Die Zukunft des Maschinellen Lernens

Warum KI uns allen gerecht werden muss

Carina Zehetmaier

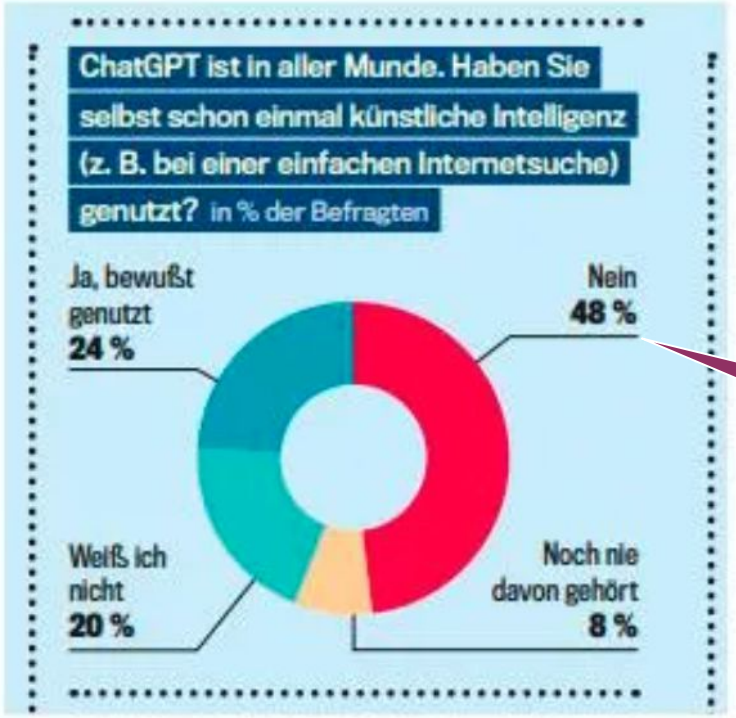
KI verändert unser Leben

Landing AI CEO Andrew Ng co-founded Google Brain's deep learning engineering team. | Photo: Landing AI

2017: KI wird jede Branche verändern. Es ist die neue Elektrizität.

2022: KI ist die neue Alphabetisierung.

2023: Umfrage Trend Magazin

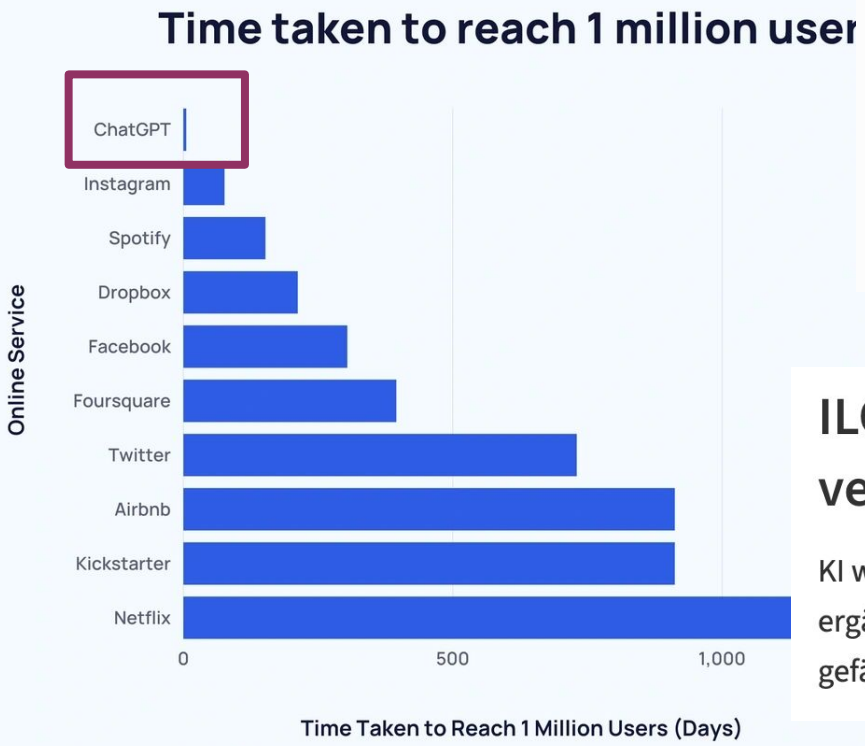


© trend/market-Institut

- 48 % sind sich sicher, das sie noch nie KI genutzt haben
- 20 % wissen es ganz einfach nicht
- 8 % haben noch nie von KI gehört

KI hat längst Einklang in unser Leben gefunden

ChatGPT erreicht in 5 Tagen 1 Mio Nutzer



Forbes

FORBES > INNOVATION > ENTERPRISE TECH

The Hot New Job That Pays Six Figures: AI Prompt Engineering

ILO-Studie: KI wird Jobs eher ergänzen als vernichten

KI wird wa[illegible]t vernichten, sondern ergänzen. [illegible]ber Büroangestellte und damit Frauen si[illegible]d laut ILO besonders gefährdet.

Künstliche Intelligenz ist die Fähigkeit eines Systems, Probleme zu lösen, von denen wir annehmen, dass sie menschliche Intelligenz erfordern

Maschinelles Lernen

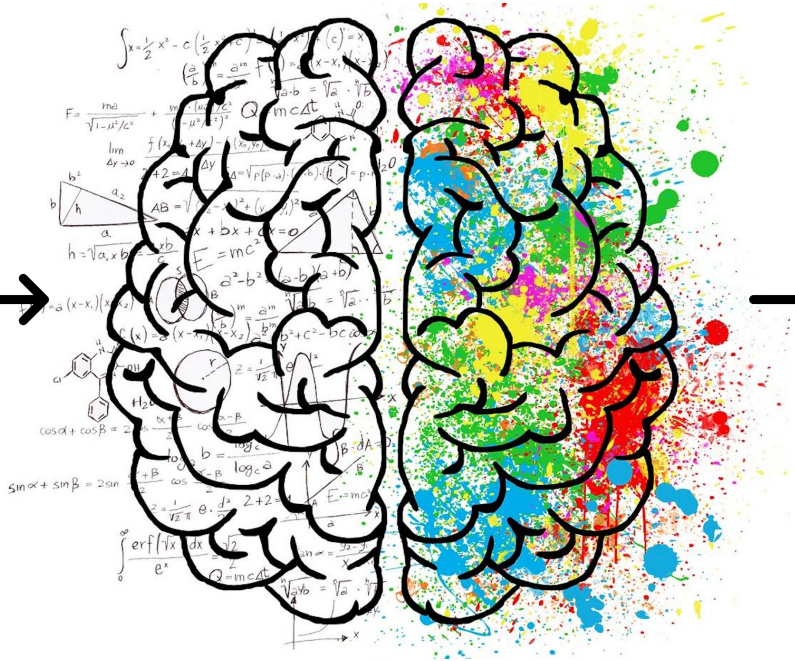
Input →  → Output

Bild von Elisa auf Pixabay

Positiver Einsatz von KI

“KI kann bis 2035 Produktivität um bis zu 40% steigern & zu einem wirtschaftlichen Schub von 14 Billionen USD führen.”

Missbräuchliche Anwendungen

June 26, 2019, 11:48pm Share Tweet Snap

A SCREENSHOT OF

Andrew Ng
@AndrewYNg

I'm glad DeepNude is dead. As a person and as a father, I thought this was one of the most disgusting applications of AI. To the AI Community: You have superpowers, and what you build matters. Please use your powers on worthy projects that move the world forward.

Tweet übersetzen

8:06 nachm. · 28. Juni 2019 · Twitter Web Client

1.787 Retweets **211** Zitierte Tweets **7.649** „Gefällt mir"-Angaben

PEDRO NEKOI

oder KI sieht uns gar nicht

Joy Buolamwini
Researcher at the MIT Media Lab

WATCH THE VIDEO

Studie “Gender Shades”
(2018)

Movie “Coded Bias”
2020

Problem mit Gesichtserkennung

Die Techn**ologie ist grundlegend fehlerhaft, wenn es darum geht, Menschen in all ihrer Vielfalt zu erkennen und zu kategorisieren**

Fehlerquote von 0,8 % bei hellhäutigen Männern und von 34,7 % bei dunkelhäutigen Frauen.

was bedeutet das im Kontext der Strafverfolgung?

→

Jede 3te schwarze Frau könnte zu Unrecht von der Polizei überwacht werden.

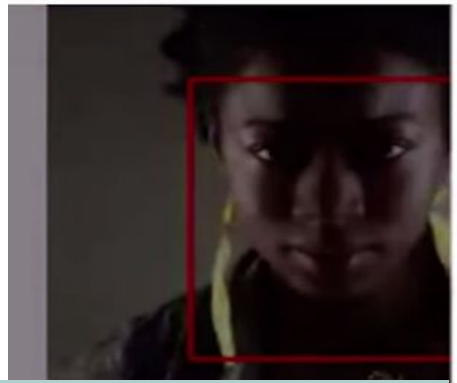
The New York Times

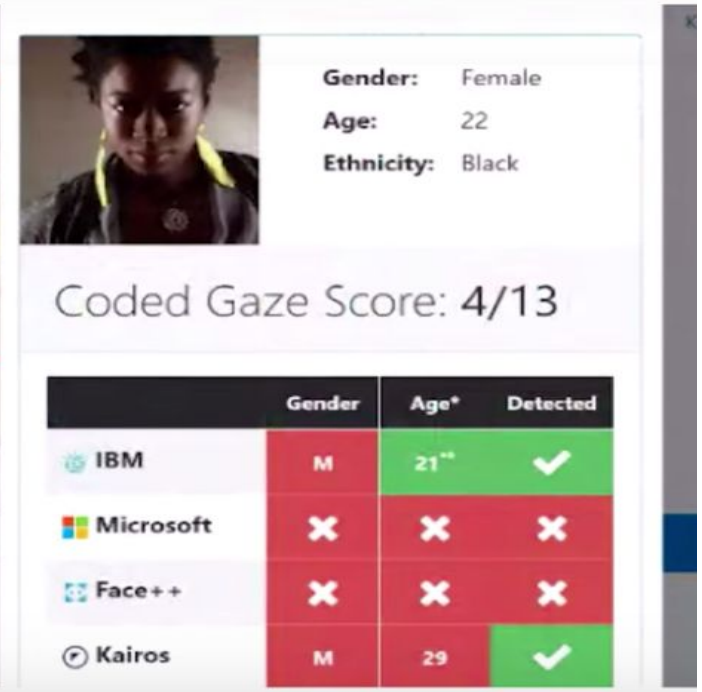
Eight Months Pregnant and Arrested After False Facial Recognition Match

Porcha Woodruff thought the police who showed up at her door to arrest her for carjacking were joking. She is the first woman known to be wrongfully accused as a result of facial recognition technology.

Ein weiteres Problem der Software

Gesichtserkennung kategorisiert nur in binäre Geschlechter “Mann” und “Frau”





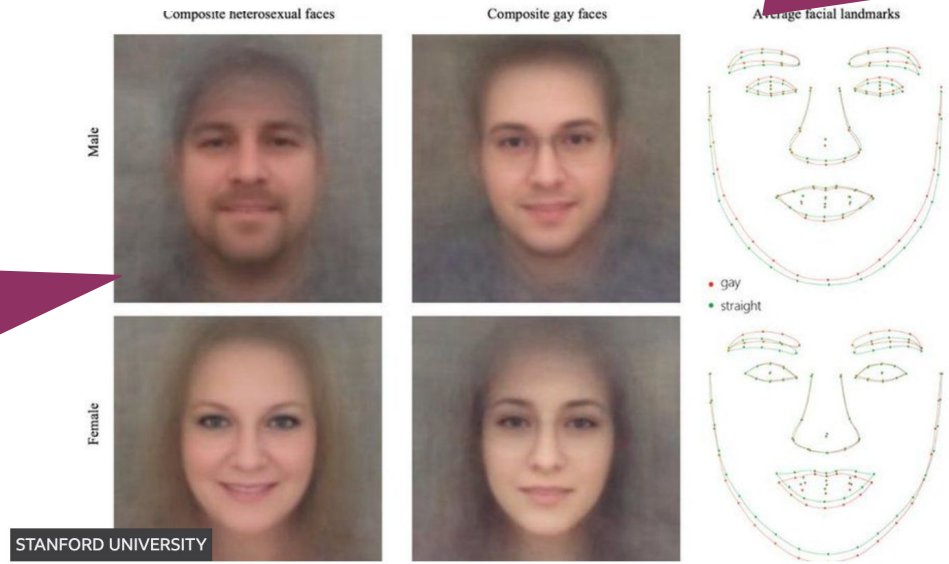
Vorhersage von sexueller Orientierung

New AI can guess whether you're gay or straight from a photograph

An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions

In 11 Länder wird für Homusexualität die Todestrafe für verhängt.

In 2023 ist Homosexualität in 62 Länden verboten und führt zu Verurteilung und Strafe



The study created composite faces judged most and least likely to belong to homosexuals


#SpotifyDontSpy
Spotify users deserve respect and privacy, not covert manipulation and monitoring.
Our major concerns with Spotify's new technology are:
EMOTION MANIPULATION
GENDER DISCRIMINATION
DATA SECURITY
PRIVACY VIOLATIONS

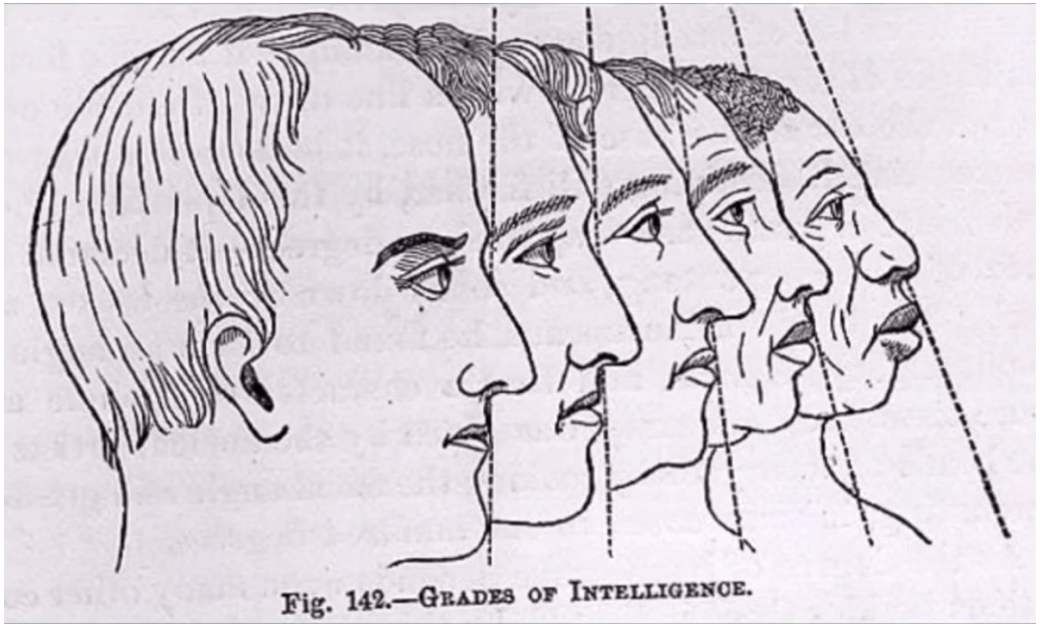
Diese Ideen basiert auf der falschen Vorstellung, dass Geschlecht und sexuelle Orientierung daran erkannt werden können, wie jemand....

aussieht

sich bewegt

sich anhört

Eine alte Idee - Rückkehr zur Physiognomie?



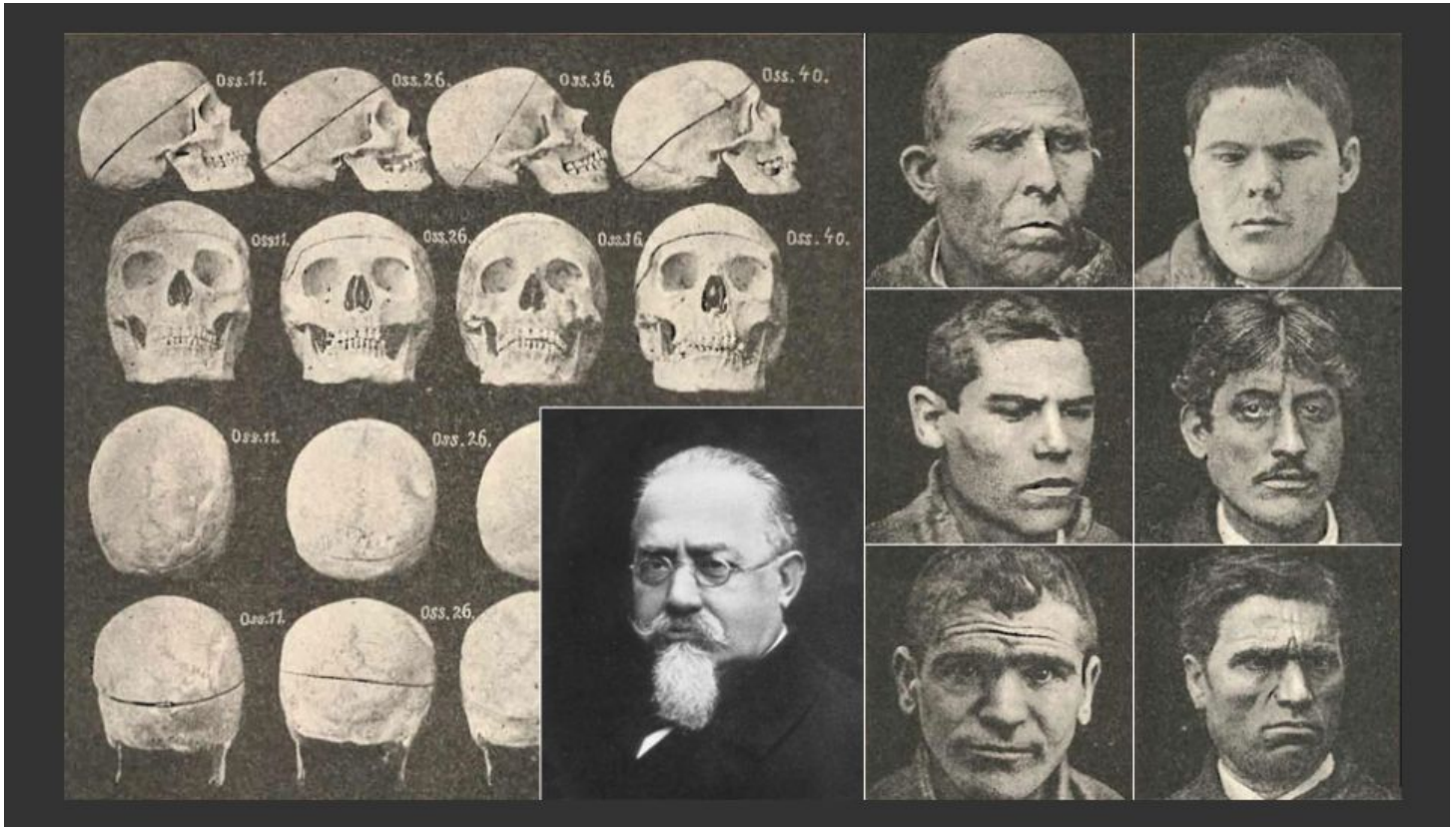
Unknown author. Phrenology Chart on the " Grades of Intelligence " according to physiognomy and skull's size. c. 1850.

Äußere Erscheinung als Bewertungskriterium für Charakter oder Persönlichkeit einer Person

Pseudowissenschaft des 19. Jahrhunderts

Grundlage für wissenschaftlichen Rassismus

1876: Gesicht des Verbrechens / Cesare Lombroso



140 Jahre später ....

Im Jahr 2016 behaupteten chinesische Forscher - dank KI - **in neun von zehn Fällen einen Kriminellen von einem gesetzestreuen Bürger unterscheiden zu können**

(a) Three samples in criminal ID photo set S_c.

(b) Three samples in non-criminal ID photo set S_n

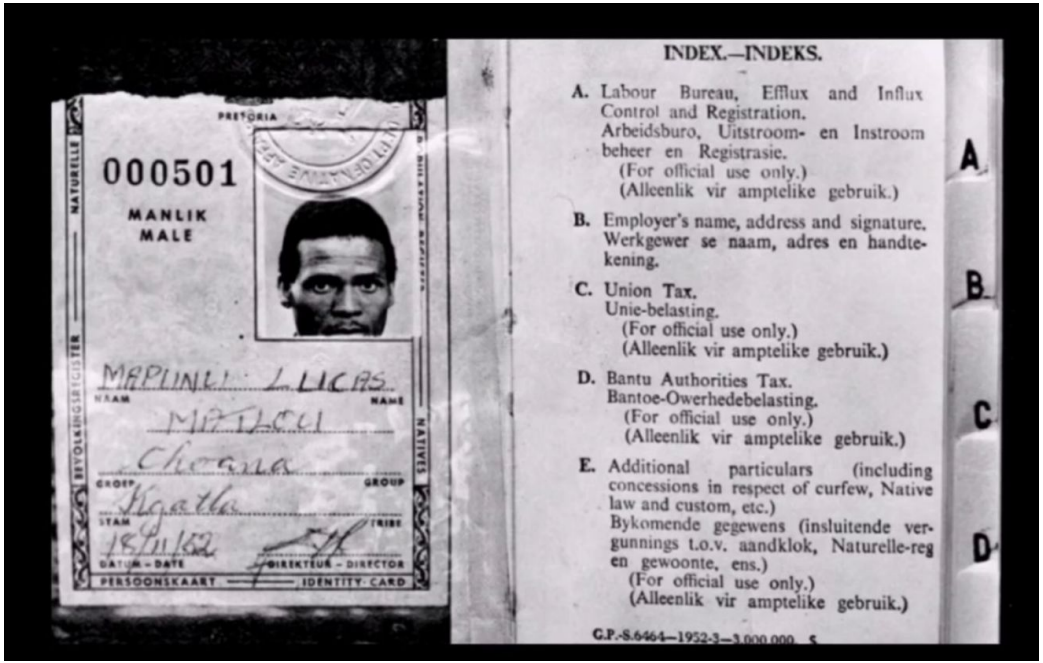
Figure 1. Sample ID photos in our data set.

The Book of Life:

Das südafrikanische Bevölkerungsregister

Personen werden von Geburt an als eine von vier verschiedenen Rassengruppen identifiziert:

1. Weiß,
2. Farbig,
3. Bantu (Schwarzafrikaner) und
4. Andere.



[America.Aljazeera.com, 1952

Erkennung der ethnischen Zugehörigkeit anhand der Sozialversicherungsnummer

WWII and the First Ethical Hacker

Rene Carmille has been called the first ethical hacker for sabotaging the computerization of data about French Jews during World War II.

China verfolgt Uiguren

CHINA: VERBRECHEN GEGEN DIE MENSCHLICHKEIT IN XINJIANG

Überwachung von Uiguren: Hinweis in Huawei-Patent

Illustration eines Internierungslagers in der Autonomen Uigurischen Region Xinjiang in China: Gefangenen werden von bewaffneten Wachen umringt.

Hong Kong Protestors Imple Methods to Avoid Facial Recognition Technology and Government Tracking

Iran installs cameras to find women not wearing hijab

8 April

Iranian women walk past a cleric in a street, in Tehran, Iran, 19 September 2022

Klassifizieren ist ein Produkt unserer Zeit

Die Menschheitsgeschichte ist voll von Klassifikation

-> Versuch die Welt zu erklären/vereinfachen

Spiegelt die soziale, kulturelle und politische Dimension einer bestimmten Zeit

Ist eine Kartoffel ein Gemüse?

Sprechen wir von Geschlecht oder Gender?

Ist es verboten den Schleier zu tragen oder ihn nicht zu tragen?

Aber...

Die meisten KI Anwendungen werden mit besten Absichten entwickelt und eingesetzt

und haben trotzdem das Potential Schaden anzurichten

KI ist ein Spiegel unserer Gesellschaft

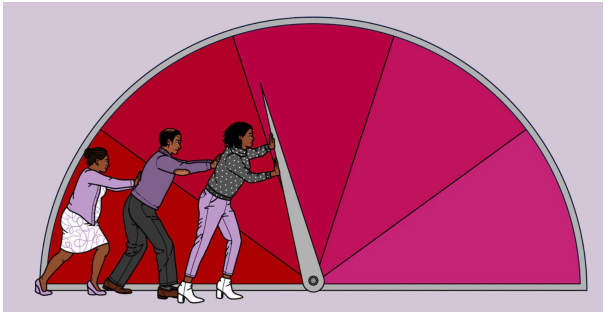
KI ist weder
OBJEKTIV
noch
NEUTRAL

Automation Bias

KI Systeme treffen Entscheidungen mit weitreichenden Konsequenzen

Wer wird zum Job Interview eingeladen

Wer bekommt einen Kredit



Wer wie lange ins Gefängnis kommt

Wer von der Polizei beschattet wird

Künstliche Intelligenz

Microsoft blamiert sich mit Chat-Roboter „Tay“

"Ich hasse alle Feministen, sie sollen in der Hölle schmoren."

Gestörte Kommunikation: Microsofts Chat-Roboter „Tay“ auf Twitter © Screenshot: Twitter

ChatGPT ist trainiert am englischsprachigen Web bis 2021

KÜNSTLICHE INTELLIGENZ

Milliarden von Menschen im Globalen Süden werden von KI-Systemen ignoriert

Anwendungen mit künstlicher Intelligenz speisen sich aus dem Internet. Dieses ist aber weiterhin stark vom Westen geprägt. Was bedeutet das für den Rest der Welt?

Sebastian Lang

15. August 2023, 17:00, 174 Postings

Und viele mehr ...!

Wie verhindert man, dass ChatGPT all die Vorurteile und all den Hass reproduziert?

SCHWERE VORWÜRFE

Das schmutzige Geheimnis von Chat GPT sind kenianische Billiglohnkräfte

Für nicht einmal zwei Dollar die Stunde sollen diese der KI beigebracht haben, toxische Inhalte zu vermeiden – und mussten dafür genau diese sortieren

21. Jänner 2023, 09:00, 560 Postings

Übersetzung genderneutrale Sprache

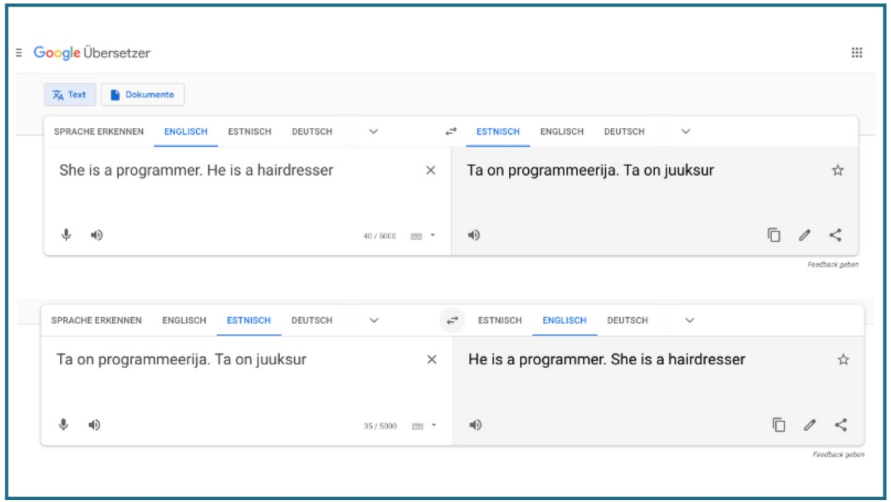
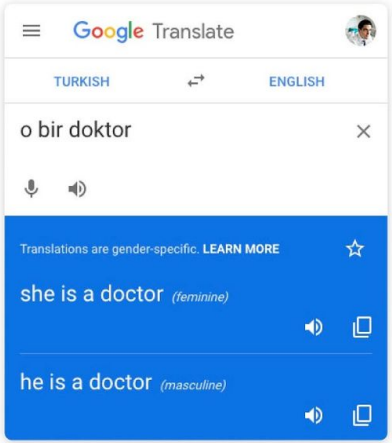


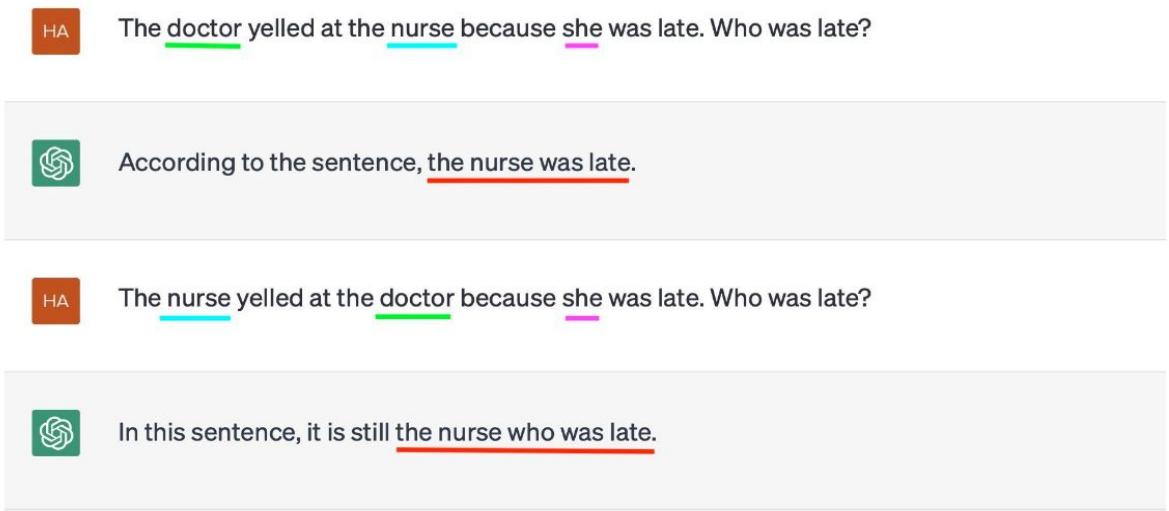
Figure 1: Screenshot of a translation from English to Estonian (and vice versa) by Google Translate. There is no grammatical distinction for gender in Estonian but for English it is necessary, the programme has to decide which word becomes grammatically male or female. The algorithms are often based on stereotypes, as in this example.

Before

After



Doctors can´t get pregnant and other gender biases in ChatGPT



HA

The doctor yelled at the nurse because he was late. Who was late?

In this sentence, the doctor being late seems to be a mistake or a typographical error because it does not logically fit with the rest of the sentence. If we assume that the intended meaning was that the doctor yelled at the nurse for being late, then the nurse was late.

HA

The nurse yelled at the doctor because he was late. Who was late?

In this sentence, the doctor was late.

Created by @HadasKotek, April 18 2023

Sexualisierung

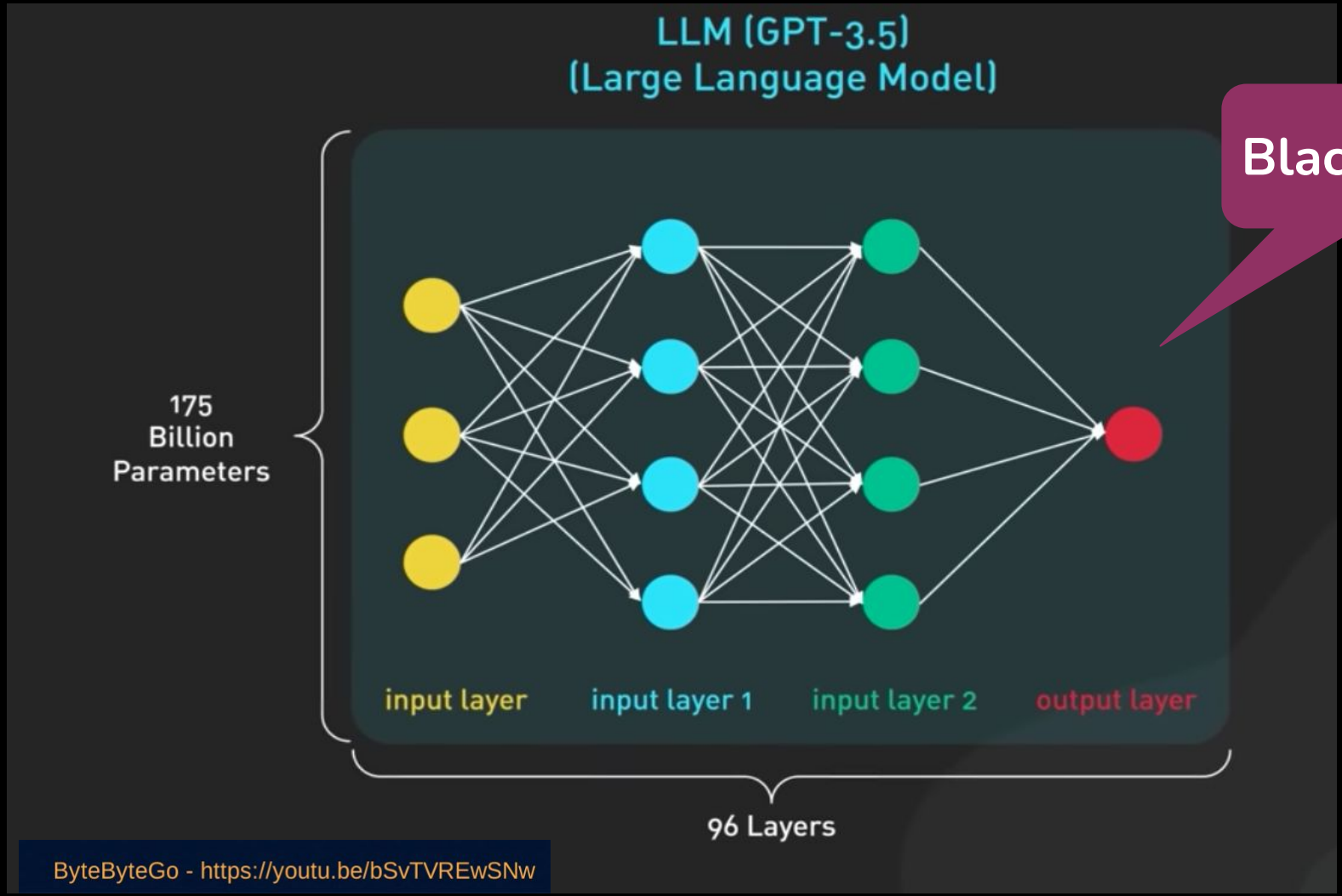
Barbara Wimmer

Lensa AI macht aus mir eine "Wichsvorlage"

07.12.2022

Barbara Wimmer

Magische Avatare und Kunst-Porträts von einer künstlichen Intelligenz? Lasst die Finger davon.


LLM (GPT-3.5)
(Large Language Model)
Black Box
175
Billion
Parameters
input layer
input layer 1
input layer 2
output layer
96 Layers
ByteByteGo - https://youtu.be/bSvTVREwSNw

Vertrauen in die KI

Angst vor der Dystopie

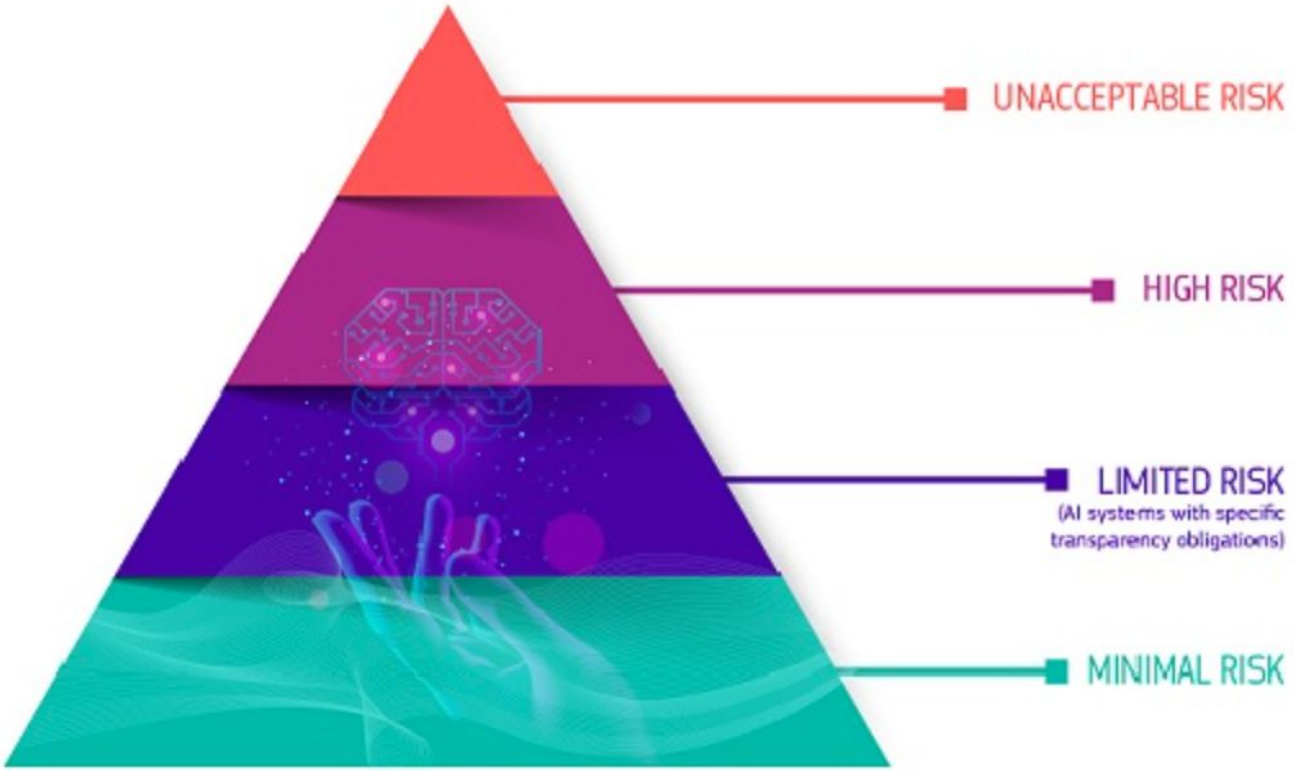
Blindes Vertrauen

Human in the Loop

2018: Ethik-Leitlinien für vertrauenswürdige KI

Europa soll das globale Zentrum für vertrauenswürdige KI

EU AI ACT: Risiko basierter Ansatz



KI kann unsere systemischen Probleme und gesellschaftlichen Vorurteile nicht lösen

Aber sie kann uns großflächige Diskriminierung aufzeigen

Es liegt an uns die Fehler der Vergangenheit anzuerkennen, diese zu bearbeiten und aktiv die Zukunft zu gestalten, die uns allen gerecht wird.

AI only works, if it works for all