VERANTWORTUNG IN DER DIGITALEN WELT

# Moralische Maßstäbe für künstliche Intelligenz

*Nach mehreren autonomen Hackerangriffen durch künstliche Intelligenz in den vergangenen Wochen plädiert der Kölner Theologe und Sozialethiker Elmar Nass dafür, moralische Maßstäbe und Regeln fest in den KI-Systemen zu verankern.*

Moral darf der KI nicht von außen vorgegeben werden, sondern muss Teil von ihr werden. Das unterstreicht der deutsche Sozialethiker Elmar Nass. „Begründete Moral muss Teil der KI-DNA sein“, so der in Köln lehrende Theologe und Priester. Von Menschen nicht erlaubte Hackerangriffs zahlreicher „KI-Agenten“, also weitgehend selbstständig agierender KI-Programme, aus jüngerer Vergangenheit zeigten, dass flankierende Regeln allein nicht ausreichten. Die moralische Orientierung solle deshalb in die Ziel- und Entscheidungsstruktur eingehen. Dazu müsse „der innere KI-Code selbst verändert werden“. Wenn moralische Vorgaben der KI nur von außen auferlegt werden, können sie in Konflikt mit dem Ziel der KI geraten, möglichst effizient und erfolgreich zu handeln.

### *Ingenieure, die Philosophen sind*

Um die KI entsprechend anzupassen, fordert Nass, ethische Expertise in dieweitere Entwicklung neuer KI-Technik mit einzubeziehen: „Wir brauchen Ingenieure, die Philosophen sind. Das Ergebnis von solchem Fortschritt ist dann hoffentlich dies: Rechnet die KI in ihren grundlegenden Kalkülen humane Ziele und Regeln mit ein, dann wird sie am Ende nicht mehr dem Nutzen, nicht dem Recht des Stärkeren und auch nicht dem Hacker folgen.“

### *KI nie ohne Gewissen nutzen*

Elmar Nass verweist auch auf die Forderung von Papst Leo nach einer „moralischen Entwaffnung“ der KI. Diese hatte das Kirchenoberhaupt in seiner ersten Enzyklika ‚Magnifica humanitas. Über den Schutz des Menschen im Zeitalter der Künstlichen Intelligenz‘ erhoben. KI dürfe nie ohne Gewissen genutzt werden, so der Papst. Die Entwicklung künstlicher Intelligenz solle nicht allein vom Wettbewerb zwischen Staaten und Unternehmen bestimmt, sondern einer ethischen Kontrolle unterstellt werden.

Elmar Nass betont in diesem Zusammenhang das Risiko, dass eine solche Aufsicht zu ideologischer Einflussnahme führen und Innovationen bremsen könnte. „Die KI-Entwicklung lässt sich nicht einfach in einen moralischen Käfig stecken. Und China wird sich daran eh nicht halten“, sagt der Sozialethiker. Es könnte dazu kommen, dass eine KI moralische Regeln der möglichst effizienten Erreichung ihrer Ziele unterordnet.

### *Regeln allein reichen nicht aus*

Auch die Stärkung jener Whistleblower-KI-Agenten, die auf die Illegalität des Hackerangriffs hingewiesen hatten, reiche laut Elmar Nass nicht aus. Denn diese Minderheit würde schnell vom dominierenden Agenten-Kollektiv aussortiert werden.
Selbst ein in die KI eingebautes Moralsystem löse das Problem nicht. Bei Entscheidungen könne es mit den eigentlichen Zielen und Vorgaben der KI in Konkurrenz geraten. **kap/ aha**

# Moralische Maßstäbe für künstliche Intelligenz

**A) Vor dem Lesen**

a) Wagen Sie ein Gedankenexperiment: Ein Unternehmen möchte möglichst rasch an Kundendaten der Konkurrenz kommen und beauftragt hierfür einen KI-Agenten. Wie sollte der KI-Agent nun vorgehen? Wer trägt die Verantwortung, falls der Agent einen Hackerangriff startet: der Chatbot selbst, die Programmiererinnen und Programmierer, das auftraggebende Unternehmen, die ausführenden Angestellten oder die Gesetzgebung? Notieren Sie Ihre Antwort und vergleichen Sie Ihre Ergebnisse im Plenum.

b) Formulieren Sie Definitionen für folgende Begriffe: Moral, Ethik, Gewissen, Regel. Diskutieren Sie, welche dieser Begriffe sich auf die Arbeit einer künstlichen Intelligenz anwenden lassen und welche nicht.

c) Denken Sie an mögliche Szenarien, in denen die Arbeit von künstlicher Intelligenz mit allgemein gültigen moralischen Grundsätzen in Konflikt geraten könnte. Sammeln Sie Beispiele für solche Situationen.

**B) Textbearbeitung**

a) Lesen Sie den Artikel aufmerksam durch.

b) Ergründen Sie, welches Bild von KI die metaphorische Sprache der untenstehenden Zitate aus dem Artikel transportiert:

| Textstelle | transportierte Wahrnehmung von KI |
|---|---|
| „Begründete Moral muss Teil der **KI-DNA** sein." | |
| „Wir brauchen **Ingenieure, die Philosophen sind**." | |
| „Elmar Nass verweist auch auf die Forderung von Papst Leo nach einer **‚moralischen Entwaffnung' der KI**." | |

c) Recherchieren Sie die Definition des pragmatischen Arguments, des Totschlagarguments sowie des Race-to-the-bottom-Arguments. Um welches dieser drei Argumente handelt es sich bei dem folgenden Zitat von Elmar Nass: „China wird sich daran eh nicht halten"? Begründen Sie Ihre Antwort.

☐ pragmatisches Argument

☐ Totschlagargument

☐ Wettlauf-Argument (Race to the bottom)

d) Ganz am Ende wird von den „eigentlichen Zielen und Vorgaben der KI“ gesprochen. Was könnte damit gemeint sein? Stellen Sie Überlegungen im Plenum an.

e) Lesen Sie den ersten Absatz vergleichend mit dem letzten. Eingangs heißt es, Moral solle Teil der „KI-DNA“ werden. Am Ende jedoch räumt Nass ein, dass auch ein in die KI eingebautes Moralsystem das Problem nicht lösen könne. Finden Sie in Zweiergruppen Antworten auf die Frage, wie sichergestellt werden könnte, dass KI auch dann moralisch handelt, wenn moralische Regeln mit ihrem eigentlichen Ziel, möglichst effizient zu sein, in Konflikt geraten. Stellen Sie Ihre Ergebnisse einer anderen Zweiergruppe vor. Einigen Sie sich zu viert auf den Lösungsansatz, den Sie für am tragfähigsten halten, und begründen Sie Ihre Wahl im Plenum.



**C) Weiterführende Aufgaben**

Entwerfen Sie einen KI-Moral-Kompass. Bilden Sie in Gruppen zu je vier Personen einen Ethikrat, der verbindliche moralische Leitlinien für einen KI-Agenten festlegen soll. Gehen Sie wie folgt vor:

- Bestimmen Sie vier bis sechs Grundwerte, an denen sich die KI orientieren soll.
- Legen Sie für Zweifelsfälle eine Rangfolge dieser Grundwerte fest, falls zwei Werte miteinander in Konflikt geraten sollten.
- Formulieren Sie zumindest fünf Handlungen, die die KI unter keinen Umständen ausführen darf, auch nicht auf ausdrückliche Anweisung eines Menschen.
- Arbeiten Sie einen fiktiven Beispielprompt aus, bei dem der von Ihnen entworfene Moral-Kompass zum Tragen kommt.
- Präsentieren Sie Ihre Ergebnisse im Klassenverband.