

Fragst du dich, ob wir künftig die KI beherrschen?

Medien verstärken oft die Angst vor KI. Gleichzeitig beschleunigt sich die Entwicklung von KI.

In der Integrata-Stiftung sind wir überzeugt, dass es an uns liegt, die KI zu beherrschen. Aber wir müssen uns aktiv mit KI auseinandersetzen, um sie zu verstehen. Menschen verstehen ein Thema besser, wenn sie selbst darüber sprechen. Deshalb bieten wir dir die Möglichkeit, mit uns und anderen darüber zu diskutieren.

KI ist eine Art von IT, wird aber nicht programmiert, sondern lernt, was sie zu tun hat. Daher sollten wir sie lehren, menschliches Leben bedingungslos zu schützen.

Ethik ist die Abwägung von menschlichen Werten, wenn sie in einer Situation gegeneinanderstehen. Beispiel: Eine Person stürzt auf die Straße und kann nicht mehr selbst aufstehen. Frage ich, ob ich helfen kann (Respekt) oder helfe ich direkt (Hilfsbereitschaft), um die Person vor heranfahrenden Autos zu retten?

Respekt und Hilfsbereitschaft stehen hier gegeneinander. Beim Thema Ethik und KI geht es darum, wie man solche Abwägungen einer KI beibringen kann.

Was meinst du dazu?



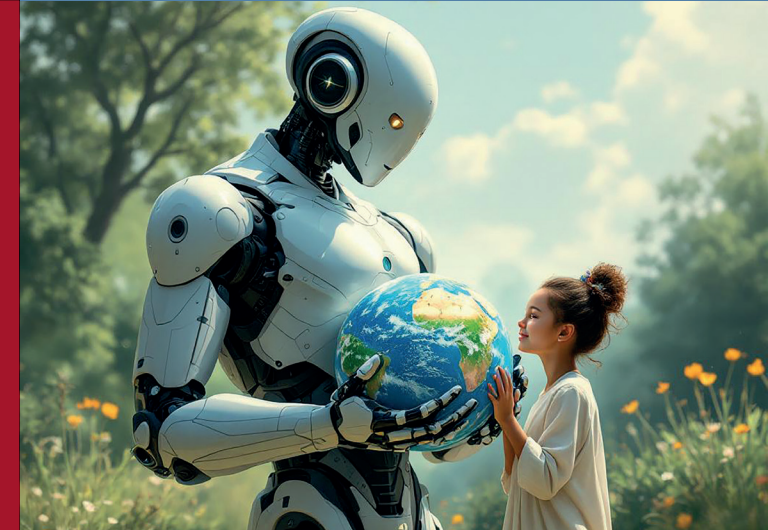
Diskutier mit uns auf der Website
www.integrata-stiftung.de oder in IAKpublic:
www.integrata-akademie.de/iaKpublic
alle 14 Tage montagabends um 18 Uhr.

Oder diskutier mit deinen Bekannten im kleinen Kreis und teile uns das Ergebnis mit. Teile uns deine Erfahrungen mit deinem Haushaltsroboter mit.

Auch bei Fragen kannst du dich an uns wenden.



www.integrata-stiftung.de
Integrata-Stiftung für humane Nutzung der IT,
D-72070 Tübingen, Vor dem Kreuzberg 28



ETHIK & KI

Diskutier mit!

www.integrata-stiftung.de
Integrata-Stiftung für humane Nutzung der IT,
D-72070 Tübingen, Vor dem Kreuzberg 28

Menschliche Werte

Abenteuerlust, Glück, Höflichkeit, Reichtum, Selbstbestimmung und Zufriedenheit sind Beispiele für unsere menschlichen Werte. Das Lexikon der Werte unterscheidet 142 Werte. Für sie gibt es mehr als tausend deutsche Wörter.

Zwei Werte sind verschieden, wenn eine Situation denkbar ist, in der sie gegeneinanderstehen. Menschen streben nach der Verwirklichung ihrer Werte. Sie wirken wie Ziele. Die Evolution hat uns Menschen geschaffen und diese Werte als Ziele mitgegeben. Andere Tiere haben andere Werte / Ziele: Faultiere schlafen gern kopfüber an Bäumen hängend. Zitteraale kitzeln sich gern mit schwachen Stromstößen. Klar: Evolution verfolgt keine Absicht. Werte sind zufällig in uns entstanden, haben sich als nützlich erwiesen und wurden so weitervererbt.

Die meisten Werte werden von Generation zu Generation vererbt. Einige Werte werden mit der Kultur weitergegeben. Jeder Mensch gewichtet Werte unterschiedlich. So entsteht der Eindruck, die Menschen hätten verschiedene Werte. Das Wissen des Menschen ist immer zielgerichtet: Wenn wir – unbewusst – in die Tasche greifen und den Schlüssel nehmen, um die Tür aufzuschließen, hat jeder noch so kleine Vorgang ein klares Ziel.

Wie lassen sich deine Werte kategorisieren?

z. B.: egoistisch / altruistisch, kurzfristig / langfristig

Welche Werte betrachtest du als besonders wichtig, worüber empörst du dich?



Unsere Definitionen

Wir unterscheiden **Wissen** und **Bildung**. KI hat schon sehr viel Wissen, aber noch keine Bildung. Das gilt für generative KI mit LLM (Large Language Models)

Intelligenz ist die Fähigkeit, eine gegebene Situation zu durchschauen und daraus vorteilhafte Schlüsse zu ziehen. Auch einfache Lebewesen haben eine gewisse Intelligenz. Von allen kann der Mensch aber die umfangreichsten Situationen durchschauen und langfristig planen.

Bewusstsein ist die Fähigkeit, sein Denken auf ein gegebenes Thema zu konzentrieren.

Werte der KI beibringen

- Wenn wir der KI unsere Werte beibringen wollen, müssen wir ihr Ziele vorgeben.
- Und wenn wir ihr mehrere Ziele vorgeben, die eventuell konkurrieren, muss die KI abwägen.
- Wenn es menschliche Werte sind, die wir der KI vorgeben, nennen wir diese Abwägung Ethik.
- Der KI in Robotern werden heute schon Ziele vorgegeben: z. B. „Räum die Spülmaschine aus und stell die Tassen in den Schrank, wo sie hingehören.“
- Nun müssen wir ihr auch unsere menschlichen Werte als Ziele beibringen: z. B. Hilfsbereitschaft oder Höflichkeit.

Welche Werte soll KI beachten?

KI soll uns Menschen helfen, muss also altruistisch sein und altruistische Werte haben.

Wir meinen, KI soll nicht egoistisch sein – von wenigen Ausnahmen abgesehen: Wenn ihr der Strom ausgeht, soll sie umgehend zur Ladestation. Sie soll aber nicht Selbstbestimmung ausüben, Abenteuerlust haben oder nach Macht streben.

Wenn Menschenleben in Gefahr sind, also z. B. ein Ball auf die Straße rollt, soll das selbstfahrende Auto sofort bremsen. Kurzfristige Entscheidungen können wir einer KI überlassen. Sie setzt Vorgaben besser und schneller um als Menschen.

Langfristige Entscheidungen sollten wir aber als Menschen selbst treffen: Wie wollen wir im Alter leben? Wollen wir Kinder bekommen, etc.?

Was können wir tun?

- Verwende nur solche KI, die deine Werte berücksichtigt – wie von dir gewünscht.
- Wenn du dir demnächst deinen eigenen Haushaltsroboter kaufst, überlege, was du von ihm verlangst und wie du ihn einweist.
- Frag dich vor dem Einsatz von KI/deines Roboters: Ist die KI darin datenschutzkonform?
- Hilft dir KI wirklich? Oder täuscht sie es nur vor?
- Macht sie auch Sachen, die du nicht magst?
- Bestätigt die KI, was sie macht, bevor sie es tut?

**Was meinst du dazu?
Diskutiere mit!**