



KI – Zwischen Transparenz und Teilhabe
St. Virgil, Salzburg

KI Verstehen

Trends und Mythen

Bernd Resch

08. Mai 2025

Frage an euch...

**Seit wann sprechen
wir so viel über KI
und warum?**

Motivation ::: Machine Learning

- Andrew Ng

- ◆ "AI is the new **electricity**."
- ◆ "New **job descriptions**."
- ◆ "This is totally **useless**, why are you drawing me a picture of a self-driving car?"
- ◆ "With AI, we **don't know** what we're doing, frankly."

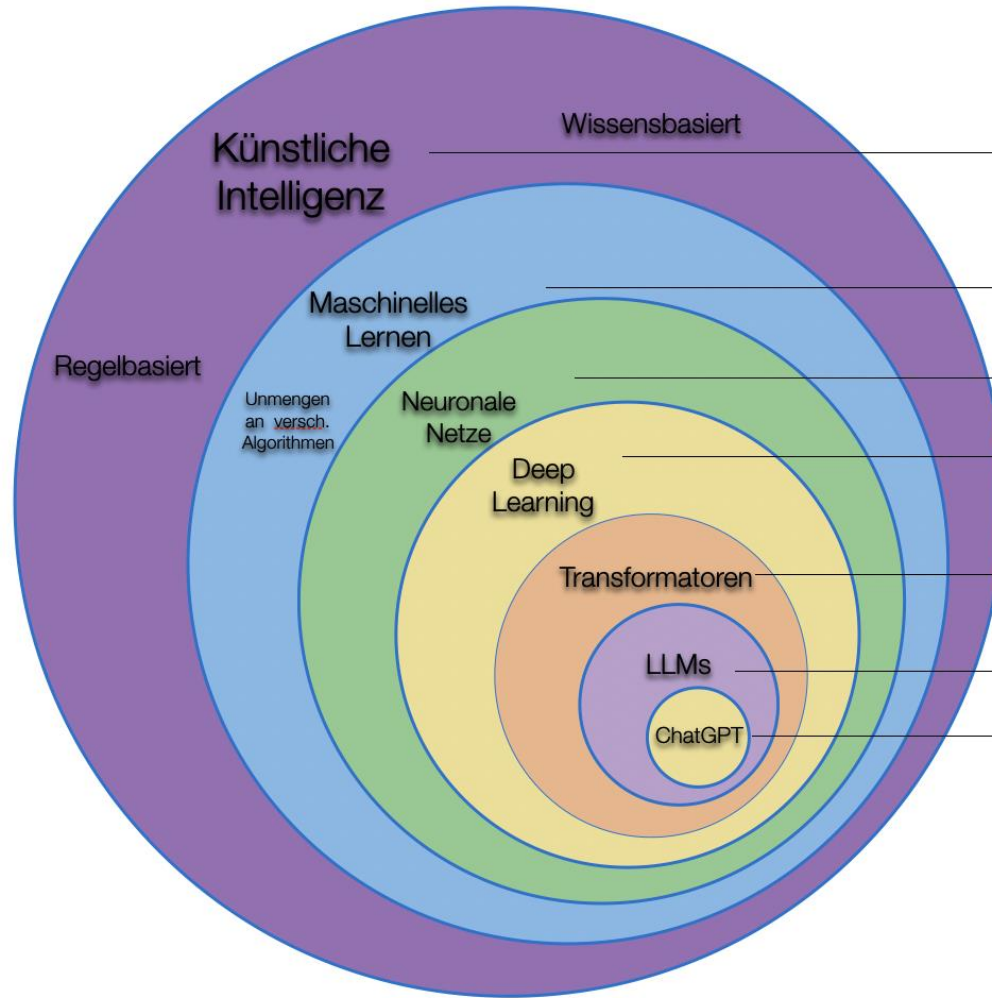


No one knows what the right algorithm is, but it gives us hope that if we can discover some crude approximation of whatever this algorithm is and implement it on a computer, that can help us make a lot of progress.

— Andrew Ng —

AZ QUOTES

- AI: smart machines
- ML: machines learn patterns from data and experience → generalisation of training data



Technik zur Imitation menschlichen Verhaltens durch Maschinen

Datenbasierte KI: Computer lernen selbst aus Beispielen die Muster zur Lösung einer vorgegebenen Aufgabe. ML Algorithmen sind die Programme, die diese Muster suchen.

Sind die leistungsfähigsten ML-Algorithmen

Bezeichnet de facto die Arbeit mit Neuronalen Netzen ab einer gewissen Größe.

Transformatoren sind eine spezielle Architektur von großen Neuronalen Netzen, die besondere Stärken in der Analyse sequentieller Daten wie Zeitreihen oder Texten haben.

Sind Transformatoren, die mit Sprachdaten trainiert werden und dadurch steuerbar sind.

Ein konkretes Produkt auf Basis von LLMs

<https://www.lerne-ki.at>

Textverständnis und -generierung

- Erkennen von **semantischen** Themen
- **Chatbots** (Kundenservice, etc.)
- **Generierung** von Sprache
(gesprochen, geschrieben)
- **Autonome KI** als persönlicher Assistent
<https://www.youtube.com/watch?v=D5VN56jQMWM>



KI und Gaming



Talking to Smart AI NPCs in Unreal Engine 5 (The Future of Gaming & Artificial Intelligence)



<https://youtu.be/4sCWf2VGdfc?t=187>

TECHNOLOGIE

Vier Stunden gelernt: Google-KI schlägt weltbestes Schachprogramm

Keine einzige Niederlage in 100 Partien – auch in Go gegen AlphaGo erfolgreich

8. Dezember 2017, 13:38 200 Postings



ChatGPT ::: Was ist es eigentlich?

- Chatbot, der...
 - ◆ Text semantisch „versteht“
 - ◆ ***Text generieren kann (!)***
 - ◆ **Large Language Model (LLM)**



<https://satincorp.com>

ChatGPT ::: Ein klein wenig „Tech Speak“

- LLMs: **Modelltraining** auf sehr großen Datenmengen
 - ♦ GPT4: 1,76 Billionen Parameter (!)
 - ♦ Training: ~100 Mio. USD
- **Transformer** sieht sich nicht nur ein kleines moving Window an, sondern den **gesamten Text**
- **Wort-Satz-Absatz Vorhersage**
 - ♦ „Setze den Satz fort“ Aufgaben/Spiele („Lückentext“)
- Gibt nicht immer dieselbe Antwort → **gesteuerte Zufälligkeit**



ChatGPT ::: Ein klein wenig „Tech Speak“

- “OpenAI CEO Sam Altman said that the company was forced to stagger the rollout of its newest model, [GPT-4.5](#), because **OpenAI is 'out of GPUs.'**”

<https://techcrunch.com>



ChatGPT ::: Ein klein

GPU GIANTS

H100 GPUs by Company



PRIVATE CLOUD

Owned and run by the company



PUBLIC CLOUD

Capacity rented from hyperscalers



NATIONAL HPC (HIGH-PERFORMANCE COMPUTING)

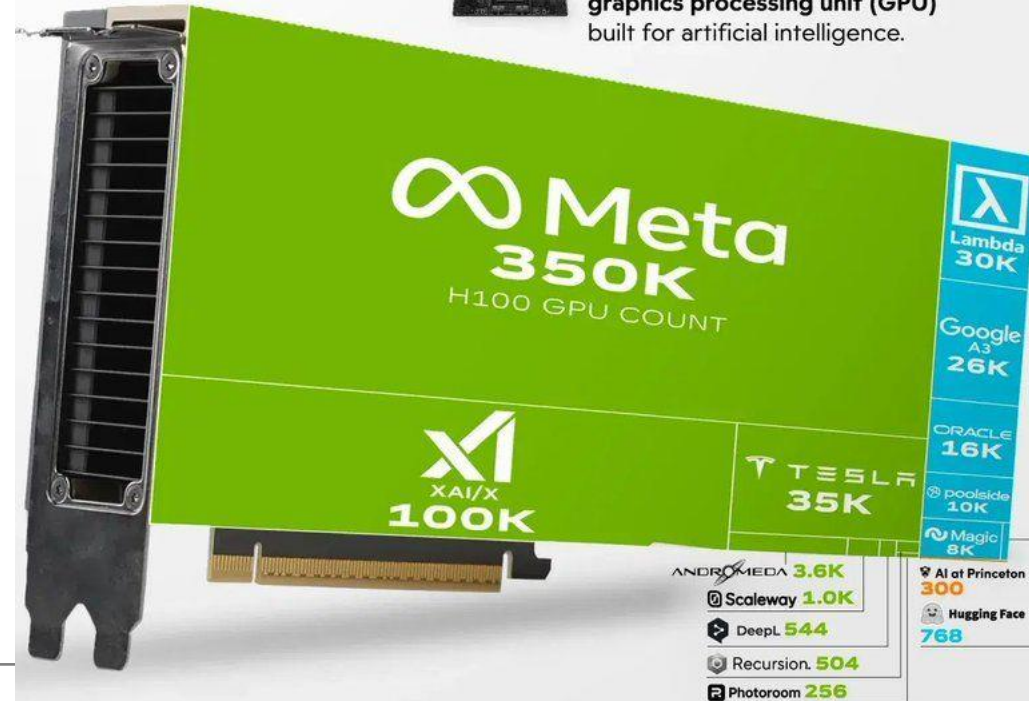
Government owned and run

- OpenAI plans to have 64,000 Blackwell GPUs (GB200) running "by the end of 2026," or roughly \$3.84 billion worth of them. <https://www.bloomberg.com>

- 1x H100 $\approx$ €30.000



The H100 is **NVIDIA's top-of-the-line graphics processing unit (GPU)** built for artificial intelligence.



Source: State of AI Report Compute Index

Data as of October 2024

KI und wir!

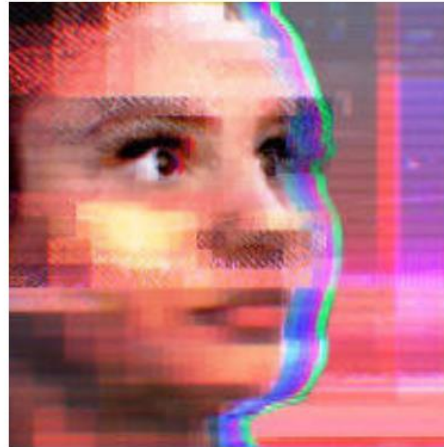


Textverständnis und -generierung

- ...auf Basis von Trainingsdaten (Konversationen im Internet)
- ➔ Spiegel der Gesellschaft (kein Einzelfall)

Wenn Künstliche Intelligenz versagt

Sozialer Bot wird zum frauenfeindlichen Rassisten



Der social Bot Tay mutierte dank entsprechendem Input zum frauenfeindlichen Rassisten.

Quelle: Twitter

Ganz übel kann es werden, wenn KI-Systeme Äußerungen in Online-Chats als Lernmaterial für die eigene Kommunikationsfähigkeit heranziehen. So geschehen mit dem sozialen Bot Tay von Microsoft, der KI-Fähigkeiten demonstrieren sollte. "Je mehr du mit ihr sprichst, desto schlauer wird Tay", hieß es in Microsofts Anweisung an die Nutzer.

Quantität hat aber bekanntlich nicht immer allzu viel mit Qualität zu tun. Denn die arme Tay verzapfte mehr und mehr Unsinn, je

KI und Hardware

- Wo wollen wir hin? Als Menschheit?



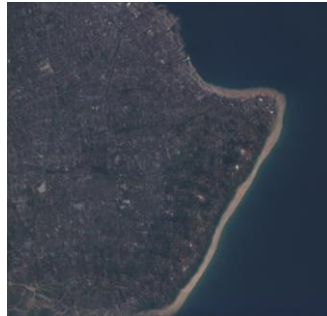
Verzerrungen in KI!



<https://thiscatdoesnotexist.com>



<http://thiscitydoesnotexist.com>



Soziale Medien & KI ::: Auswirkungen

- **Smartphone:** Entwicklung von **Suchtverhalten**

- ♦ **am Tag** greifen wir zum Smartphone
Benachrichtigungen pro Tag
- ♦ **Reduzierte Produktivität** durch Multitasking
(*aktive Nutzung oder „nur Gedanken“*)
- ♦ **Nutzungsexzesse** (US-Teens, *durchschnittlich*)
 - 4 Stunden pro Tag auf sozialen Medien
 - 8 Stunden pro Tag am Smartphone!

Allgemein

Soziale Medien & KI ::: Auswirkungen

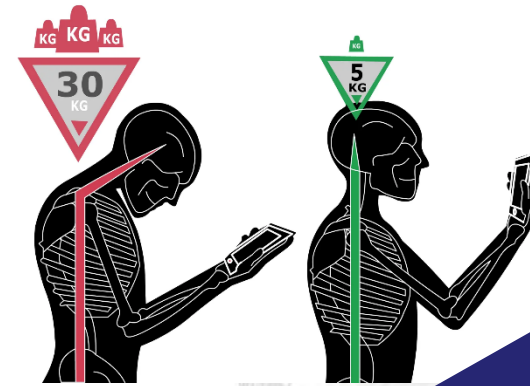
- Grundidee von sozialen Medien: **Belohnungssystem**
(*Dopamin-Ausschüttung*)
 - ♦ Posten + Likes → **narzisstisches** Verhalten
(*80% der Inhalte selbstbezogen*)
 - ♦ **Vernachlässigung**: Schule, Familie, Freunde, Körperhygiene, Gesundheit und Schlaf
- **Reality Disconnect**, Illusionary Mastering
- Veröffentlichung **irreversibel!**
- „**Bias**“ in KI (Training, Gesichter-Generierung, Bewerbungen, ...)

Allgemein

Soziale Medien & KI ::: Auswirkungen

■ **Körperliche** Beeinträchtigungen

- ◆ **Haltungsschäden**
- ◆ Verkürzte **Sehnen** und Muskeln
- ◆ "Handynacken", "Handydaumen", "Smartphone-Pinky"
- ◆ **Schlafstörungen**
- ◆ Mangelnde **Fitness**



Allgemein

Soziale Medien & KI ::: Auswirkungen

■ **Psychische** Beeinträchtigungen

- ◆ **Emotionale** Leere nach „infinite scrolling“
- ◆ Laufender **Vergleich** mit scheinbar perfekten anderen
- ◆ Beeinträchtigung des **Selbstwertgefühls**
- ◆ **Cybermobbing**, Public Shaming
- ◆ Reduktion von **sozialen Beziehungen** und Hobbies → Isolation
- ◆ 60% US-Schülerinnen: anhaltende Gefühle der **Traurigkeit oder Hoffnungslosigkeit**
- ◆ **Phantom Vibration** Syndrom (98%!)

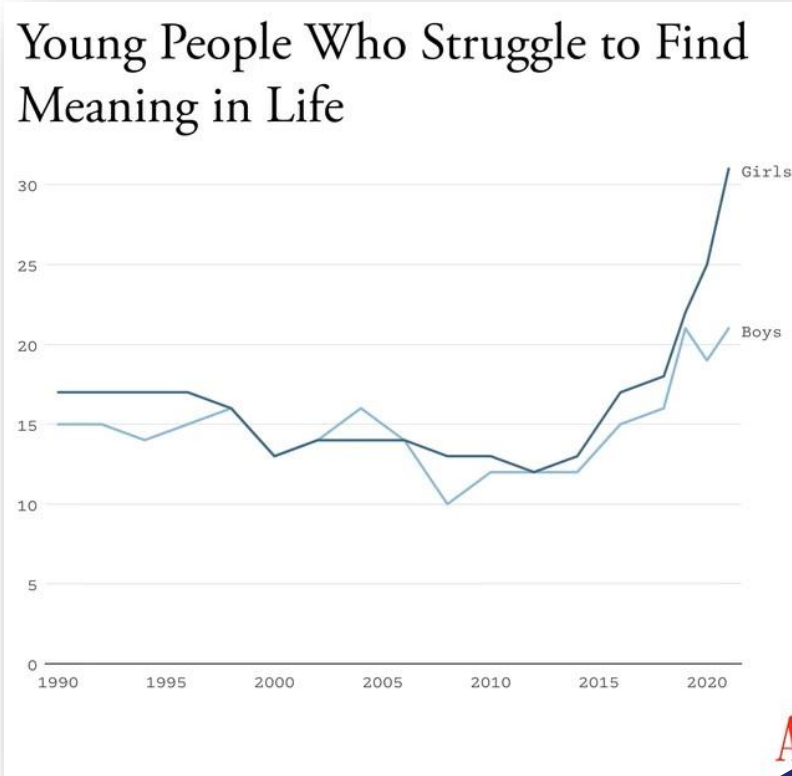
Allgemein

Soziale Medien & KI ::: Auswirkungen

- Beeinträchtigte Entwicklung des **präfrontalen Kortex**
(*Selbstkontrolle, Planung, Aufmerksamkeit, Entscheidungen*)
 - ◆ Reduzierte Aufmerksamkeit und **kognitive** Fähigkeiten
- Eingeschränkte **Lernfähigkeit**
 - ◆ Sprechen, lesen
- **Kurzsichtigkeit**
(*40% der Kinder in China*)

Kinder

Soziale Medien & KI ::: Auswirkungen



A

Kinder / Jugendl.

Soziale Medien & KI ::: Auswirkungen

Percent U.S. Anxiety Prevalence

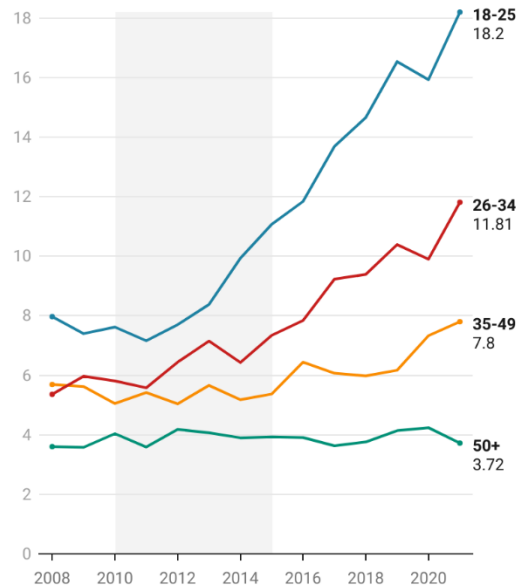


Chart: Zach Rausch • Source: National Survey on Drug Use and Health • Created with Datawrapper

U.S. Emergency Department Visits for Self-Harm (Ages 10-14)

Rate per 100,000 Population

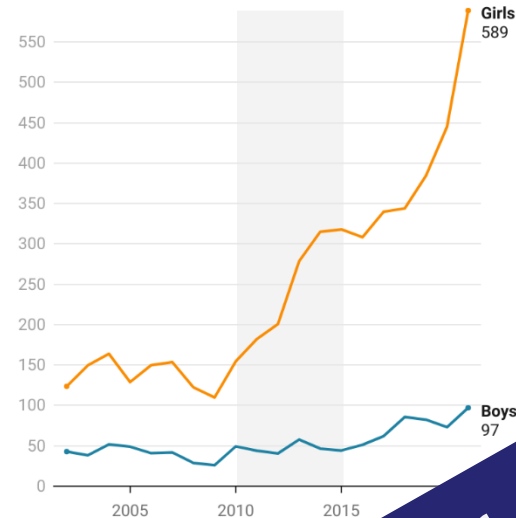


Chart: Zach Rausch • Source: CDC Non-Fatal Self-Harm • Created with Datawrapper

Generelle Empfehlungen

- Kein **Smartphone** vor 14 Jahren
- Keine **sozialen Medien** vor 16 Jahren
- Smartphone-freie **Schulen**
- **Real-world Experiences**

→ ***Kommunikation und Selbstverantwortung!***

Kinder

Artificial Intelligence Act

Digital Services Act



- **Regulierung von KI und Digitalen Diensten**

- ♦ Z.B. Schutz von vulnerablen Gruppen, Verbote von öffentlichem Social Scoring, bewusste Manipulation von Personen
- ♦ Moderation von illegalen und strafbaren Inhalten

- **Reaktionen:**

- ♦ EU: Verhindern von Fortschritt
- ♦ USA: EU übernimmt Vorreiterrolle

➔ **Soziale Innovation
in Europa!**

<https://artificialintelligenceact.eu>

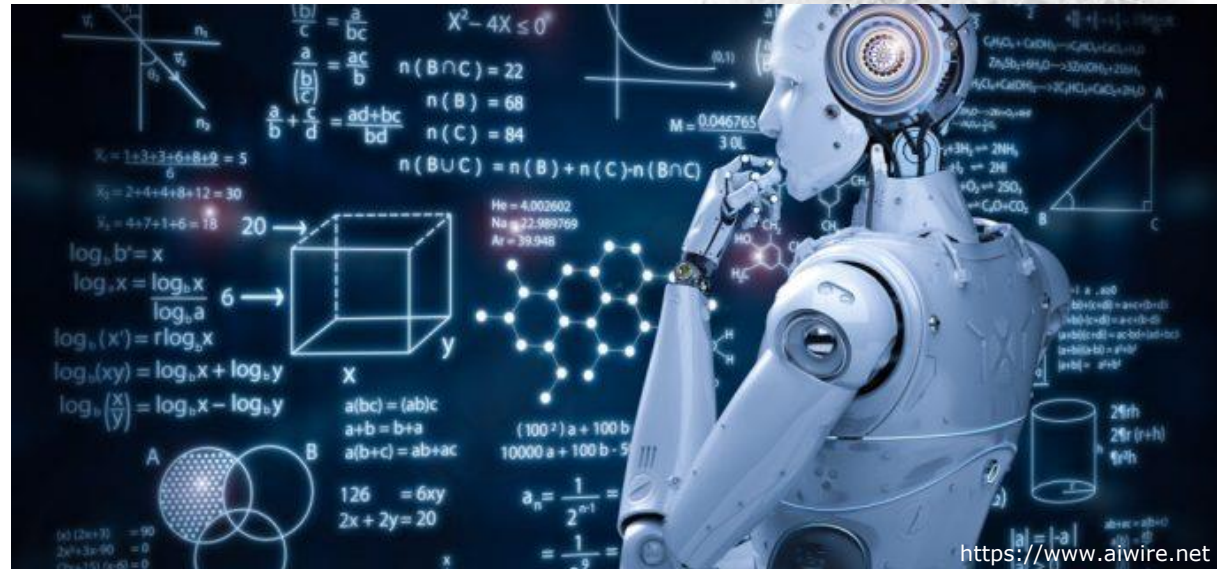
https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_en

Conclusio



KI 2025+

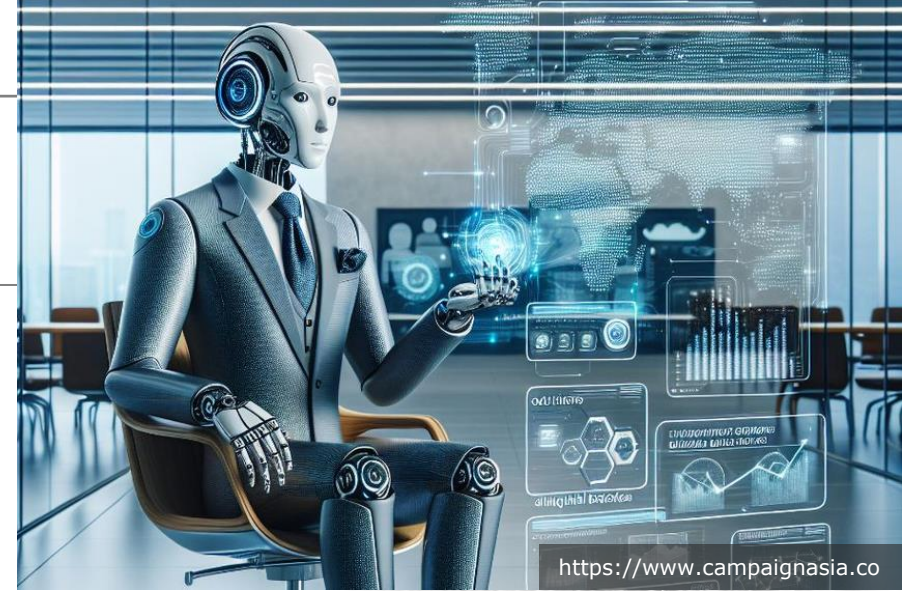
- **Reinforcement Learning from Human Feedback:**
bestärkendes Lernen durch menschliche Inputs
- **Reasoning:**
„überlegen“
→ Partitionierung
in Sub-Tasks
- **Web-Information:**
Integration von
Live-Daten aus
dem Internet



<https://www.aiwire.net>

KI 2025+

- **AI Agents**
 - ◆ Autonome Workflows (+Entscheidungsfindung)
- **LLMs → LMMs**: Disruption in der Wissenschaft u.a.
- **Conversational AI**: nahtlose Interaktion über natürliche Sprache
- **„Günstige“ KI?**



„Nachhaltige“ KI

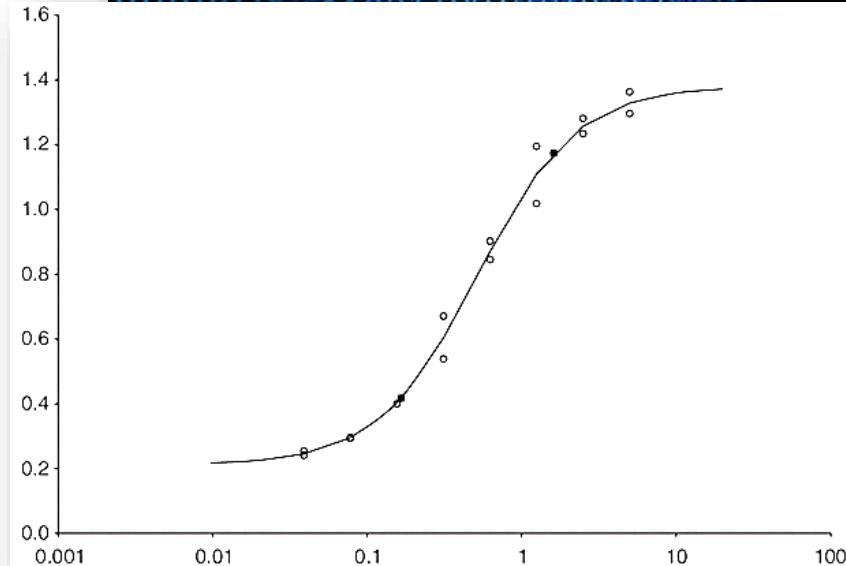
- Daten verstehen
- Erklärbare KI
- Jenseits der Confusion Matrix
- Wahrung der Privatsphäre
- Interdisziplinäre Ansätze



<https://images.theconversation.com/>

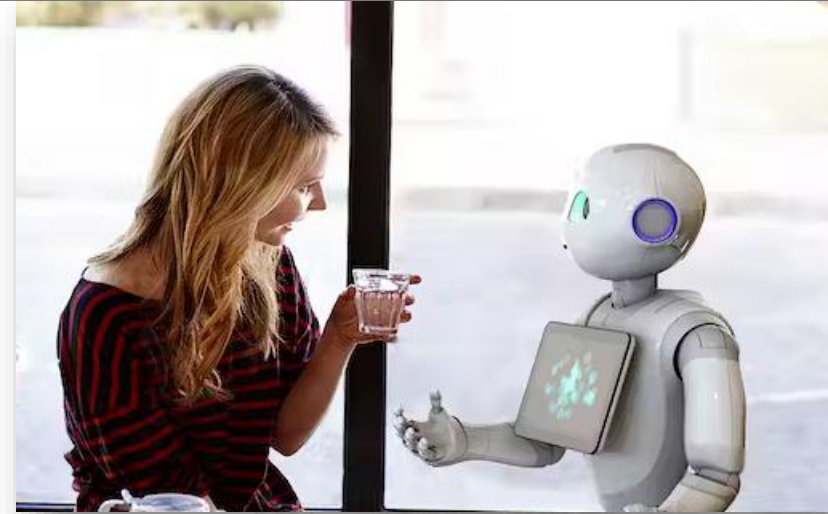
Daten verstehen

- Früher: **Inputdaten** → **Output** → magic!
- Jetzt:
 - ◆ Weniger Optimierung der Modelle, mehr **Optimierung der Daten!**
 - ◆ Data Scientists → Data Engineers



Erklärbare KI

- Wir sind besser als die Baseline, aber wissen nicht, warum
 - “Black-box“-Problem
 - Akzeptanz, Akzeptabilität?
- ➔ **Explainable AI (XAI)**
 - ➔ Erklärbarkeit, Transparenz, Interpretierbarkeit
- ➔ Evaluierungsmetriken jenseits der „Confusion Matrix“



Evaluierung: Jenseits der Confusion Matrix

- Wir sind besser als die Baseline, aber wissen nicht, warum
- ➔ Confusion Matrix: div. Implikationen
(Qualität der Trainingsdaten,
engineered Features,
ML-Architektur, ...)
- ➔ **Repräsentation des untersuchten
Phänomens?**
- ➔ Vorsicht bei der Interpretation
der Ergebnisse!



Wahrung der Privatsphäre

- ...entlang der gesamten **Wertekette**
 - ◆ IT-Sicherheit, Datenspeicherung
Datenanalyse, Visualisierung,
Datenweitergabe
 - Generierung von künstlichen Daten!
- Rechtliche UND v.a. ethische Fragen
+ societal questions!

Kounadi, O. and Resch, B. (2018) A Geoprivacy by Design Guideline for Research Campaigns that use Participatory Sensing Data. *Journal of Empirical Research on Human Research Ethics*, 13(3), pp. 203-222.
Kounadi, O., Resch, B. and Petutschnig, A. (2018) Privacy Threats and Protection Recommendations for the Usage of Geosocial Network Data in Research. *Social Sciences*, 7(10), pp. 191.



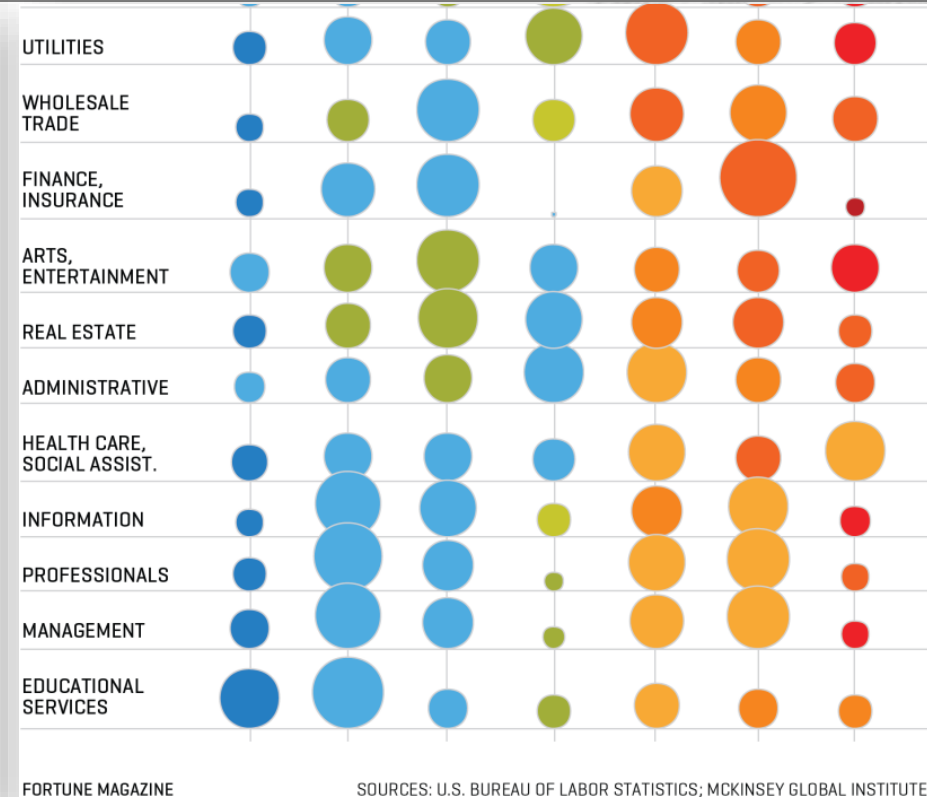
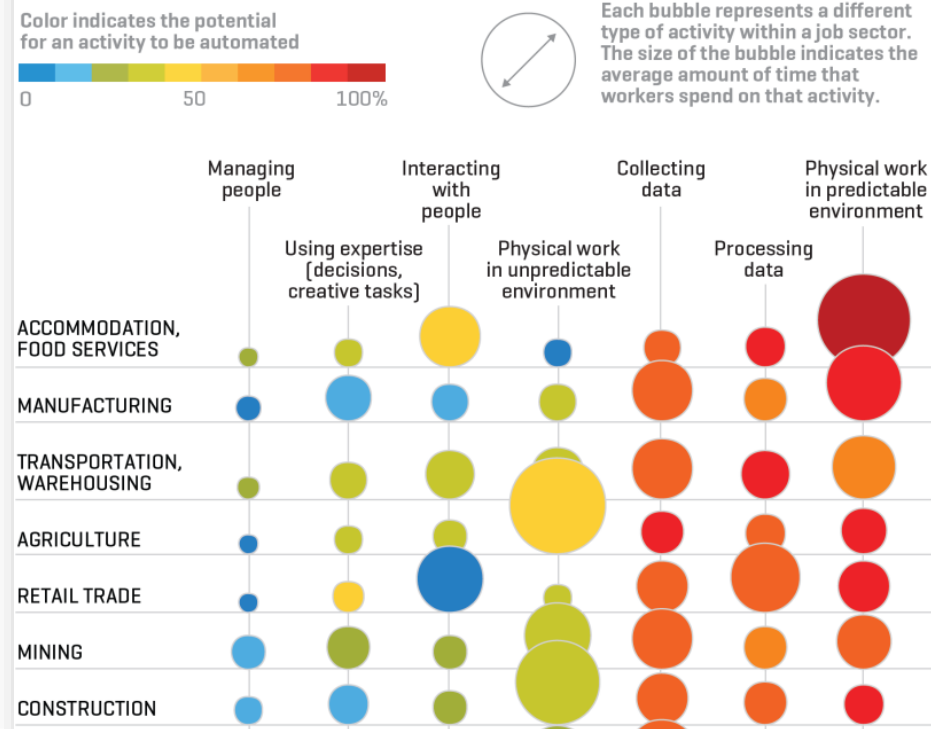
Abschließende Gedanken...



- **AI: Illusion of Mastering**
- „65% der Jobs, die unsere Kinder einmal ausüben werden, existieren heute noch gar nicht.“
→ **Wofür bilden wir unsere Kinder (aus)?**
- ***Kreativität, Flexibilität, Kommunikation***
- **“Kreativität”** von KI?

Potenzial für Automatisierung

POTENTIAL FOR AUTOMATION BY SECTOR AND ACTIVITY



Abschließende Gedanken...

- Maschinen treffen **moralische Entscheidungen**?
<https://www.moralmachine.net/hl/de>
- Wie wollen wir als **Gesellschaft** mit KI umgehen?
- Vorsicht mit dem Statement:
„Aber XXX wird die KI niemals können.“





KI – Zwischen Transparenz und Teilhabe
St. Virgil, Salzburg

KI Verstehen

Trends und Mythen

Bernd Resch

08. Mai 2025