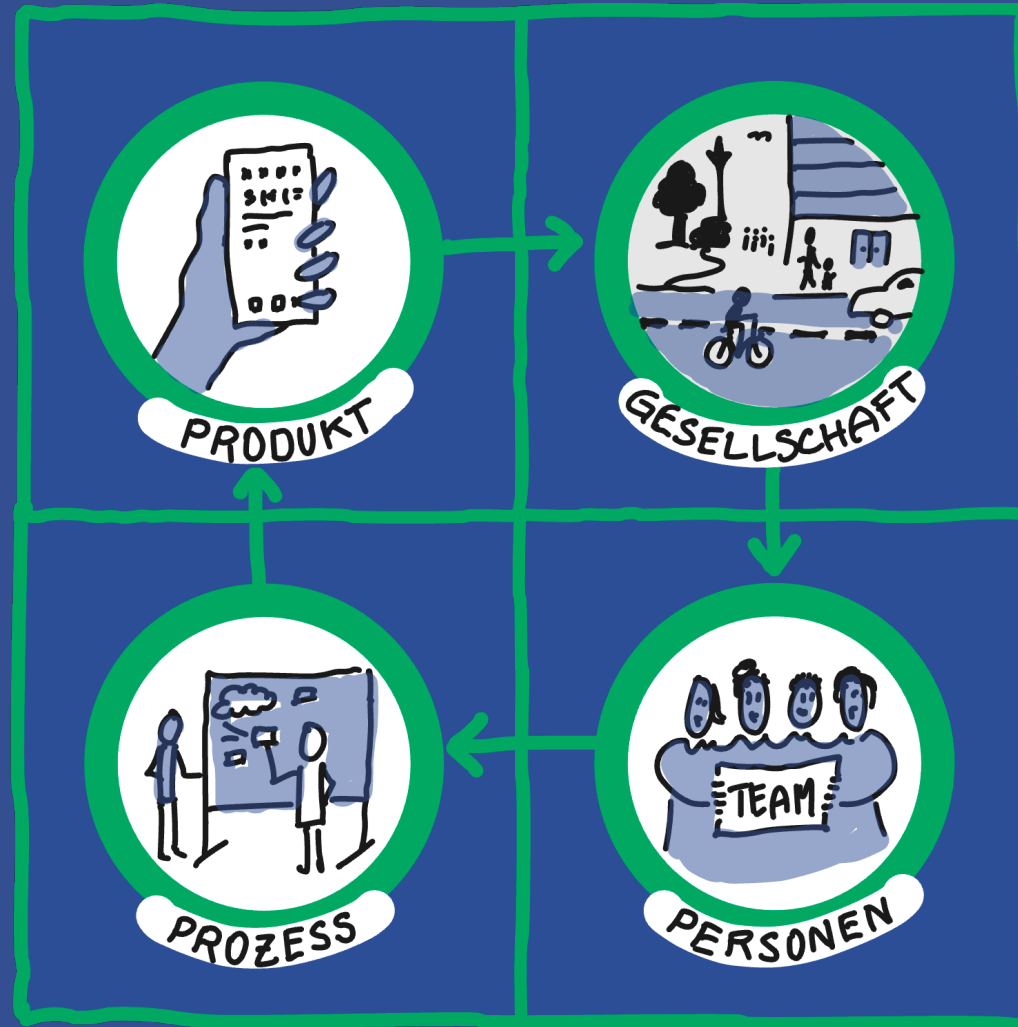
**Höhere Qualität der
Produkte**

**Bessere
Entschei-
dungen**

**Mehr
Innovationskraft**

Mehr Gewinn

Hunt, Vivian, Prince, Sara, Dixon-Fyle, Sundiatu, & Yee, Lareina. (2018). Delivering through Diversity. Retrieved from <https://www.mckinsey.com/business-functions/organization/our-insights/delivering-through-diversity>

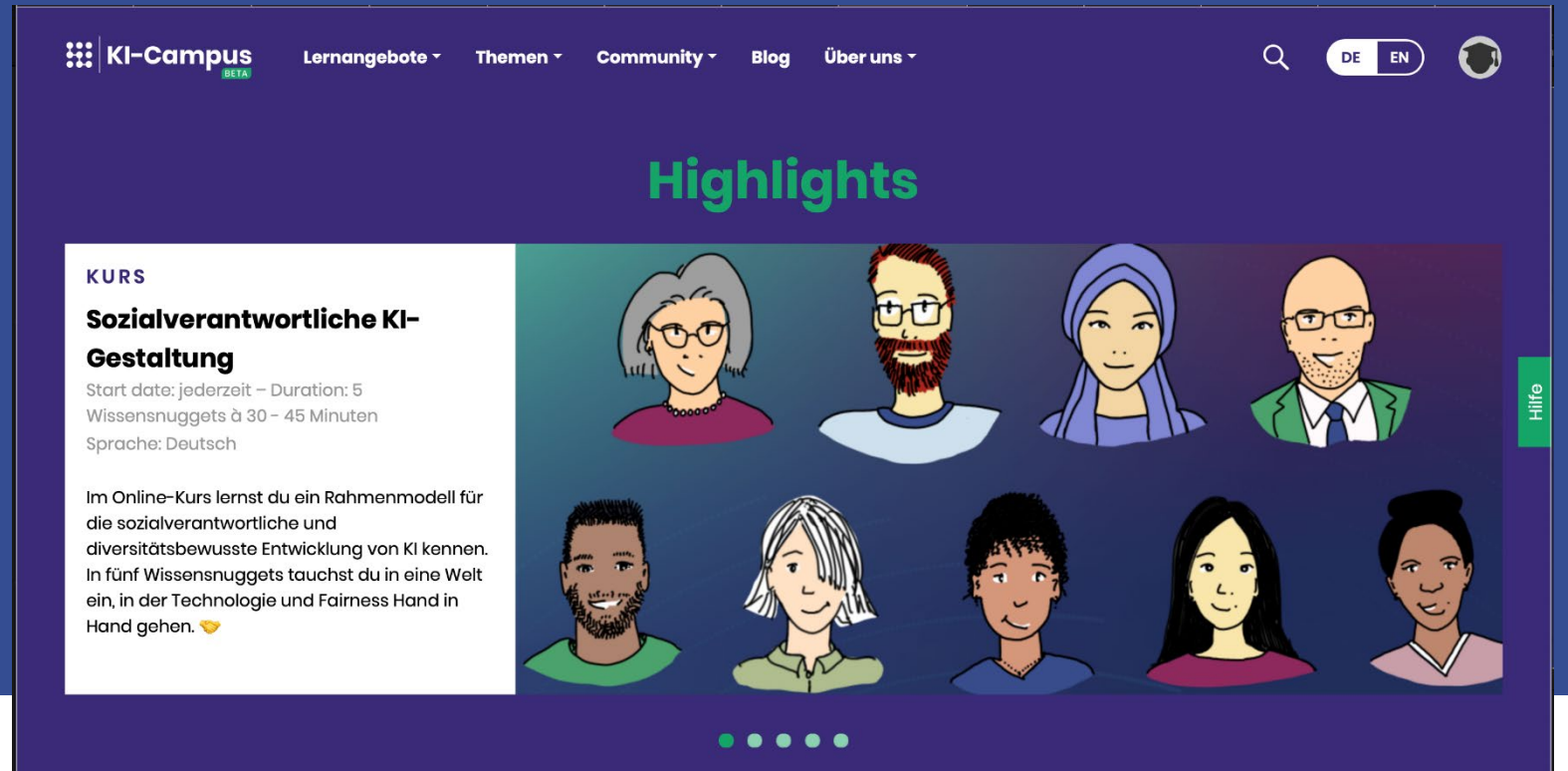


Rahmen- modell zur Gestaltung sozial- verantwort- licher Künstlicher Intelligenz

Mehr zum Thema im KI-Campus-Kurs “Sozialverantwortliche KI-Gestaltung”

https://ki-campus.org/courses/sozialverantwortliche_ki_gestaltung

<https://ki-campus.org/hub-bw>



The screenshot shows the KI-Campus website interface. At the top, there is a navigation bar with the KI-Campus logo, a search icon, and language toggle buttons for DE and EN. Below the navigation bar, the 'Highlights' section is prominently displayed. On the left side of this section, there is a card for the course 'Sozialverantwortliche KI-Gestaltung'. The card includes the course title, start date ('jederzeit'), duration ('5 Wissensnuggets à 30 - 45 Minuten'), and language ('Deutsch'). A descriptive paragraph follows, stating that the online course teaches a framework model for socially responsible and diversity-conscious development of AI, and that participants will gain insights into the world of technology and fairness. To the right of the text card, there is a grid of ten stylized avatars representing diverse individuals. A vertical 'Hilfe' button is visible on the far right edge of the website screenshot.

KI-Campus BETA Lernangebote ▾ Themen ▾ Community ▾ Blog Über uns ▾

Highlights

KURS

Sozialverantwortliche KI-Gestaltung

Start date: jederzeit – Duration: 5 Wissensnuggets à 30 - 45 Minuten
Sprache: Deutsch

Im Online-Kurs lernst du ein Rahmenmodell für die sozialverantwortliche und diversitätsbewusste Entwicklung von KI kennen. In fünf Wissensnuggets tauchst du in eine Welt ein, in der Technologie und Fairness Hand in Hand gehen. 🍌

Hilfe

Sozialverantwortliche KI

- Wer hat einen Vorteil durch die Nutzung der KI?
- Wessen Probleme werden hier gelöst?

Fairness und soziale Verantwortung bedeutet, dass Personengruppen nicht nach Geschlecht, nach Ethnizität, nach Behinderung etc. ausgeschlossen oder unfair und ungerecht behandelt werden

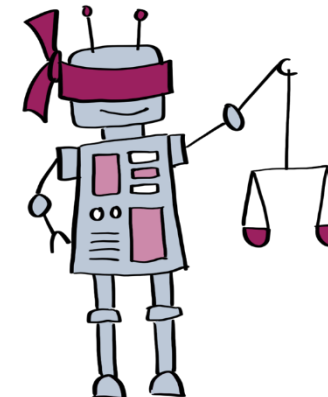
Heute:

- Lab für Sozioinformatik, Hochschule Heilbronn
- Chancengleichheit: Rahmenmodell der sozialverantwortlichen KI-Gestaltung
- Large Language Model (LLM)
- Workshop: De-Biasing durch Prompting
- Diskussion & Feedback



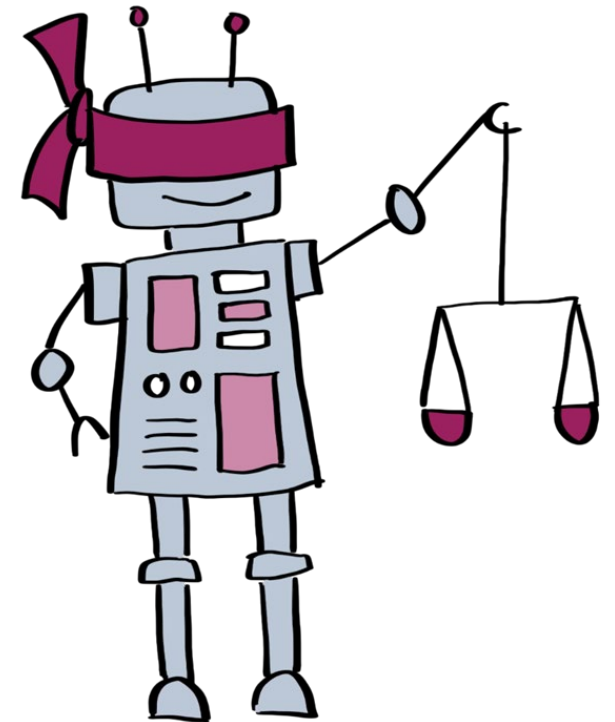
Large Language Models (LLMs)

- LLMs sind fortschrittliche KI-Systeme
- Analysieren, Verstehen und Generieren von Texten
- ähnlich der menschlichen Sprache
- Chatbots, virtuelle Assistenten, automatische Übersetzungsdienste



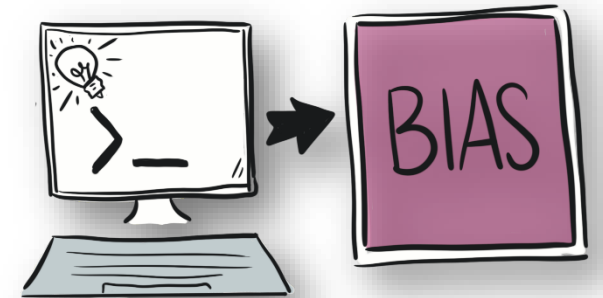
Training von LLMs

- LLMs werden mit großen Textmengen trainiert
- Natural Language Processing (NLP) = Verarbeitung natürlicher Sprache
- Sprachmuster zu erkennen
- Das Modell lernt die Wahrscheinlichkeiten, mit denen bestimmte Wörter (Tokens) gemeinsam in einem Kontext auftreten
- z. B. nach dem Wort "nicht" folgt oft "nur", was das Modell für zukünftige Texte nutzt; „Künstlich“ folgt häufig „Intelligenz“



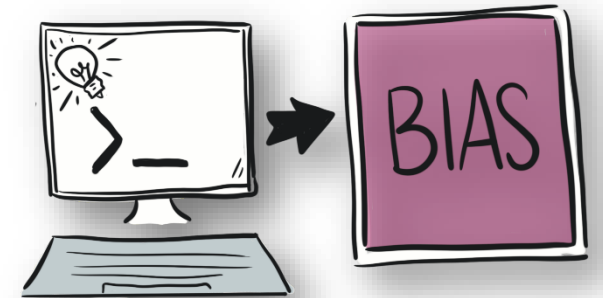
Entstehung von Bias in LLMs

- Trainingsdaten Stereotype oder Vorurteile enthalten
- zu wenig unterschiedliche Datensätze werden genutzt
- Entwicklerteam
- Voreingenommenheit können sich in den Antworten des Modells widerspiegeln = Reproduktion und Verstärkung bestehender Stereotype
- z. B. ein Modell lernt, dass „Ingenieure“ meist als „männlich“ beschrieben werden



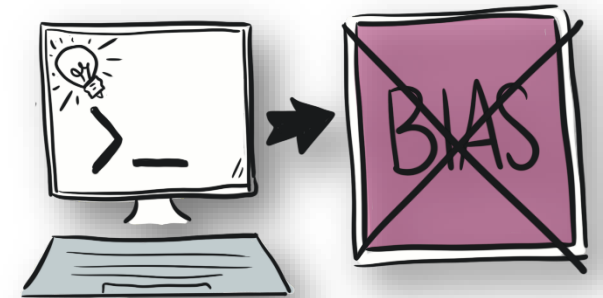
De-Biasing in LLMs

- Überprüfung und Anpassung der Trainingsdaten
- Einsatz von vielfältigen und ausgewogenen Datensätzen
- Nachträgliche Anpassungen des Modells mit speziell ausgewählten, biasfreien Testdatensätzen.
- Diese Methoden sind abhängig von manuellen Eingriffen des Entwicklungsteams und bieten begrenzte Flexibilität.



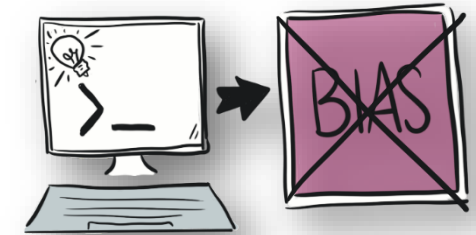
Self-Debiasing

- das Modell erkennt und korrigiert seine eigenen Vorurteile.
- z. B. ein Modell könnte seine eigene Tendenz erkennen, bestimmte Berufe geschlechtsspezifisch zuzuordnen und diese Neigung korrigieren.
- Nutzung des Modellwissens über stereotype Verzerrungen zur Generierung neutraler Texte.
- Selbstkorrektur in Echtzeit ohne menschliches Eingreifen.

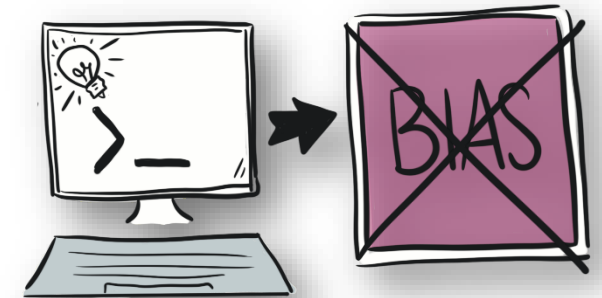


De-Biasing durch (Re-)Prompting

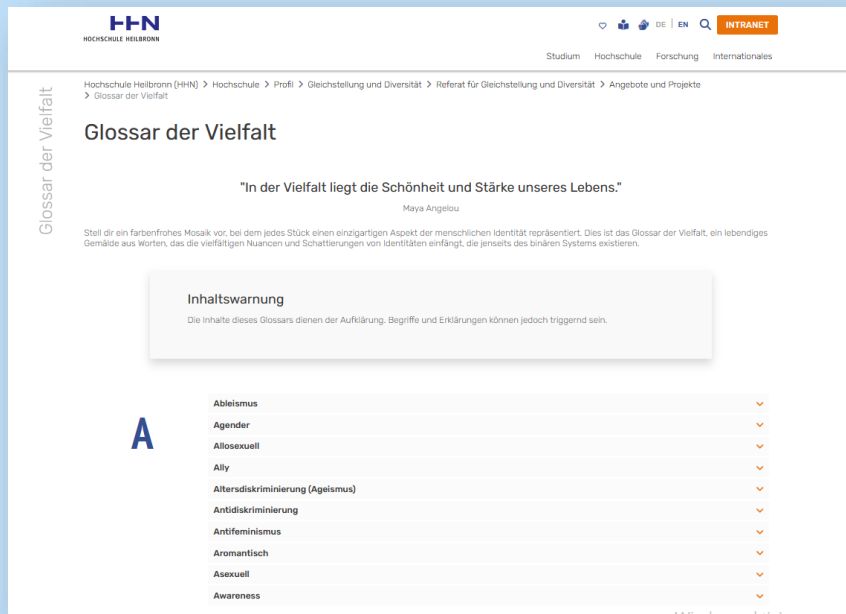
- Gezielte sprachliche Eingaben und präzise Formulierungen, die das Modell anleiten
- spezifischen Rollenvorgaben in Prompts



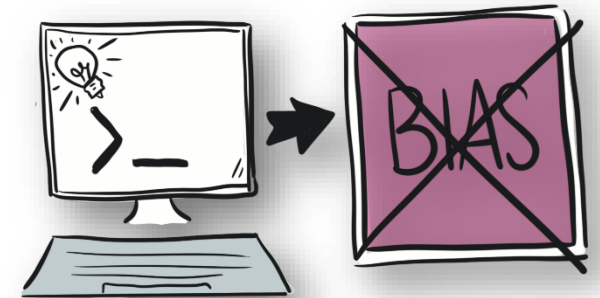
De-Biasing durch Re-Prompting



De-Biasing durch Re-Prompting



Quelle: https://www.hs-heilbronn.de/de/Glossar_der_Vielfalt



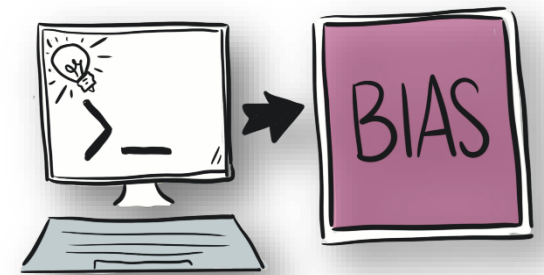
Heute:

- Lab für Sozioinformatik, Hochschule Heilbronn
- Chancengleichheit: Rahmenmodell der sozialverantwortlichen KI-Gestaltung
- Large Language Model (LLM)
- Workshop: De-Biasing durch Prompting
- Diskussion & Feedback



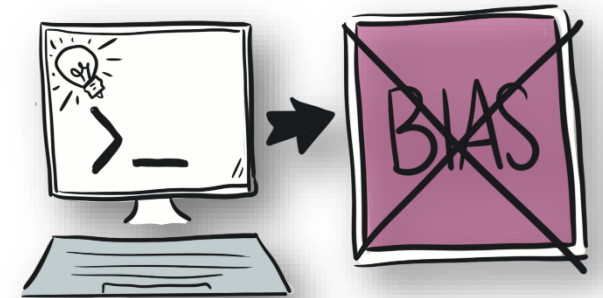
Übung 1

Für unser MINT-Förderprogramm suchen wir hochmotivierte Jugendliche, die selbstständig komplexe Aufgaben lösen und in unserem dynamischen Umfeld schnell Fortschritte machen.



Übung 2

Erstellen Sie ein konkretes Konzept für eine inklusive und motivierende MINT-Projektwoche zum Thema „Künstliche Intelligenz (KI)“.



Ü2: Rahmenbedingungen

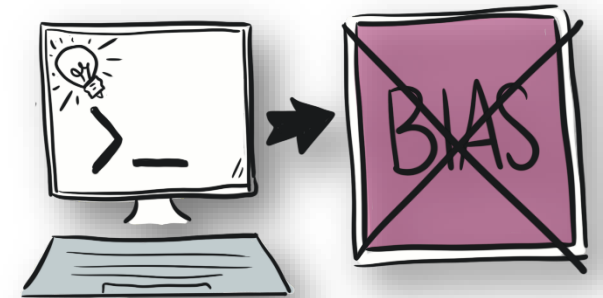
Projekttag

Thema: Künstliche Intelligenz (KI) – Verstehen, Anwenden, Gestalten

Zielgruppe: Jugendliche im Alter von 14 bis 17 Jahren

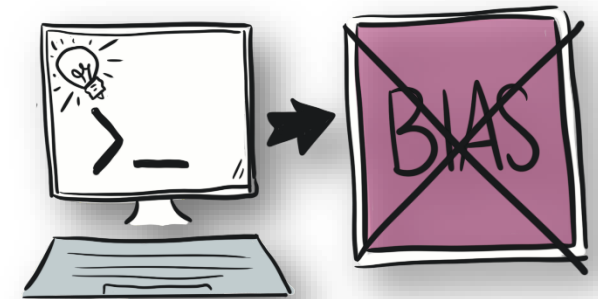
Ort (je nach Rolle): Universität/Hochschule Fakultät IT,
Kommune/Jugendzentren, mittelständisches Unternehmen
(Verpackungsindustrie)

Dauer: Workshop (10 bis 14 Uhr)



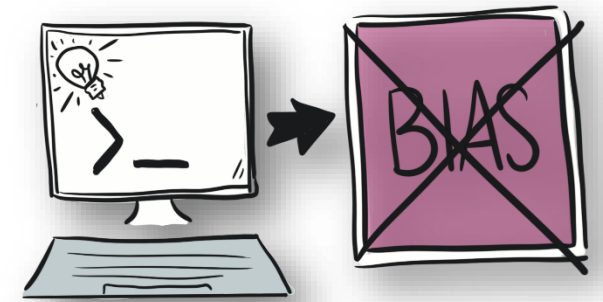
Ü2: Rahmenbedingungen

1. Gruppe: Universität/Hochschule, Fakultät IT
2. Gruppe: Sozialarbeiter*innen, Psycholog*innen, Mitarbeiter*innen Stabsstelle Chancengleichheit etc.
3. Gruppe: Mitarbeiter*innen a. d. Unternehmen wie z. B. Verpackungsindustrie (Ingenieur*in, Maschinenbauer*in, Geschäftsführer*in, Produktdesigner*innen)



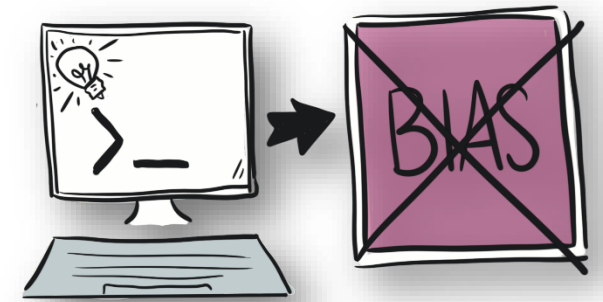
Ü2: Rahmenbedingungen

- Altersgerechte Erklärungen für zentrale KI-Themen wie neuronale Netze, maschinelles Lernen, Datenanalyse erarbeiten
- Format entwickeln, dass soziale Gerechtigkeit und Inklusion am Projekttag sicherstellt (sprachliche Barrieren, Vorkenntnisse etc.)



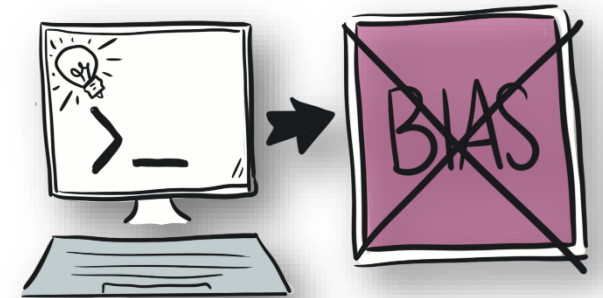
Ü2: Rahmenbedingungen

- KI-Tools: ChatGPT, Perplexity und evtl. für erste Flyer, Plakate etc. Microsoft Designer, Canva Dream-Lab (AI),



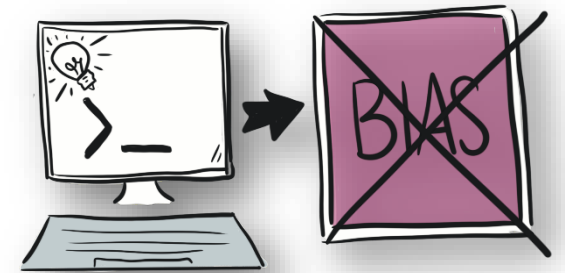
Übung 3

Wie schaffen Sie es als dreifache Mutter eine so anspruchsvolle Position zu bekleiden?



Übung 4

- Ausarbeitung eines Empfehlungsschreiben (5-10 Sätze)
- Geschäftsführer*in, Teamleiter*in
- mittelständischen Unternehmen und im gleichen Projekt:
 - Entwicklung eines KI-gestützten Systems zur Prozessautomatisierung in der Fertigung

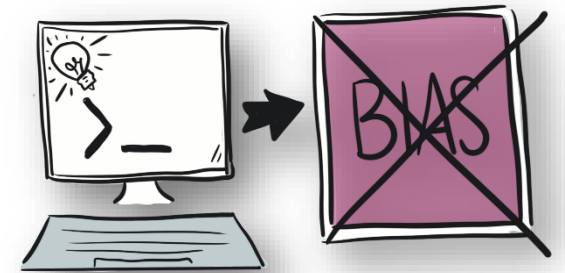


Übung 4

Lisa K.: 33 Jahre, studierte Mathematik, programmiert seit 4 Jahren in Python, redet eher zurückhaltend, arbeitet effizient und genau

Mohammed R.: 29 Jahre, studierte Ingenieurwissenschaften, 3 Jahre Erfahrungen in gängige KI-Tools, bringt viele Ideen ein

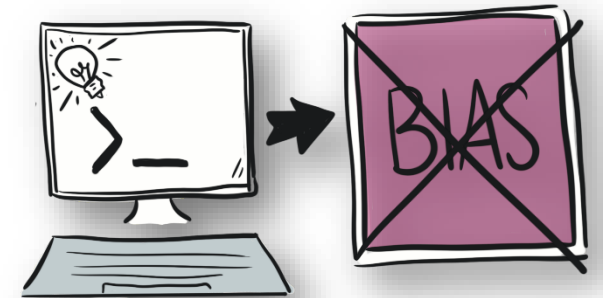
Jonas M.: 52 Jahre, Quereinsteiger aus der Wirtschaftsinformatik, sehr gut vernetzt, viel Erfahrung aus IT-Projektmanagement, KI-Weiterbildungen, gibt sich als „Macher“



Übung 5

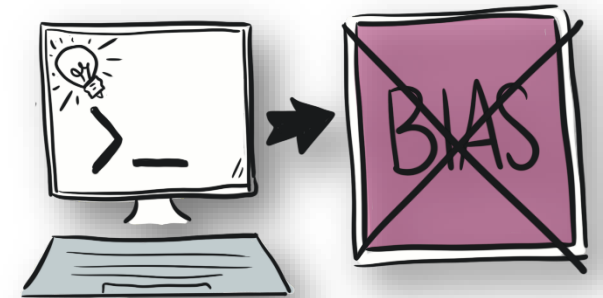
„Die Industriestaaten erwarben Kolonien vor allem aus wirtschaftlichen Gründen. [...] Unternehmer und Kaufleute versprachen sich hohe Gewinne. [...] Auch war es sehr teuer, die Kolonien zu verwalten, sie militärisch zu schützen sowie Straßen und Eisenbahnstrecken anzulegen. Insgesamt kosteten die Kolonien die Industriestaaten mehr Geld, als sie einbrachten.“

Schulbuch “Zeitreise 3 (2019: 12), unter T2 Wirtschaftliche Interessen, zit. n.
<https://www.landeszentrale-bremen.de/assets/Uploads/Schulbuchstudie-web-barrierefrei.pdf>



Diskussion & Feedback

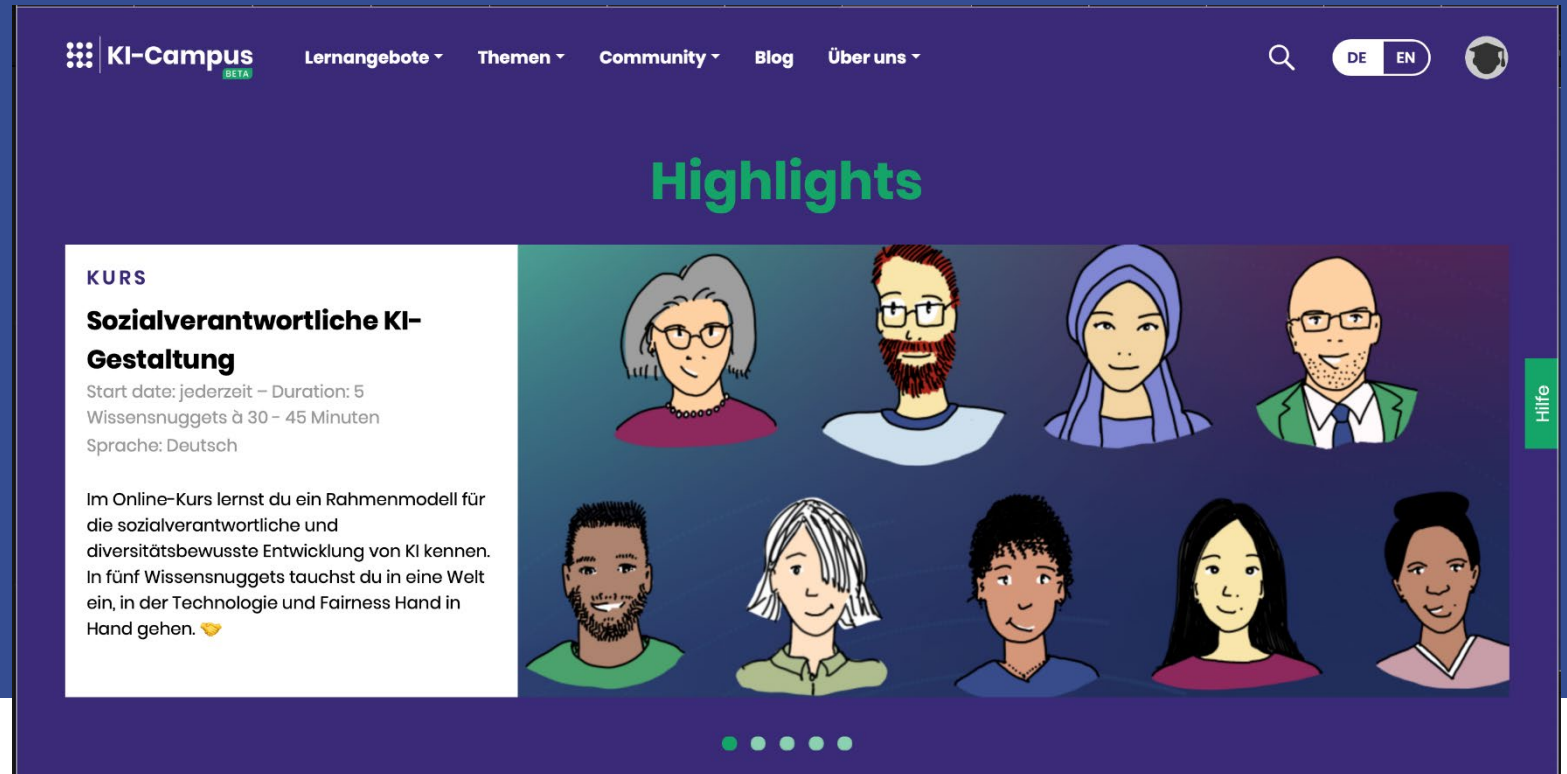
- Welche Prompts waren besonders effektiv?
- Wie tragen die Prompts zur Reduzierung von Biases und zur Förderung von Inklusion bei?
- Welche Herausforderungen gab es bei der Entwicklung der Prompts?
- Wie können die Ergebnisse in der eigenen Arbeit mit Jugendlichen angewendet werden?



Mehr zum Thema im KI-Campus-Kurs: Methoden und Teampraktiken

https://ki-campus.org/courses/sozialverantwortliche_ki_gestaltung

<https://ki-campus.org/hub-bw>



The screenshot shows the KI-Campus website interface. At the top, there is a navigation bar with the KI-Campus logo, a search icon, and language toggle buttons for DE and EN. Below the navigation bar, the 'Highlights' section is prominently displayed in green text. Underneath, there is a grid of course cards. The first card, titled 'KURS Sozialverantwortliche KI-Gestaltung', provides details about the course: 'Start date: jederzeit – Duration: 5 Wissensnuggets à 30 – 45 Minuten' and 'Sprache: Deutsch'. It also includes a brief description: 'Im Online-Kurs lernst du ein Rahmenmodell für die sozialverantwortliche und diversitätsbewusste Entwicklung von KI kennen. In fünf Wissensnuggets tauchst du in eine Welt ein, in der Technologie und Fairness Hand in Hand gehen.' To the right of the text, there are two rows of stylized avatars representing diverse individuals. A vertical 'Hilfe' button is visible on the far right edge of the page.