

# Was ist, was kann, was soll KI?

Ein philosophisches Gespräch

Vincent C. Müller, Martin Hähnel



Vincent C. Müller | Martin Hähnel

# **Was ist, was kann, was soll KI?**

Ein philosophisches Gespräch

Meiner

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der  
Deutschen Nationalbibliographie; detaillierte bibliographische  
Daten sind im Internet über <https://portal.dnb.de> abrufbar.

ISBN 978-3-7873-4672-1

ISBN eBook 978-3-7873-4673-8

© Felix Meiner Verlag Hamburg 2024. Alle Rechte vorbehalten. Dies gilt  
auch für Vervielfältigungen, Übertragungen, Mikroverfilmungen und die  
Einspeicherung und Verarbeitung in elektronischen Systemen, soweit es  
nicht §§ 53, 54 UrhG ausdrücklich gestatten. Satz: Jens-Sören Mann.  
Druck und Bindung: Stückle, Ettenheim. Gedruckt auf alterungsbestän-  
digem Werkdruckpapier, hergestellt aus 100% chlорfrei gebleichtem Zell-  
stoff. Printed in Germany. [www.meiner.de](http://www.meiner.de)

## INHALT

VORWORT .....	7
DAS GESPRÄCH .....	9
TEIL I: PHILOSOPHIE UND KI .....	17
TEIL II: ETHIK, RECHT UND ÖKONOMIE DER KI .....	25
TEIL III: KI ZWISCHEN ÖFFENTLICHKEIT UND PERSÖNLICHER LEBENSWELT .....	75
TEIL IV: LETZTE FRAGEN .....	93
ANMERKUNGEN .....	99
REGISTER .....	103

## VORWORT

Dieses Interview findet in einem Zeitraum statt, der nicht besser hätte gewählt werden können. Nicht nur, dass sich der diesjährige Kongress der Deutschen Gesellschaft für Philosophie ausführlich mit den Themen Digitalisierung und Künstliche Intelligenz auseinandersetzt, was zeigt, dass die Eule der Minerva ihren nächtlichen Flug begonnen hat und die Zeit reif ist, um eine umfassende, nicht von Euphorie oder Skepsis geleitete philosophische Reflexion zu einem Forschungsfeld anzustellen, das sich weiterhin stark verändert und dessen technologische, politische, soziale und ethische Tragweite wir immer noch nicht gänzlich ermessen können. Auch gesellschaftlich befinden wir uns in einer (Zwischen-)Phase, in welcher erste »Normalisierungsprozesse« einsetzen: Eine anfangs als disruptiv erfahrene Technologie »veralltäglich« sich allmählich und wird eingesetzt, um das Leben vieler Menschen »smarter« zu machen, während auf der einen Seite die ökonomischen Effizienzsteigerungspotentiale, die in KI-Technologien beschlossen liegen, gezielter freigesetzt werden, auf der anderen Seite aber auch immer mehr diejenigen menschlichen Lebensbereiche identifiziert oder neu geschaffen werden, die möglichst frei von digitalen Anwendungen sind bzw. sein sollen. Ferner nehmen erste rechtliche Regelungen Gestalt an und der akademischen Ethik und Technikphilosophie ist es in den letzten Jahren gelungen, eine elaborierte deskriptive und normative Sprache zu entwickeln, die es ermöglicht, Chancen und Risiken neuer digitaler Technologien besser zu erfassen und zu bewerten. Gleichwohl bleibt vor allem das Thema der Künstlichen Intelligenz ob seiner anhaltenden transakademischen Faszination, der wachsenden Unübersichtlichkeit ihrer Anwendungen und einer der modernen Technologieentwicklung innewohnenden und verschiedene Ambiguitäten erzeugenden Dynamik ein Dauerbrenner.

Das folgende Interview verdankt sich vor allem der Notwendigkeit, das Thema KI auf den philosophischen Prüfstand zu bringen, ohne in feuilletonistischen Vermutungen zu verharren oder sich in fachwissenschaftlichen Detaildebatten zu verlieren. Wir sind uns darüber im Klaren, dass das Gespräch nur einen kleinen Ausschnitt aus dem gesamten Spektrum einer philosophischen Diskussion zur Künstlichen Intelligenz bietet. Allerdings erhoffen wir uns, dass der im Titel dieses Buches angekündigte Dreischritt, der keine Schlussfolgerung mit logischer Notwendigkeit darstellen soll, etwas mehr Ordnung und Licht in die aktuelle Diskussion bringt: Denn zunächst gilt es zu klären, was Künstliche Intelligenz dem Begriff und der Technologie nach eigentlich *ist*. Nur wenn wir wissen, was KI ist, können wir auch abschätzen, wozu sie fähig bzw. nicht fähig ist. Was KI demzufolge *kann* oder nicht kann, bietet die Grundlage für normative Überlegungen, denn KI-Systeme *sollen* letztlich nur das tun, was sie gut können – nicht mehr und nicht weniger.

Vielelleicht werden sich einige vor oder bei der Lektüre fragen, wieso und unter welchen Umständen das Interview zustande gekommen ist. Die Antwort ist ganz einfach: Zwei Philosophen, der eine international renommierter KI-Experte und Inhaber der Alexander-von-Humboldt-Professur an der Friedrich-Alexander-Universität Erlangen-Nürnberg, der andere Medizinethiker, ebenfalls im KI-Bereich tätig und mit großem Interesse an »öffentlicher Philosophie«, treffen sich im digitalen Raum. Letzterer stellt in bewusster Abgrenzung zu rein journalistischen Interviews dem Experten philosophische Fragen rund um das Thema KI. Das Interview bietet nicht nur eine Momentaufnahme bzw. einen groben Überblick über die aktuellen philosophischen und ethischen Trends in Bezug auf KI, sondern zeigt auch, inwieweit KI Fragen berührt, die man in der Philosophie schon seit Aristoteles diskutiert. Ob wir diesem Anspruch, über KI angemessen zu philosophieren, gerecht geworden sind, mag die Leserin oder der Leser beurteilen. Jedenfalls danken wir dem Felix Meiner Verlag, vor allem seinem Lektor Marcel Simon-Gadhof, sich auf dieses Experiment eingelassen zu haben, und Max Hellrigel-Holderbaum für die Unterstützung bei der Transkription. KI war nicht gut genug!

M. H.

# DAS GESPRÄCH

**MARTIN HÄHNEL:** Herr Professor Müller, ich begrüße Sie zu unserem kleinen Interview zum Thema KI und Philosophie. Ich habe mir vorgestellt, dass wir das Interview in vier Teile aufgliedern. Nach ein paar allgemeinen Fragen soll der erste Teil vom Verhältnis Philosophie und KI handeln. Der zweite Teil wird ethische, rechtliche, ökonomische Aspekte von KI beleuchten und auf bestimmte neuralgische Punkte in der Diskussion eingehen. Der dritte Teil wird künstliche Intelligenz im Bezug zur Öffentlichkeit und zur Lebenswelt der Menschen thematisieren. Und am Ende gibt es noch einen vierten Teil mit kurzen lockeren Fragen. Ich freue mich sehr, dass wir die Möglichkeit haben, mit Ihnen hier dieses Gespräch zu führen, und ich würde direkt einsteigen mit den drei zentralen Fragen, die im Titel dieses Interviews schon erwähnt werden und die sich an die berühmten Kant'schen Fragen anlehnen: Was ist für Sie Künstliche Intelligenz? Und dann: Was kann und soll sie?

**VINCENT C. MÜLLER:** Vielen Dank. Ich habe diese Kant'schen Fragen schon verschiedentlich verwendet, als charakteristisch für die Beschäftigung mit der KI. In der Tat ist die erste Frage schon einmal gleich recht schwierig. Was ist KI? Dummerweise ist das keine Frage der KI, sondern eher eine Frage der Philosophie, also sind wir zuständig. In einer ersten Näherung ist klar, dass wir uns gegenwärtig KI immer als etwas vorstellen, was auf Rechenmaschinen stattfindet, also erstens auf *Maschinen* und zweitens auf einer bestimmten *Sorte* von Maschinen, nämlich auf denen, die *rechnen*. Und diejenigen, die wir da gegenwärtig verwenden, sind stets *digitale* Rechenmaschinen. Nun, was macht man mit diesen digitalen Rechenmaschinen? Meine Auffassung ist, dass man sie am besten als Entscheidungsmaschinen ansehen kann, also Maschinen, die irgendeine Art von Frage als Input bekommen (einschließlich normativer Aspekte) und daraus eine Entscheidung generieren. Also zum Beispiel: »Ist das, was ich da sehe, eine Katze?« oder »Sollte dieser Mitarbeiter befördert werden?« oder »Sollte ich jetzt diesen Satz sagen?« Wenn man sich das kurz überlegt, stellt man fest,

dass das, was wir den ganzen Tag machen, im Wesentlichen aus Entscheidungen besteht, genauer aus Entscheidungen, bestimmte Handlungen durchzuführen. Oft sind die natürlich nicht explizit, aber man kann das ganz gut als allgemeines Modell verwenden. Also in der KI geht es um die Erzeugung von künstlichen Entscheidungsmaschinen. Was KI kann, ist eigentlich damit schon geklärt. In einem abstrakten Sinne kann sie alles tun, was man als Entscheidung ansehen könnte. *Und nur das.* Also zum Beispiel kann sie keinen Stein hochheben, weil das zwar eine Entscheidung beinhaltet (den Stein hochzuheben), aber es beinhaltet auch eine physische Interaktion. Das kann eine KI zunächst nicht. Das kann sie nur dann, wenn sie mit einer entsprechenden *Hardware* verbunden wird.

Kann die KI alles das machen, was wir jetzt unter dieser Klammer der »Entscheidung« gefasst haben? Das ist natürlich eigentlich eine technische Frage. Es ist aber auch eine interessante Frage, ob es Grenzen dafür gibt, was vielleicht eine Maschine nicht kann, *weil* sie eine Maschine ist oder *weil* sie eine *digitale* Maschine ist. Ich glaube, es gibt tatsächlich gute Hinweise darauf, dass das der Fall ist. Es gibt bekannte Probleme in der Philosophie, denen zu folge eine Maschine zum Beispiel nicht meinen kann, was sie sagt, und keine Intentionen oder Absichten haben kann. Es ist eigentlich für eine Maschine nicht eines besser als das andere. Sie hat also eigentlich keine *Bewertung* der Umwelt, außer einer Bewertung, die wir ihr einbauen. Und mir scheint tatsächlich, dass das ein wesentlicher Unterschied zwischen einer Maschine, insbesondere einer digitalen Maschine, und einem Lebewesen ist. Es gibt noch eine Reihe von anderen Unterschieden, wie zum Beispiel, dass die persönliche Identität einer Maschine sehr viel, sagen wir mal, unklarer ist als die eines biologischen Wesens. Und das hängt wahrscheinlich auch mit dem ersten Punkt zusammen.

Der letzte Punkt: Was soll eine KI? Das ist die grundlegende Frage der KI-Ethik und es ist ja bekannt, dass es eigentlich in fast allen menschlichen Bereichen einen Unterschied gibt zwischen dem, was man machen kann, und dem, was man machen soll. Sonst hätte man ja auch wenig Entscheidungsfreiheit. Und die systematische Reflexion darüber, was man machen *soll*, nennt man *Ethik*. Dementsprechend, scheint mir, ist es bei der KI nicht prinzipiell anders als bei Menschen. Das heißt, man überlegt sich, was ein KI-System

tun soll. Nach welchen Prinzipien soll ein KI-System entscheiden? Also soll es zum Beispiel regelbasiert sein? Oder soll man Berechnungen für den Nutzen von Konsequenzen aufstellen; oder wie soll man das machen? Welche Form der normativen Ethik wendet man an und wie geht man mit Konflikten oder sogenannten Dilemmata um, die da immer auftreten? Und dann gibt es noch eine speziellere Frage, nämlich: Welche *Form* hat diese Ethik überhaupt? Denn normalerweise macht man ja eine Ethik natürlich für Akteure (*agents*), wie man heutzutage sagt, also verantwortliche Systeme, Systeme, die für ihre Handlungen verantwortlich sein können. Und das würde man ja meistens von technischen Systemen nicht sagen. Wenn mein Laptop irgendeinen Unsinn macht, dann würde ich mich vielleicht darüber ärgern und sagen, »das blöde Ding tut heute nicht, was ich will«, aber eigentlich weiß ich ganz genau, dass das Ding selbst nicht für irgendetwas verantwortlich ist und dass normative Ansprüche ihm gegenüber vollkommen sinnlos sind.

Daher ist die Ethik der KI eigentlich sehr viel besser als eine Ethik *für Menschen* zu beschreiben, die KI herstellen oder benutzen. Die Herstellung oder das Design ist, glaube ich, die Hauptfrage, die da eine Rolle spielt, aber natürlich spielt auch die Benutzung eine Rolle, denn man kann ja ein Werkzeug für einen Zweck herstellen und für einen anderen Zweck benutzen. Es gibt insofern einen erheblichen Unterschied zwischen der KI-Ethik und der gewöhnlichen Ethik für Menschen. Es gibt noch einen weiteren Unterschied, und zwar dass es in der KI-Ethik einen Teilbereich gibt, den man *Maschinenethik* nennt, also die Ethik für die Maschine selbst. Hier gibt es eine Diskussion darüber, inwieweit so etwas überhaupt möglich ist, und wenn es möglich ist, welche Form das annehmen sollte. Ob man in der Maschinenethik zum Beispiel annehmen muss, dass KI-Systeme selbstverantwortliche Akteure sind oder ob man irgendwelche anderen Annahmen machen muss. Es ist wichtig, zu verstehen, dass der Großteil der KI-Ethik *nicht* Maschinenethik ist, denn das wird oft durcheinandergeworfen.

Um das zusammenzufassen: Die KI-Ethik ist eigentlich eine Ethik für Menschen; spezifisch für die Herstellung und den Gebrauch von KI-Systemen durch Menschen. Zusätzlich umfasst sie noch den Aspekt der Maschinenethik.

**MARTIN HÄHNEL:** Danke. In diesem Zusammenhang ist es interessant, auf eine wichtige Unterscheidung zu sprechen zu kommen. Wir sprechen – das hatten Sie ja bereits erwähnt – einerseits von Softwareprogrammen und andererseits von Hardware und im ausgearbeiteten Fall von Robotern. Wie nimmt eine KI-Ethik darauf Bezug und inwieweit ist dieser Unterscheidung Rechnung zu tragen? Vielleicht können Sie darauf noch ein bisschen genauer eingehen.

**VINCENT C. MÜLLER:** Die KI ist eine Wissenschaft, die sich eigentlich nur mit Software beschäftigt. Das stimmt insofern nicht ganz, als man häufig eine Designentscheidung treffen muss, ob bestimmte Aspekte des Verhaltens eines Systems in der Hardware oder in der Software verankert werden. Dennoch ist KI prinzipiell eine Frage der Funktion und nicht der physischen Realisierung; insofern sind KI-Systeme Softwaresysteme. Sie werden aber natürlich implementiert, denn eine Software macht an sich ja erstmal gar nichts: Sie muss irgendwo physisch auf einem Rechner laufen und manchmal kann sie dann auch direkt mit der Welt interagieren. Meistens passiert das über ein sogenanntes Interface; mit anderen Worten, eine Schnittstelle zur physischen Welt, insbesondere zum Menschen. Zum Beispiel zeigt sie uns etwas auf dem Bildschirm an, also mein Telefon zeigt zum Beispiel »16:15«. Diese Anzeige wirkt natürlich in einem minimalen Sinn bereits physisch auf die Umwelt ein. Es gibt darüber hinaus Systeme, die in einem substanzialeren Sinne auf die Umwelt einwirken, nämlich über physische Kräfte. Diese Systeme nennt man *Roboter*. Folglich sind Roboter Systeme, die etwas greifen können, etwas wegschieben können oder, wie autonome Autos, irgendwo hinfahren können.

Die KI-Ethik beschäftigt sich also eigentlich mit der KI in den Robotern, sofern sie sich überhaupt mit Robotern beschäftigt. Es gibt eine interessante Entwicklung in dem Bereich: Die Ethik der Robotik war eine Zeit lang in Mode, weil man sich irgendwie immer einen Akteur wie du und ich in der Welt vorstellen möchte – also mit einem Körper, der sich durch die Welt bewegt, welcher möglichst auch noch menschlich aussieht, wie ein humanoider Roboter. Man hat aber dann relativ bald festgestellt, dass die wirklich interessanten ethischen Fragen nicht mit dieser verkörperlichten

Form der KI zusammenhängen, sondern schlicht mit dem, was die KI macht und warum. Ich glaube auch, dass die ethischen Probleme bei den Systemen sehr viel drängender sind, die keine Roboter sind. Das liegt daran, dass, zumindest gegenwärtig, robotische Systeme weniger Einfluss auf die Welt haben als nicht-robotische.

TEIL I:  
PHILOSOPHIE UND KI

**MARTIN HÄHNEL:** Da wir hier ein dezidiert philosophisches Gespräch führen, bin ich natürlich daran interessiert zu erfahren, was Sie für einen Philosophiebegriff haben, das heißt was Sie allgemein unter Philosophie verstehen?

**VINCENT C. MÜLLER:** Da beginnen wir gleich mit einer notorisch schwierigen Frage. Denn die Frage, was die Philosophie ist und wie man philosophieren sollte, ist ja bekanntlich selbst eine philosophische Frage. Darin unterscheidet sich die Philosophie von anderen Wissenschaften, dass sie diese Arbeit selbst machen muss. Mein grober Begriff von Philosophie ist, dass philosophische Fragen immer dann auftauchen, wenn man eine beliebige Frage bis zum Grunde weiterverfolgt; also immer tiefer gräbt. Wenn man das tut, dann kommt man irgendwann zu Fragen, die sich darum drehen, was ein bestimmter Begriff bedeuten soll. Also man fragt sich zum Beispiel: Was ist Freiheit oder was ist Schönheit? Philosophische Fragen haben eigentlich alle diese Form: Was ist  $x$ ? Und für dieses  $x$  kommt dann in der Regel ein relativ abstrakter Begriff ins Spiel. Diese Art von Fragen sind also eigentlich begriffliche Fragen, nach meiner Auffassung. Begriffliche Fragen sind in der Geschichte der Philosophie natürlich auf verschiedene Arten und Weisen beantwortet worden.

Meine Herangehensweise an solche Fragen ist, dass wir uns erstens darum kümmern müssen, wie diese Begriffe verwendet werden, also so, wie es die klassische *Ordinary Language Philosophy* vorschlägt – das wäre die *Analyse*. Zweitens müssen wir aber darauf schauen, wie es um die Referenten dieser Begriffe steht, wie die Welt beschaffen ist. Und das wird wiederum die Begriffe selbst beeinflussen. Das heißt, man muss die jeweils relevanten wissenschaftlichen Erkenntnisse einbauen, um den Begriff wirklich besser zu verstehen – das wäre die empirische Wissenschaft. So weit handelt es sich bei dieser Auffassung um genau das, was Quine in *Two Dogmas of Empiricism* gefordert hat.<sup>1</sup> Drittens habe ich eine Sympathie dafür, die Begriffe so auszustalten, wie wir sie *brau-*

*chen* – das wäre der pragmatische Aspekt. Man nennt das heutzutage »conceptual engineering«<sup>2</sup>, es ist aber eigentlich nichts Neues. Die Idee ist, dass man bei der Methode der Begriffsanalyse nicht erwarten kann, dass sich die richtige Analyse gewissermaßen vorgefertigt irgendwo finden lässt, sondern man muss die Begriffe in einem gewissen Maße erst für unsere Zwecke formen. Und man kann zum Beispiel zu dem Schluss kommen, dass es eigentlich verschiedene Begriffe gibt, die für verschiedene Zwecke nützlich sind. Ich habe zum Beispiel vorhin den Begriff der Freiheit erwähnt. Es ist relativ klar, dass es einen Begriff von politischer Freiheit gibt, der erhebliche Bedeutung hat. Und es gibt einen Begriff von Willensfreiheit, der eine andere Bedeutung in der Philosophie hat, und man sollte, glaube ich, bei solchen Gelegenheiten erkennen und anerkennen, dass es verschiedene Verwendungen bzw. Zwecke für diese Begriffe gibt, und die Begriffe dann dementsprechend ausformen. Schließlich bin ich der Meinung, dass diese grundlegende Arbeit an den Begriffen, die also analytische, empirische und pragmatische Aspekte hat, auch durch die philosophische Arbeit an der KI gemacht werden kann. Ich nenne das *KI-Philosophie*, weil auch die KI eine sehr nützliche Methode ist, mit der man diese philosophische Arbeit verrichten kann.

**MARTIN HÄHNEL:** Das bringt mich gleich zu einer Anschlussfrage, wenn Sie von KI-Philosophie sprechen. Diese scheint mir ja vielleicht so eine Art neue philosophische Subdisziplin, ja vielleicht sogar neues Paradigma zu sein, oder? Und da ist natürlich die Frage interessant, wie sich eine KI-Philosophie, die Sie vielleicht dann auch gar nicht so gern als Philosophie *der* KI bezeichnen würden, in den Kanon der anderen Disziplinen einordnet. Wie verhält sich ihre Vorstellung von KI-Philosophie zum Beispiel zu klassischen Fragen der theoretischen Philosophie? Sie hatten in Bezug auf den Freiheitsbegriff bereits erwähnt, dass eine KI-Philosophie sich mit solchen tradierten Konzepten auseinandersetzen muss, ohne diese übernehmen zu müssen. Und betrachten wir die Ethik, zu der Sie ja auch arbeiten: Wie positioniert sich da eine KI-Philosophie? Ist sie eine neue Form oder weitere Spielart der angewandten Ethik oder angewandten Philosophie?