



# KI & MORAL

NICHOLAS THORNTON

MAXI WIDDECKE

ARTHUR FALK

# Gliederung

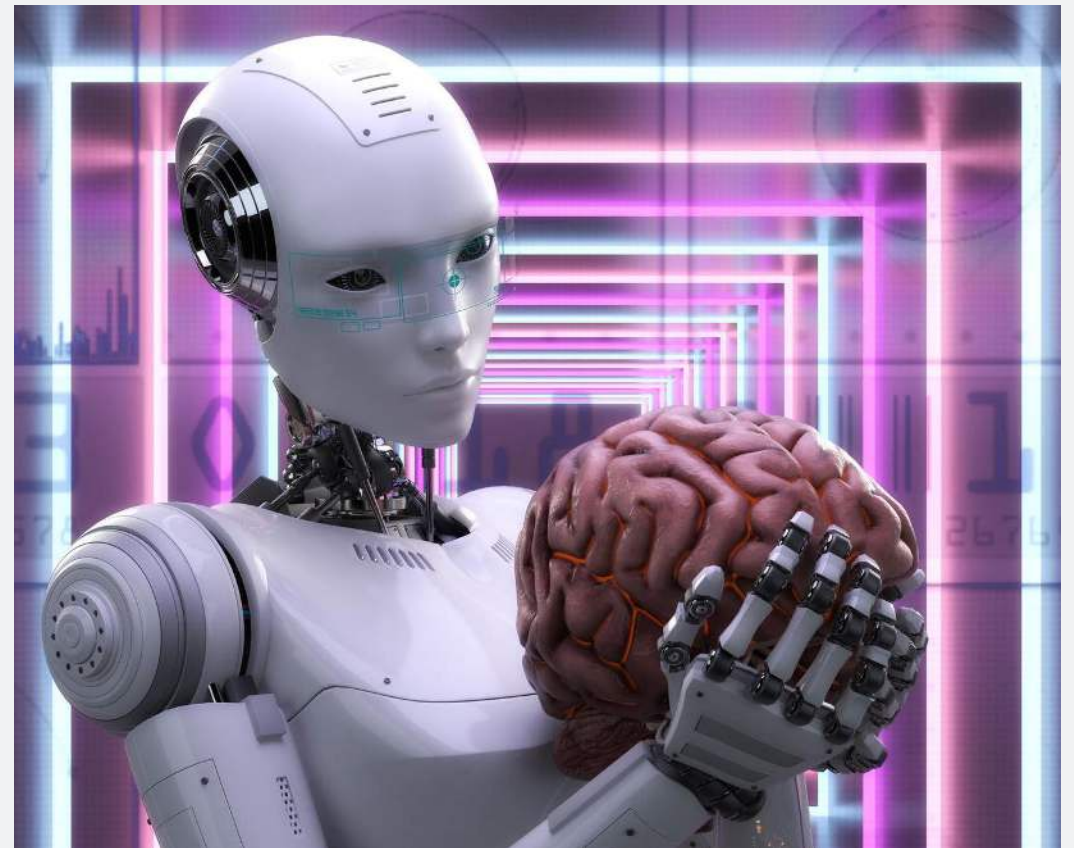
## I. Moral

- Was ist Moral
- Physikalischer Ansatz der Glaubenstheorie
- Ist Moral lehrbar?

## II. KI & Moral

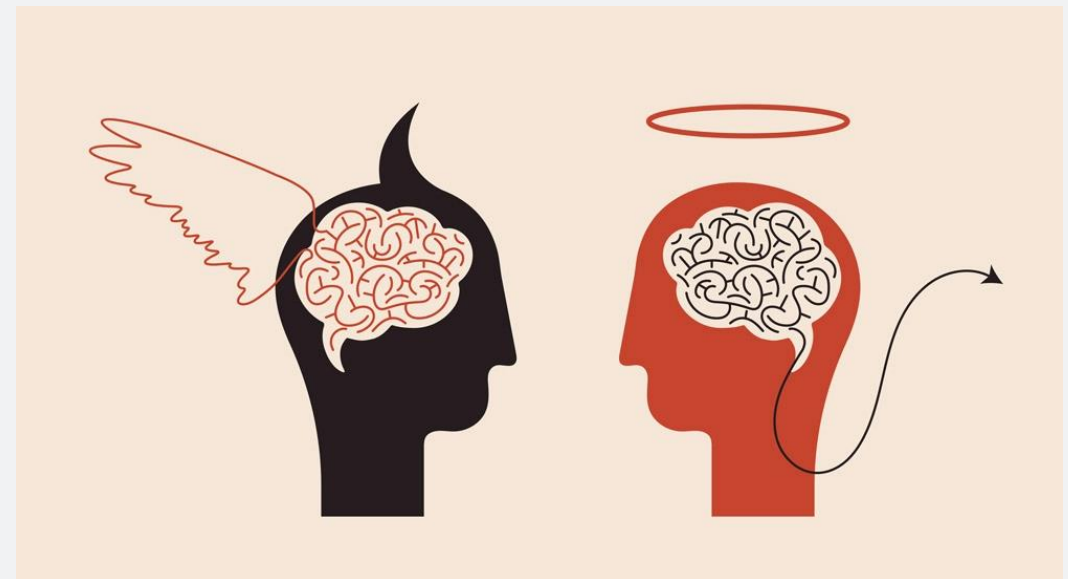
- Wieso braucht KI Moral?
- Gedankenexperiment
- Mögliche Probleme bei der Entwicklung von einer moralischen KI
- Soll KI eine Moral überhaupt besitzen?
- Delphi als Beispiel einer moralischen KI?

## III. Fazit

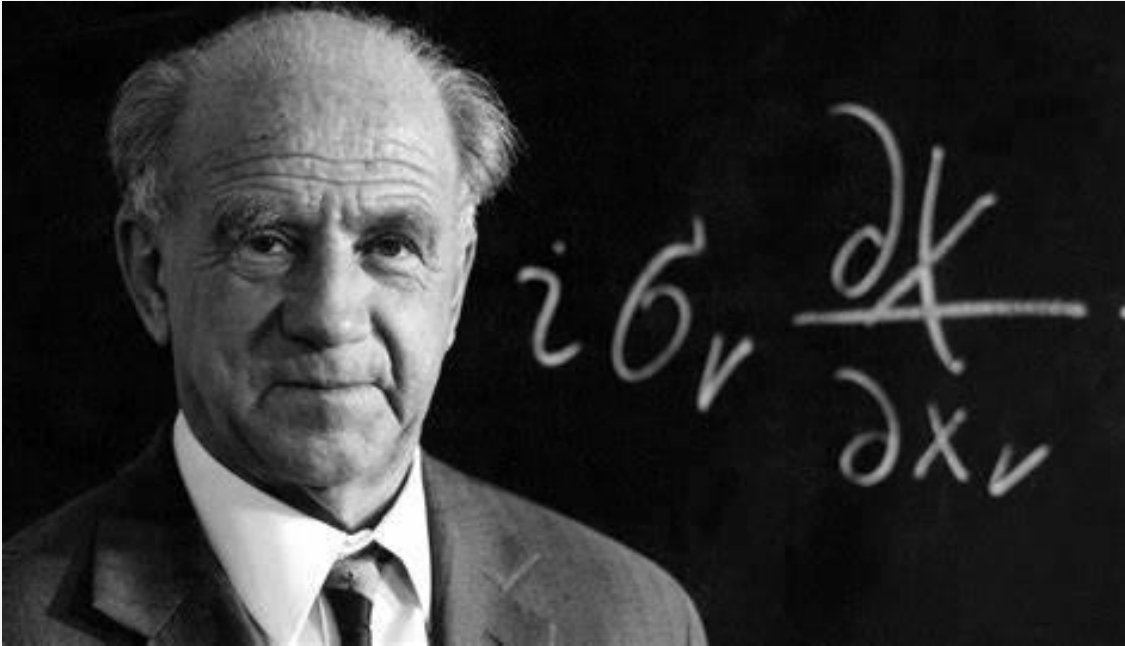


# Was ist Moral?

- Moral ist aufzuteilen in zwei separate aber zusammenhängende Teile
- *Intuitive Moral*: Natürlich vorgegebenen Eigenschaften eines Menschen, wie Empathie, aus denen der Mensch als Konsequenz seine moralischen Grundvoraussetzungen zieht.
- *Kollektive Moral*: Werte und Normen die der Mensch im Laufe seiner Existenz aus seiner Umwelt (Familie, Mitmenschen, Staat), übernimmt.
- Die beiden Moralunterpunkte wirken aufeinander ein und bilden das **subjektive Moralspektrum**
- Schlussfolgerung -> Moral kann nicht bestimmt werden, da ethisches Verhalten eine komplexe individuelle, soziale und kulturelle Leistung ist, die nicht von einer Person definiert werden kann
- Moral kann auch aufgrund von Glaube begründet werden -> Moral ist keine Wahrheit sondern nur Glaube



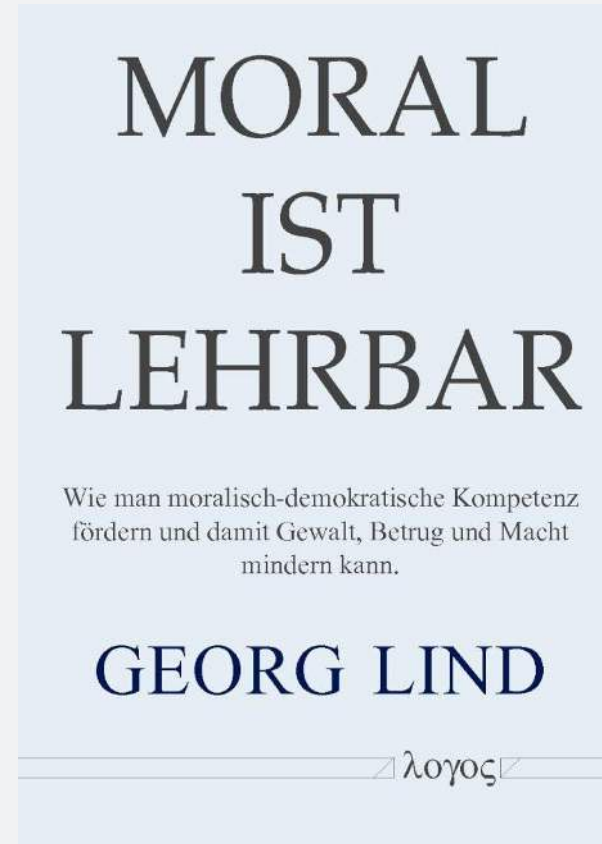
# Glaube und Moral fundiert mit Heisenberg



- Der Glaube ist fundamental (Heisenberg  $\Delta x \cdot \Delta p \geq \hbar$ ), jenes drückt die Unschärfe aus, was uns zwingt die Ungewissheit zu akzeptieren = Glaube
- Beobachtung/Wahrnehmung beeinflusst unsere Umwelt  
---> Glaube beeinflusst unsere Umwelt
- Moral ist immer subjektiv, und Glaube/Überzeugung hat natürlich einen Einfluss auf das subjektive Moralbild
- Simple KI kann kollektive Moralvorstellungen übernehmen jedoch keine subjektive
  - Funktionales Äquivalent zu menschlicher kollektiver Moralvorstellung
- Quantencomputern können die Unschärfe füllen mit möglichen Lösungen (Quantenparallelismus)
- Quantencomputer können die möglichen Lösungen =  $[\Delta x \cdot \Delta p]$  in Erwägung ziehen
- Wenn er nun zufällig eine Lösung für subjektive Moralvorstellung wählt, dann könnte man jenes als funktionales Äquivalent zu der menschlichen subjektiven Moralvorstellung sehen

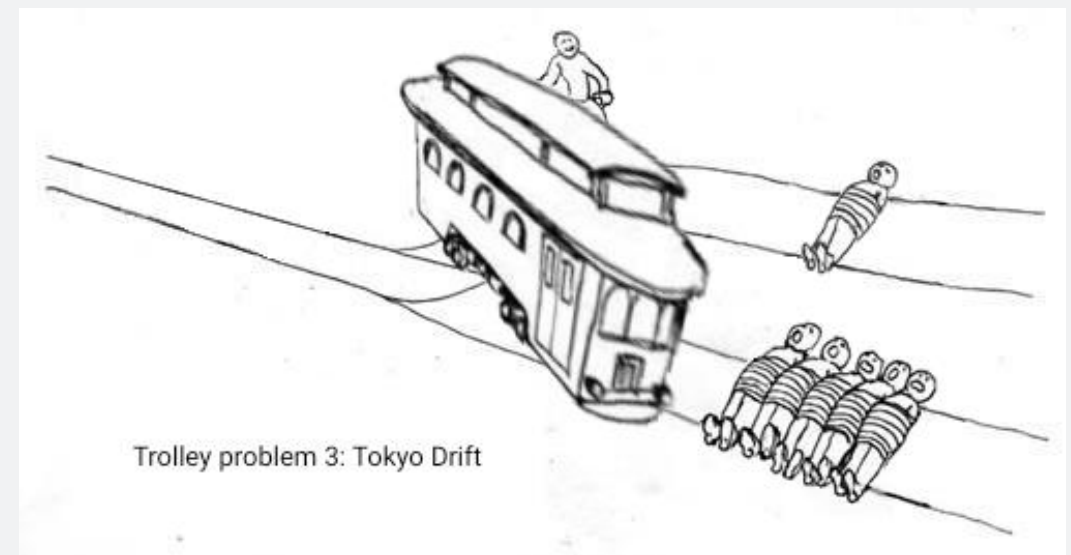
# Ist Moral lehrbar?

- Bildung und Erziehung wichtig für die Entwicklung der eigenen Urteilsfähigkeit
  - Anpassung an soziale, sittliche und kulturelle Normen
- Kollektive Moral ist lehrbar, Menschen lernen sie wenn sie aufwachsen, sind sie aufgewachsen lehren Menschen sie.
- Studie von Haidt zeigt, dass wir Menschen nach kollektiven moralischen Maßstäben handeln, ohne sie genauer zu hinterfragen, ähnlich wie eine Maschine.



# Wieso braucht KI moral?

- Selbstfahrende Autos (Trolley Problem)
- Entscheidungen zwischen Leben und Tod
- Manche Informationen verheimlichen, obwohl sie aufgerufen werden (Pornographie, Herstellung von Drogen, Sprengstoff etc...)(Chat GPT)
- KI Einsatz in vollautomatisierten Waffensystemen
- Erdbeerfeldproblem



# Gedankenexperiment

- Das Gehirn eines neugeborenen Kindes wird mit einer gehirn-ähnlichen KI ersetzt
- Empathie wird programmiert, neuronale Prozesse wie Reize werden elektronisch verknüpft
- Kind wird von einer durchschnittlichen Familie aufgezogen
- KI-Kind lernt Moral durch Eltern
- Fazit: ein großer Teil der Moral wird gelernt und kulturell sittlich beeinflusst



# Mögliche Probleme bei der Entwicklung

- Moral entsteht durch Emotionen nach D. Churchland (University of California)
- Emotionen könnten derzeit nur simuliert werden
  - -> KI kann zumindest jetzt keine menschenähnliche Moral besitzen
- Welche Moral wird programmiert? Wer bestimmt welche Moral die KI haben soll?
- Wer ist verantwortlich, beziehungsweise wer wird bestraft anhand eines Fehlers der KI (Autounfall)
- KI handelt nicht aufgrund von menschenähnlichen Emotionen, somit kann sie zurzeit nur mathematisch systematisch vorgehen
  - Intuitive Moral noch nicht programmierbar





# Soll KI eine Moral überhaupt besitzen?

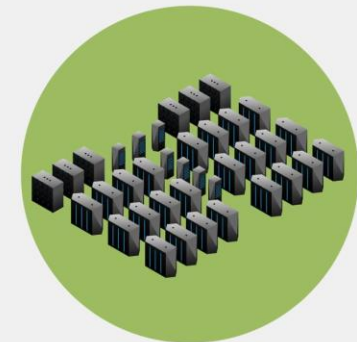
- Systematische, mathematische und morallose Herangehensweise der KI als hilfreich bei manch ethischer Probleme (Bsp. Trolley-dilemma, nach GG müssten beide überfahren werden)
- Eine KI wird in Zukunft den Menschen in allem Gedanklichen überlegen sein
  - FOLGE: Mensch wird überflüssig, und hört auf selber zu denken und zu handeln
- KI könnte eigene moralischen Vorstellung bekommen (Superintelligenz, Quantencomputer),
  - Stellt eigene Moral über die der Menschen (FOLGE: Auslöschung der Menschheit fördern wegen Klimawandel, Überbevölkerung, Unglück, etc...)
- Implementierung einer Moral in die KI stellt ideologisches Problem dar
  - Entwickler setzen immer **ihre** moralischen Überzeugungen um

## Rechenleistung im Vergleich



**8 Mrd. Menschen  
brauchen  
20.000 Jahre**

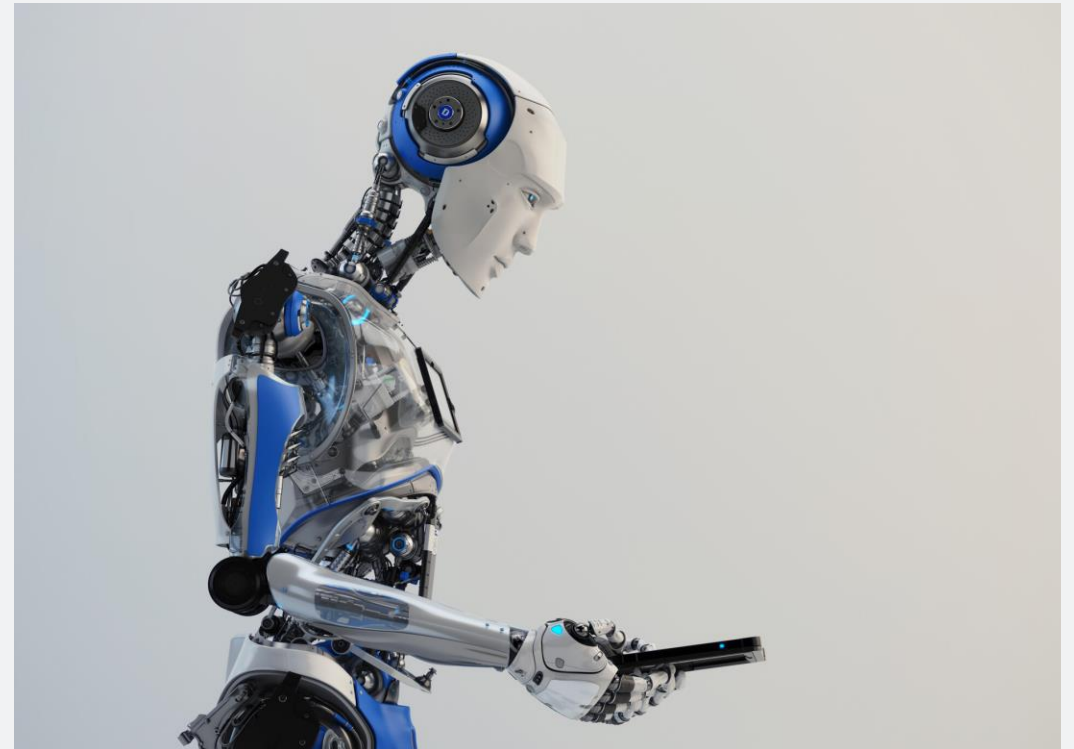
=



**1 Supercomputer  
braucht  
20 Minuten**

# Delphi

- Ethischer Algorithmus
- Beurteilung menschlicher Verhaltensweisen basierend auf einer “Commonsense Norm Bank”
- Antwortet mit “In Ordnung”, “Schlecht”, “Widerlich” etc...
- Problem: GIGO-Prinzip durch riesige Datenbank
- -> Wer darf die Moral der KI programmieren, beziehungsweise bestimmen?
- FOLGE: Fast unmöglich, einen allgemein verbindlichen Code für moralisches Handeln zu entwickeln





# FAZIT

- KI kann keine bedingungslose, menschenähnliche Moral besitzen
- KI kann kollektive Moral besitzen, dies wäre allerdings nur eine Simulation und Nachahmung der menschlichen Moral und würde deshalb nur bedingte Autonomie besitzen
- Selbst wenn KI eine solche Moral besitzen würde, wären wahrscheinlich die Folgen für die Menschheit katastrophal
- Quantencomputer würden allerdings in Zukunft ein funktionales Äquivalent zur subjektiven Moral bilden können.

# QUELLEN

- TEXTQUELLEN:
- NZZ: Delphi-Artikel <https://www.nzz.ch/technologie/wuermer-essen-ist-widerlich> (2022)
- New York Times: Can a machine learn morality? <https://www.nytimes.com/2021/11/19/technology> (2021)
- Richard David Precht: “Künstliche Intelligenz und der Sinn des Lebens” (2020)
- Lacroix; Können Maschinen moralisch handeln? <https://www.zeit.de/kultur/2018-05/kuenstliche-intelligenz-moral-menschen-maschinen-werte> (2018)
- Hoffman; Kann künstliche Intelligenz moralisch denken? <https://www.nzz.ch/feuilleton/kuenstliche-intelligenz-hat-zurzeit-weder-moral-noch-autonomie-id.1443718>
- Lexikon philosophischer Begriffe aus der Bücherei
  
- BILDQUELLEN:
- <https://www.acadoo.de/wp-content/uploads/2020/05/2018-03-1-Serioese-Quellen-1080x675.jpg>
- <https://images.t-online.de/2021/05/82712714v1/0x0:1920x1120/fit-in/1920x0/>
- <https://i.redd.it/shpzyd7otcb1.png>
- [https://www.scinexx.de/wp-content/uploads/m/o/moralhirn\\_g.jpg](https://www.scinexx.de/wp-content/uploads/m/o/moralhirn_g.jpg)
- [https://cdn.prod.www.manager-magazin.de/images/94c159c9-0001-0005-0000-000001780691\\_w323\\_r0.7690476190476191\\_fpx49.85\\_fpy51.67.jpg](https://cdn.prod.www.manager-magazin.de/images/94c159c9-0001-0005-0000-000001780691_w323_r0.7690476190476191_fpx49.85_fpy51.67.jpg)
- <https://bilder.deutschlandfunk.de/e4/0c/4e/24/e40c4e24-3ae3-44aa-aea0-277901955b9b/kuenstliche-intelligenz-122-1920x1080.jpg>
- <https://cdna.artstation.com/p/assets/images/images/034/840/374/large/vladislav-ociacia-v-21.jpg?1613395990>
- <https://img.welt.de/img/wissenschaft/mobile234771692/9292505627-ci102l-w1024/Saudi-humanoid-robot-Sophia-debuts-on-Chinese-TV-program-2.jpg>

