Karlsruher Institut für Technologie

Transkript zum Podcast

Von Kühlschranklichtern, KI-Pubertät und Turnschuhen

Thomas Metzinger im Gespräch mit Karsten Wendland

Zur Podcast-Folge

Podcast-Reihe „Selbstbewusste KI“
Folge 5

Erstveröffentlichung: 13.10.2020

Forschungsprojekt:

Abklärung des Verdachts aufsteigenden Bewusstseins in der Künstlichen Intelligenz – KI-Bewusstsein

www.ki-bewusstsein.de

Projektleitung:

Prof. Dr. Karsten Wendland
Institut für Technikfolgenabschätzung und Systemanalyse (ITAS)

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

Förderkennzeichen: 2016ITA202

Herausgeber:

Karsten Wendland, Nadine Lahn, Pascal Vetter

Empfohlene Zitationsweise/Suggested citation:

Wendland, K., Lahn, N. & Vetter, P. (Hg.) (2021). Von Kühlschranklichtern, KI-Pubertät und Turnschuhen. Thomas Metzinger im Gespräch mit Karsten Wendland. Karlsruhe: KITopen.
https://doi.org/10.5445/IR/1000139797

Inhalt

1 Zum Projekt „KI-Bewusstsein“

Zum Projekt

Im Projekt „Abklärung des Verdachts aufsteigenden Bewusstseins in der Künstlichen Intelligenz (KI-Bewusstsein)“ am Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) des Karlsruher Instituts für Technologie (KIT) untersuchen und kartieren wir, welche Gruppen wissenschaftlich, wirtschaftlich und weltanschaulich zu „aufsteigendem Bewusstsein“ in der KI arbeiten. Wir fragen danach, welche Motive, Intentionen und Verankerungen jeweils dahinterstecken und welche Zukunftsszenarien angedacht sind oder auch in Zweifel gezogen werden. Dabei klären wir technische Entwürfe ab und hinterfragen Mythen und Narrative, die in die Welt gesetzt werden und bestimmte Zuschreibungen auslösen.

Die Idee einer „erwachenden“, sich selbst bewusst werdenden Künstlichen Intelligenz hat in den vergangenen Jahren zunehmend Popularität erfahren, u.a. durch verbrauchernahe KI-gestützte Systeme wie *Siri* von Apple und den auf *Alexa* hörenden Smart Speaker, den eingebürgerten Roboter *Sophia* und auch IBMs dialogfähigen *Watson*. Renommierte KI-Akteure melden sich begeistert, mahnend oder warnend zu Wort und stellen die Entwicklung einer „Superintelligenz“ in Aussicht, die alles planetarisch Dagewesene in den Schatten stellen und den Menschen in seinen Fähigkeiten überholen werde.

In der KI-Community ist die Fragestellung zum sogenannten „maschinellen Bewusstsein“ zwar bekannt, aber kaum systematisch erforscht – das mystisch aufgeladene Nischenthema geht mit dem Risiko einher, sich einen wissenschaftlichen Reputationsschaden einzuhandeln. Gleichzeitig nähren KI-Forschung und -Marketing genau diese Mystik, indem sie vermenschlichende Sprachbilder verwenden, die ein aufkeimendes Bewusstsein verheißen, wenn etwa der Roboter „denkt“ oder „fühlt“, das autonome Fahrzeug mit einer „environment perception“ ausgestattet ist oder das Smart Home „weiß“, wie es seinen Bewohnern „helfen“ kann. Hierdurch werden Redeweisen und Narrative aufgebaut, die in der (medialen) Öffentlichkeit Vorstellungen zu einer „bewussten KI“ verbreiten, ohne dass hierzu wissenschaftlich belastbare Aussagen geliefert würden. Auch der transdisziplinäre Dialog zur Frage, was am sogenannten maschinellen Bewusstsein „dran“ sein könnte, ist bislang kaum vorhanden.

An diesem Defizit setzt das Projekt an mit dem Ziel, eine Abklärung zum Verdacht aufsteigenden Bewusstseins in der KI herbeizuführen, indem bestehende Diskurse analysiert, empirisch untersucht, einschlägige Akteure fächerübergreifend zusammengeführt, offene Fragen identifiziert und bearbeitet werden, ein gemeinsames, transdisziplinär tragfähiges Verständnis erarbeitet wird und die Ergebnisse in den öffentlichen Diskurs eingebracht werden.

„KI-Bewusstsein“ im Netz:
Web: www.ki-bewusstsein.de
Twitter: @KIBewusstsein

Institut für Technikfolgenabschätzung und Systemanalyse (ITAS):
http://www.itas.kit.edu/

2 Podcast-Serie „Selbstbewusste KI“

Zur Podcast-Serie

Kann Künstliche Intelligenz ein Bewusstsein entwickeln? Wie könnte das überhaupt funktionieren, und was würde das für uns bedeuten? 12 Folgen, 12 Gespräche mit Expertinnen und Experten und eine gemeinsame Abschlussrunde.

Folge	Titel	Gast
01	Ohne Leben kein Bewusstsein (01:10:29)	Thomas Fuchs
02	Roboter bekommen eine menschliche Aura (59:37)	Andreas Bischof
03	In der heutigen KI ist kein Geist (57:45)	Ralf Otte
04	Die Gründerväter der KI machten sich über Bewusstsein keine Gedanken (01:03:36)	Christian Vater
05	Von Kühlschranklichtern, KI-Pubertät und Turnschuhen (01:08:58)	Thomas Metzinger
06	Wir müssen auf Weitsicht fahren und fragen: Was wäre wenn? (41:31)	Frauke Rostalski
07	Bewusstsein ist eine kausale Kraft und kein cleverer Programmier-Hack (55:24)	Christof Koch
08	Wir müssen Maschinen bauen, die Gefühle haben (39:04)	Antonio Chella
09	Dass Roboter uns Emotionen vorgaukeln, kann sehr wichtig sein (45:06)	Janina Loh
10	Die größte Hoffnung wäre, die Dystopien zu verhindern (52:46)	Joachim Weinhardt
11	Die meisten SF-Romane sind als Warnung gedacht, nicht als Gebrauchsanleitung (55:14)	Andreas Eschbach
12	Roboter werden bald ein Bewusstsein besitzen (34:55)	Junichi Takeno
Bonus	Das große Staffelfinale – Diskussionsrunde zu bewusster KI (01:41:08)	Gesprächsrunde

Verfügbarkeit der Audiodaten

Zu KITopen

KITopen: https://publikationen.bibliothek.kit.edu/serie/649
Anchor.fm: https://anchor.fm/kibewusstsein
Spotify: https://open.spotify.com/show/4BzUdFgR6o74H5sS2ait9z
Apple Podcasts: https://podcasts.apple.com/de/podcast/selbstbewusste-ki/id1530833724

3 Bibliometrische Angaben zur Folge

Teasertext

Für den Philosophen Prof. Dr. Thomas Metzinger von der Universität Mainz ist es keineswegs ausgeschlossen, dass Künstliche Intelligenz irgendwann Bewusstsein haben könnte. Schon jetzt sollten wir uns Gedanken über unsere Verantwortung und zur Ethik machen, etwa in Form einer globalen Charta mit KI-Regeln, die auch das Bewusstseinsthema umfassen. KI-Systeme lernen viel über uns Menschen lernen, müssen aber nicht immer an das Gemeinwohl gekoppelt sein. Es hängt von uns ab, unsere Digitale Souveränität wiederzuerlangen.

Metadaten

Titel:	Von Kühlschranklichtern, KI-Pubertät und Turnschuhen
Dauer:	01:08:58
Erstveröffentlichung:	13.10.2020
Autor:	Karsten Wendland
Gesprächsgast:	Thomas Metzinger
Fragensteller:	Michael Mörike
Redaktion, Aufnahmeleitung und Produktion:	Robert Sinitsyn
DOI Audiofile:	https://doi.org/10.5445/IR/1000124512
DOI Transkription:	https://doi.org/10.5445/IR/1000139797

Folgenlogo

4 Transkription des Gesprächsverlaufs

Zur Podcast-Folge

Karsten Wendland: Hallo und herzlich willkommen bei Selbstbewusste KI, dem Forschungspodcast an der Grenze zwischen Mensch und Maschine. [00:00:10]

Thomas Metzinger: Wir quälen Nutztiere und wir wissen es ganz genau, um bestimmte Ziele zu erreichen. [00:00:15] Warum sollten wir das mit Maschinen nicht machen? [00:00:17]

Karsten Wendland: Mein Name ist Karsten Wendland, ich bin Forscher am Karlsruher Institut für Technologie und gehe Fragen nach wie Technik, die gerade erst noch erforscht wird, morgen vielleicht schon unseren Alltag prägen könnte. [00:00:29] Kann Künstliche Intelligenz ein Bewusstsein entwickeln, wie könnte das überhaupt funktionieren und was würde das für uns bedeuten? [00:00:39] Mein heutiger Gast zu dieser spannenden Frage arbeitet seit vielen Jahren zur Philosophie des Geistes, zur Philosophie der Kognitionswissenschaft, zur Wissenschaftstheorie der Neurowissenschaften mit ihren philosophischen Problemen und auch zur Neuroethik. [00:00:54] Er ist einer der wenigen Philosophen, die mit Technik- und Naturwissenschaftlern gemeinsam und eng in Projekten und Reflexionen zusammenarbeiten und er sagt, eine gute Philosophie muss zur Empirie passen, anstatt sie zu ignorieren. [00:01:08] Er ist hochrangiges Mitglied in Expertengremien und er mischt sich ein, etwa mit der Forderung, der Industrie die Ethik wieder wegzunehmen. [00:01:17] Er ist, man wird es ahnen, Professor für theoretische Philosophie, und zwar an der Universität in Mainz. [00:01:24] Schön, dass wir heute miteinander sprechen können, herzlich willkommen Thomas Metzinger! [00:01:29]

Thomas Metzinger: Schönen guten Tag! [00:01:30]

Karsten Wendland: Herr Metzinger, wie machen Sie das, dass Sie sich selbst als lebend empfinden? [00:01:34]

Thomas Metzinger: Das weiß ich nicht, das weiß ja niemand von uns. [00:01:37] Vielleicht lebe ich ja gar nicht, vielleicht bin ich ja nur eine Simulation in einem großen Rechner von einem Lebewesen. [00:01:46] Jedenfalls ist dieses bewusste Erleben der eigenen Existenz, wenn man es sich genauer überlegt, etwas sehr Mysteriöses. [00:01:58] Morgens wachen wir auf und

kommen wieder zu uns selbst, aber wir machen das nicht, das Aufwachen, das geschieht uns. [00:02:05] In meiner eigenen Sprache heißt das, dass der Organismus, der Sie sind, sein phänomenales Selbstmodell wieder hochfährt. [00:02:13] Also es ist so ähnlich wie der Bootvorgang beim Computer und dann erscheint auch der bunte Desktop mit all seinen Farben.[00:02:20] Also eine Idee, wie wir uns Bewusstsein vorstellen können, ist also eine Benutzeroberfläche auf dem eigenen Gehirn, die manchmal angeschaltet ist und manchmal ausgeschaltet ist. [00:02:31] Und auch interessant ist es, dass wir jeden Abend wieder ins Bett gehen und das Bewusstsein verlieren, wir wissen dann nicht, dass wir jemals existiert haben und jemals wieder existieren werden, aber haben komischerweise keine Angst davor. [00:02:46] Vorm Sterben haben wir Angst, aber ins Bett gehen wir jeden Abend mit tiefem Vertrauen und geben all das auf, wonach Sie eben gefragt haben, das Erleben der eigenen Lebendigkeit. [00:02:59]

Karsten Wendland: Und wir erwarten, dass wir am nächsten Morgen wieder aufwachen und der oder die Gleiche sind in der gewohnten Umgebung und das sich nichts geändert hat. [00:03:07] Das sind die Erfahrungen, die wir gemacht haben und die stellen wir auch gar nicht so infrage. [00:03:10]

Thomas Metzinger: Nein, und die Frage ist auch, woher wissen wir das eigentlich, also dass man denkt, man sei dieselbe Person wie gestern, heißt ja gar nicht, dass man sie auch wirklich ist. [00:03:22] Philosophen nennen das das Problem der Personalidentität und jeder kennt das mit älteren Leuten wie mir selbst zum Beispiel, die immer noch nicht verstanden haben, wie alt sie schon sind und immer irgendwie noch meinen, sie wären ganz jugendlich. [00:03:38] Menschen können sich auch über ihre Personalidentität täuschen. [00:03:39] Manchmal, das wissen wir auch zum Beispiel bei schweren Geisteskrankheiten, kann man völlig wahnhafte Vorstellungen darüber entwickeln, wer man ist und die auch fest glauben. [00:03:54] Dann kann man aber auch zum Beispiel in der Demenz die Erinnerung daran, wer man war, vollständig verlieren. [00:04:02] Und das, wonach Sie ganz am Anfang gefragt haben, Herr Wendland, dieses Gefühl am Leben zu sein, das kann man in seltenen psychiatrischen Störungsbildern auch verlieren, nämlich im sogenannten Cotard-Syndrom. [00:04:18] Es gibt tatsächlich Leute, die bei bestimmten Störungsbildern das robuste Erleben haben, dass sie tot sind, dass ihr Körper tot ist. [00:04:27]

Karsten Wendland: Sie haben eben den Begriff des Bewusstseins erwähnt, der in diesem Zusammenhang eine große Bedeutung zu haben scheint. [00:04:35] Wo ist denn Ihrer Einschätzung nach das Bewusstsein zu finden, oder wo sollte man danach suchen? [00:04:40] Wir leben ja in einer sehr gehirnfixierten Welt momentan, zum Bewusstsein ist oft der Frontalbereich des Gehirns im Gespräch, andere suchen weiter hinten im Kopf, manche arbeiten mit Rückkopplungsschleifen, die sie auch versuchen, in KI-Systemen und in Hardware zu implementieren. [00:04:57] Und seit einigen Jahren ist nun sogar auch der Darm im Gespräch und vom Darmbewusstsein oder vom Darmhirn die Rede, ja bis hin zu Überlegungen, dass manche sagen, das Bewusstsein, das kommt von außen, das durchdringt uns, solange wir leben – wenn wir nicht mehr leben, ist es irgendwann wieder weg. [00:05:15] Können Sie das mal für uns sortieren, wo ist denn das Bewusstsein und welche dieser Ansätze sollten wir besser verwerfen? [00:05:22]

Thomas Metzinger: Vor 350 Jahren, da hat man gedacht, das war der kartesianische Dualismus, dass der Geist, das Bewusstsein sich dadurch auszeichnet, dass er keine räumliche Ausdehnung hat. [00:05:33] Körperliche Dinge kann man teilen, durchschneiden, die sind im Raum, und das Bewusstsein ist eben nirgendwo im Raum, der Geist ist nirgendwo im Raum. [00:05:44] Und das war der Grund, warum ich weiter Philosophie studiert habe. [00:05:48] Im fünften Semester war ich nämlich völlig verzweifelt und fand das alles völlig blöd und auch die Professoren ziemlich eingebildet und dann habe ich mal ein Seminar über die Leidenschaften der Seele, ein Buch von Descartes gemacht, und da hat mir der junge Privatdozent etwas erklärt, was ich noch nicht verstanden hatte bis dahin, dass nämlich, wenn der Geist nicht im Raum ist, dass es dann auch gar keinen Ort gegeben kann im Gehirn oder anderswo, wo er die materielle Welt berührt [Quellenverweis 1]. [00:06:17] Descartes hat gemeint, das wäre in der Zirbeldrüse, wo die Lebensgeister die Seele drücken und stoßen, und da habe ich dann später auch eine Doktorarbeit über das Leib-Seele-Problem geschrieben, da habe ich dann doch wieder Feuer für die Philosophie gefangen, weil wie soll man sich das vorstellen, dass etwas, das nicht im Raum ist, etwas beeinflusst wie das Gehirn, was im Raum ist. [00:06:41] Heute ist das ganz anders, heute sind fast alle Philosophen, die in dem Bereich arbeiten, Materialisten der einen oder anderen Art, die streiten sich nur noch darüber, welche Form des Materialismus die richtige ist und ein ganz wichtiger Begriff ist heute das minimal

hinreichende neuronale Korrelat des Bewusstseins. [00:07:03] Was heißt das? Das heißt, dass viele annehmen, dass es für das bewusste Erleben als Ganzes, aber zum Beispiel auch für eine Rotempfindung oder das süße Geschmackserlebnis, das ich habe, wenn ich Schokolade esse, einen Bereich im Gehirn gibt, der diesem Erleben entspricht, korreliert und dass es einen Bereich gibt, der so klein ist, dass man ihn nicht mehr kleiner machen kann, minimal hinreichend, um dieses Erleben zu erzeugen. [00:07:32] Das heißt, wenn man da eine Elektrode reinstecken würde ins Gehirn, würde man ein Roterlebnis erzeugen. [00:07:30] Wenn man diesen Bereich kaputtmachen würde oder durch ein Narkosemittel unterdrücken würde in seiner Aktivität, dann würde das Roterleben oder das Bewusstsein verschwinden. [00:07:52] Das heißt, viele Leute suchen jetzt die hinreichenden Bedingungen im Gehirn für bestimmte Bewusstseinsinhalte oder für Bewusstsein als Ganzes. [00:08:02] Und einerseits ist es so, dass wir da große Fortschritte gemacht haben, da gibt es viel, auf der anderen Seite ist es so, dass wir das einfach nicht wissen, und dass es einen Wettstreit zwischen ganz verschiedenen Modellen gibt. [00:08:17]

Karsten Wendland: Also das Rätsel ist noch da, es gibt viele Ansätze und die Grundsatzfragen sind uns erhalten geblieben. [00:08:23]

Thomas Metzinger: Das würde ich nicht sagen. [00:08:26] Die Grundsatzfragen haben sich verwandelt in viele präzisere Fragen, also aus dem, was wir für ein Problem gehalten haben, das Bewusstseinsproblem, ist auch nicht ein philosophisches oder wissenschaftliches Problem, sondern es ist ein ganzes Bündel von Problemen. [00:08:44] Also wenn man zum Beispiel dann fragt: Haben Fische Bewusstsein oder können Maschinen Bewusstsein haben, da muss man immer nachfragen: Was denn jetzt genau, Farben sehen, Gefühle, Ich-Gefühl, Erinnerung, was genau meinen wir damit? [00:09:04] Das heißt, eigentlich ist der wissenschaftliche Fortschritt so verlaufen, dass man manchmal gar nicht mehr weiß, worüber man eigentlich spricht, weil die ursprünglichen Fragen, die Urfragen, die Anfangsfragen sehr viel feinkörniger gestellt werden auf ganz viele verschiedene Arten heute. [00:09:22]

Karsten Wendland: Wir haben auch im heutigen Gespräch wieder einen Fragensteller, der sich Ihnen kurz vorstellt und einige Spezialfragen an Sie hat. [00:09:31]

Michael Mörike: Mein Name ist Michael Mörike, ich bin Vorstand der Integrata-Stiftung für humane Nutzung der IT. [00:09:38] Ich habe Kernphysik studiert und bin seit 1964 in der IT tätig, die damals EDV hieß. [00:09:48] Für die Integrata-Unternehmensberatung habe ich vordem große technische Projekte geleitet, darunter auch Teile zur Entwicklung von BTX, dem Vorläufer des heutigen Internet. [00:10:01] Und nun meine Fragen, Herr Professor Metzinger: Kann es Bewusstsein ohne Selbstbewusstsein geben, haben Sie sich das auch schon mal gefragt? [00:10:11]

Thomas Metzinger: Ich bin mir ganz sicher, dass es Bewusstsein ohne Selbstbewusstsein geben kann, aber da gibt es ein paar Stolpersteine, über die man nachdenken muss. [00:10:23] Erst mal muss man natürlich wissen, was Selbstbewusstsein eigentlich ist, das hat beim Menschen ganz viele Bestandteile. [00:10.32] Es gibt zum Beispiel das körperliche Selbstbewusstsein, das emotionale Selbstbewusstsein, es gibt die Fähigkeit, Ich-Gedanken zu denken oder das Personalpronomen der ersten Person richtig zu verwenden, das heißt Ich zu sagen, um auf sich selbst Bezug zu nehmen. [00:10:50] Das ist vielleicht etwas, was die Maschine auch könnte. [00:10:53] Und dann gibt es aber ein etwas tieferes Problem: Wir können uns Bewusstsein ohne Selbstbewusstsein nicht vorstellen und das hat viele Philosophen dazu verleitet, diesen Fehlschluss zu machen, dass etwas, was man sich nicht vorstellen kann, auch logisch unmöglich ist. [00:11:17] Warum kann man sich das nicht vorstellen? Na, weil einfach der Versuch, das vorzustellen, schon ein kleines Gefühl einer geistigen Anstrengung mit sich bringt, jemanden, der das will, jemand, der versucht, einen Ich-losen Zustand zu simulieren, und dann haben wir natürlich schon wieder ein kleines Ego da drin. [00:11:37] Natürlich haben sich über die Jahrhunderte viele Philosophen versucht rauszudenken, aus ihrem Bewusstsein und die absolute Außenperspektive einzunehmen, den Gottesstandpunkt, aber das kann man halt nicht willentlich machen. [00:11:53] Trotzdem gibt es eine große Klasse von Bewusstseinszuständen, zum Beispiel bei mystischen Erfahrungen, in denen das Ich-Gefühl verschwindet. [00:12:02] Dazu gehören zum Beispiel spirituelle Erlebnisse vom sogenannten nicht dualen Bewusstsein oder religiöse Einheitserfahrungen, Ganzheitserfahrungen. [00:12:14] Solche Erfahrungen treten auch häufig auf bestimmten psychoaktiven Substanzen auf, wie Psilocybin oder LSD. [00:12:21] Darüber gibt es viel Wissenschaft, insbesondere gibt es sehr viel neue Wissenschaft, die Forscher nennen das Ego-Dissolution, das heißt, sie

untersuchen den Vorgang der Ich-Auflösung, schieben Leute in den Hirnscanner. [00:12:37] Und ich selbst habe ein großes Forschungsprojekt gegründet, das MPE-Projekt an der Universität Mainz, das läuft seit anderthalb Jahren, wo wir die Erfahrung reinen Bewusstseins in der Meditation untersuchen und da werden wir vielleicht später sehen, dass das mit unserem Thema ganz viel zu tun hat, das heißt, die Frage ist: Was ist eigentlich die einfachste Form von Bewusstsein? [Quellenverweis 2] [00:13:01] Können wir so was wie ein Minimalmodell von Bewusstsein bauen? [00:13:06] Und da stellt sich halt heraus, um Bewusstsein zu haben, muss man kein Zeitempfinden haben, es gibt ganz klare Zustände in der Meditation ohne Zeiterleben. [00:13:19] Man muss dazu auch keine Gedanken und keine Gefühle haben, auch das Körperbewusstsein, das Gefühl im Raum lokalisiert zu sein, kann vollständig abwesend sein. [00:13:31] Damit habe ich mich die letzten Monate sehr intensiv beschäftigt, wir haben zum Beispiel psychometrische Studien mit Meditierenden gemacht und es ist ganz klar, dass es so etwas wie reines Bewusstsein ohne Selbstbewusstsein gibt. [00:13:46] Natürlich kann man, wenn dieser Zustand da ist, nicht über ihn berichten, weil allein der Versuch mal kurz hinzugucken, ja, also das ist so ähnlich, ich nenne das immer das Kühlschranklichtproblem. [00:13:59] Sie wissen, Sie haben den Kühlschrank zugemacht, das Licht ist aus, aber so ganz glauben Sie es nicht, und dann ganz vorsichtig öffnen Sie die Tür, um zu gucken, ob das Licht wirklich aus ist und dann geht es natürlich immer sofort an. [00:14:12] Und genau so geht es Ihnen natürlich, wenn Sie glauben, es gibt Bewusstsein ohne Selbstbewusstsein, aber Sie wollen immer mal kurz gucken, das funktioniert nicht, da geht das Kühlschranklicht an. [00:14:24] Und trotzdem wissen wir natürlich von den Mystikern, von den Mönchen, von den Heiligen, von den Meditierenden aller Kulturen und aller Jahrhunderte, dass es Bewusstsein ohne Selbstbewusstsein gibt. [00:14:38] Nur so ein paar Kollegen von mir, die ein eingeschränktes Vorstellungsvermögen haben, die meinen, Selbstbewusstsein wäre eine notwendige Bedingung für Erleben. [00:14:49]

Karsten Wendland: Sie haben eben das Kühlschranklicht erwähnt, das hatte auch einige mysteriöse Eigenschaften, die man so landläufig nicht kennt, ich plaudere jetzt mal aus dem Nähkästchen. [00:14:58] Ich habe mit einem Hersteller von Kühlschränken gesprochen, der gesagt hat, dass das Licht sogar manchmal dann brennt, wenn die Tür zu ist. [00:15:07] Das ist etwas, was die meisten nicht wissen,- (**Thomas Metzinger:** Da haben wir es schon.) -weil es

dann als Wärmequelle verwendet wird, um die Temperatur im Kühlschrank zu regeln, also eine Sonderfunktion, die Lampe kann auch an sein, wenn man nicht hinguckt (**Thomas Metzinger:** Das ist ja interessant). [00:15:21] Und das führt jetzt auch zu der nächsten Frage von Michael Mörike, der sich Gedanken darüber macht, welche emotionalen Zustände denn Maschinen haben könnten. [00:15:32]

Michael Mörike: Könnten Maschinen stolz sein? [00:15:36]

Thomas Metzinger: Ja, da weiß man natürlich nicht, was das jetzt heißt, heißt das, dass Maschinen ein bisschen eingebildet sein könnten, also Stolz im schlechten Sinne? [00:15:47] Oder dass sie, das fände ich viel interessanter, Selbstachtung vielleicht im Sinne von Immanuel Kant besitzen könnten, indem sie sich selbst einen Wert zuschreiben? [00:15:59] Und in der Tat halte ich das für ein Risiko, das überhaupt nicht wahrgenommen worden ist in der aktuellen Diskussion um die KI-Ethik. [00:16:10] Es könnte ja sein, dass, und das ist jetzt eine etwas längere Geschichte, dass eine Maschine sich selbst, ihrem eigenen Leben, ihrer eigene Existenz, leben tut sie ja nicht, einen hohen Wert zuschreibt. [00:16:26]

Karsten Wendland: Also Sie würden sagen, die Maschine lebt als solches nicht, also ist-. [00:16:29] (**Thomas Metzinger:** Nein, Maschinen sind keine Lebewesen.) [00:16:31] Sie ist kein Lebewesen? Okay, ich wollte es nur noch mal klären, ich wollte es kurz noch mal geklärt haben, wir machen ja einen großen Bogen und aber dennoch könnte sie sich selbst ein Ich zuschreiben und möglicherweise auch einen Wert, sagen Sie, im Sinne eines positiven Wertes, Würde, Selbstachtung und so weiter zuschreiben. [00:16:52] Dann wäre da auch eine gewisse Empfindlichkeit im Spiel, denn diese Würde wäre auch verletzbar. [00:16:55]

Thomas Metzinger: Ja, das ist jetzt eine sehr schwierige Frage, ich fange mal lieber von vorne an. [00:17:03] Also, was wir anständig behandeln müssen, was die Philosophen sagen, ein Gegenstand ethischer Überlegungen ist, ist in einer ersten Annäherung alles, was leiden kann. [00:17:15] Das heißt, wir müssten eine Theorie über bewusstes Leiden haben. [00:17:19] Das wäre zum Beispiel in der Tierethik auch so wichtig, weil es völlig klar ist, dass die Art und Weise, wie wir mit Tieren umgehen, ethisch nicht haltbar ist.

[00:17:28] Die meisten Tiere, die wir essen, sind leidensfähige, selbstbewusste Kreaturen und so was Ähnliches könnte uns mit Maschinen auch passieren. [00:17:36] Ich nenne jetzt mal vier notwendige Bedingungen, die eine Maschine haben müsste, um leiden zu können. [00:17:45] Die Erste ist die B-Bedingung, die Bewusstseinsbedingung, die müsste wach sein, die müsste Bewusstsein haben, wir wissen im Moment nicht, was es bedeuten würde, wir haben keine Theorie. [00:17:56] Die zweite notwendige Bedingung wäre die S-Bedingung, die Selbstbewusstseinsbedingung, die Maschine müsste ein Ich-Gefühl haben. [00:18:05] Ich glaube, das werden Maschinen sehr schnell entwickeln, zum Beispiel indem sie Roboterkörper steuern, Bewegung auf Wahrnehmung abbilden, Maschinen mit Selbstmodellen gibt es schon lange, darüber habe ich auch viel geschrieben. [00:18:18] Die dritte Bedingung wäre die V-Bedingung, das heißt, die Maschine müsste Valenzen darstellen können, das heißt Werte. [00:18:29] Es müsste Zustände geben, bei denen sie sagt: Der Zustand hat einen positiven Wert, zum Beispiel „Mein Akku ist voll aufgeladen, das ist gut für meine Funktionsfähigkeit und meine weitere Existenz“, oder negative Valenz, „Ich habe einen Hardwareschaden, den ich noch nicht verstanden habe.“ [00:18:48] Und diese Valenzen müssten in das Selbstmodell eingebettet werden. [00:18:54] Bei uns ist es ja so, bei uns biologischen Wesen, dass Emotion, die Wahrnehmung darum, ob etwas gut oder schlecht für das Überleben von uns und unseren Kindern ist, ins Körperempfinden eingebettet ist. [00:19:09] Also uns sinkt das Herz in die Hose, uns fährt der Schreck in die Glieder uns friert das Blut in den Adern, das heißt, für uns ist es immer ein räumliches, körperliches Erleben, eine Emotion zu haben. [00:19:22] Jetzt ist das Erste, was man verstehen muss, Maschinen könnten ganz andere Arten von Emotionen haben als wir, welche, die wir uns nicht vorstellen können. [00:19:33] Zum Beispiel, wenn sie andere Körper haben mit anderen Sinneswahrnehmungen, anderen Arten des Kaputtgehens und so weiter und so fort. [00:19:42] Und vierte Bedingung, die ganz entscheidend ist, die ist ein bisschen komplizierter zu verstehen, das ist die Transparenzbedingung. [00:19:50] Bei uns selbst ist es so, der ganz gemeine Trick der Evolution bei uns war, dass unser Selbstmodell transparent ist, das heißt, wir erkennen es nicht als ein inneres Bild von uns selbst. [00:20:04] Wir kleben sozusagen da dran, wir identifizieren uns mit dem Inhalt unseres Selbstmodells und das ist eine ganz gemeine Erfindung. [00:20:13] Wenn das Tier nämlich Hunger hat, Durst hat, Schmerzen hat, Angst empfindet, empfindet es es sozusagen ganz direkt und unvermittelt im eigenen Körper und kann

sich nicht davon distanzieren. [00:20:27] Das heißt, in der biologischen Evolution war dieses Leiden, dieses Identifiziertsein mit einem Körper und seinen Gefühlen ein ganz gemeiner und sehr wirksamer Trick, die Tiere vorwärts zu treiben, dass sie nach Lösungen suchen, dass sie sich schützen, dass sie aufpassen und so weiter, dass sie solche Zustände verhindern. [00:20:47] Und diese vier Bedingungen, ich sage nicht, dass das hinreichende Bedingungen sind, aber man kann sich meiner Meinung nach sehr gut vorstellen, dass die in der Maschine realisiert werden. [00:21:00] Also wie gesagt, Roboter mit Selbstmodell gibt es heute schon und dass man denen Valenzen einbaut, positive und negative Bewertungsfunktionen, auch überhaupt kein Problem, und wenn sie sich damit identifizieren, das heißt, wenn in der Maschine das Gefühl auftaucht, das ist meine eigene Angst, das ist mein eigener Schmerz, das ist meine eigene Verzweiflung, dann hätten wir tatsächlich so was wie bewusstes Leiden und das ist auch das, was wir auf jeden Fall verhindern sollten. [00:21:34] Was ich jetzt eben gesagt habe, das hat aber interessante Konsequenzen. [00:21:39] Wenn eine von diesen notwendigen Bedingungen nicht gilt, dann brauchen wir uns auch keine Sorgen zu machen. [00:21:46] Das heißt, wenn die Maschinen zum Beispiel kein Ich-Gefühl hätten, dann können sie auch nicht leiden, das ist die tiefste Erkenntnis der buddhistischen Philosophie von 2500 Jahren. [00:21:58] Wenn die Maschinen ein Ich-Gefühl und Bewusstsein hätten, aber wie manche Philosophen früher gesagt hatten, ihre eigene Existenz sie nichts angeht, das heißt, sie einfach keine Wertigkeit für sich selbst, aber auch keine Gefühle, die es ganz kalt und rational hinnehmen, wenn ihr Körper kaputtgeht, die zum Beispiel nicht wie Lebewesen eine Angst vor dem Tod hätten, dann könnte es doch sein, dass solche Systeme aber Bewusstsein, Selbstbewusstsein haben und nicht leiden wie wir Menschen und Tiere auch auf dem Planeten. [00:22:35] Jetzt kann man sich natürlich fragen, Sie haben eben am Anfang nach Emotionen gefragt. [00:22:40] Es haben ja wahrscheinlich auch nicht alle Tiere Emotionen auf dem Planeten, ich gehe jetzt mal so als Laie davon aus, dass Insekten keine Gefühle haben und auch Reptilien nicht. [00:22:56] Ich gehe mal davon aus, dass die Tiere, die Brutpflege hatten, Mutter-Kind-Bindung, bestimmte Arten von Wirbeltieren, dass die Emotionen entwickelt haben über die Mutter-Kind-Bindung im Wesentlichen, die das sind, was wir Menschen heute Gefühle nennen. [00:23:17] Und da Maschinen überhaupt nicht aus so einer Evolution stammen, könnte es ja sein, dass manche Maschinen eher sind wie Insekten, ja, so viele Roboter sind ja sogar Insektenkörpern nachempfunden

in der biologischen Intelligenz, oder so wie Reptilien. [00:23:37] Oder wenn Sie so einer Eidechse in ihr kaltes Auge schauen, da merken Sie schon die Intelligenz, die Eidechsenintelligenz ist irgendwie ganz anders als unsere, irgendwas ist da nicht. [00:23:50] Ja. Auch wenn sie ihren Schwanz verliert, weil eine Katze mit ihr spielt, sie hat zwar diese motorischen Fluchtreflexe, aber möglicherweise gibt es da drin keinen emotionalen Stress im engeren Sinne. [00:24:05] Und diese ganzen Fragen, die wir jetzt so ein bisschen umkreisen, die sind natürlich ganz wichtig für die KI-Ethik und für die Maschinenethik. [00:24:15]

Karsten Wendland: Sie vertreten ja unter anderem auch die Position des negativen Utilitarismus. [00:24:23] In dem geht es darum, das Leid und Leiden auf der Welt zu vermeiden im Jetzt und auch in der Zukunft. [00:24:30] Wenn jetzt eine Maschine, ein Roboter leiden würde in der Art, wie Sie es eben beschrieben haben, könnte es nicht sein, dass dieses Leiden, das der Roboter sich selbst zuschreibt, letztlich auch nur eine interne Simulation ist, also dass der Roboter sich selbst gegenüber so tut, als ob und sich zwar etwas zuschreibt, aber letztlich das ohne Wirkung bleibt. [00:24:58] Wir hätten es dann also mit einer Imitation von Leiden zu tun. [00:25:02] Ist das nicht auch eins dieser Probleme, bei denen man genauer hinschauen muss? Nicht, dass wir am Ende unseren Zuschreibungen bezogen auf Roboter selbst auf den Leim gehen. [00:25:11]

Thomas Metzinger: Ja, das ist ein großes Problem, das Sie nennen und auch ein sehr Wichtiges. [00:25:17] Das hat der amerikanische Philosoph Hilary Putnam schon Anfang der 60er-Jahre in einer Serie von neun Aufsätzen genau diskutiert, also wann wir, er war der Erste, der gefragt hat, wann sollten wir Robotern Bürgerrechte zuschreiben und hat dann erst mal gesagt, da kommen ja bei normalen Leuten, kommt also die ganz langweiligen Standardantworten sind immer: „Ja, aber die sind ja gar nicht lebendig und die haben ja gar keine Gefühle und so.“ [Quellenverweis 3] [00:25:44] Und da hat er schon damals gesagt, also das ist, wie man sagt, die haben schwarze Hautfarbe, andere Menschen, die sollten keine Bürgerrechte haben, das sind Hardwarekriterien. [00:25:53] Und jetzt ist natürlich die Frage, was ist der Unterschied zwischen einem Wesen, das Schmerzverhalten sehr erfolgreich simuliert, Philosophen nennen so was manchmal auch einen Zombie, also etwas, was funktional isomorph zu mir ist, das sogar schreien würde, aber es

gibt diesen Innenaspekt nicht, diesen phänomenalen. [00:26:14] Das führt natürlich zu noch weiterführenden Fragen: Woher wissen wir eigentlich, dass unser eigener Schmerz nicht nur eine Simulation ist und wir Maschinen sind, die sich dauernd selbst täuschen, aber das lasse ich jetzt noch mal beiseite. [00:26:30] Natürlich können wir Systeme bauen, die Schmerzverhalten simulieren und auch die Intelligenz, die durch Leiden entsteht. [00:26:40] Ich mache mal einen kleinen Schlenker, um das zu illustrieren. [00:26:44] Ich habe die letzten beiden Jahre in der europäischen High-Level Expert Group for Artificial Intelligence gearbeitet, wir hatten unsere letzte Sitzung vorgestern [Quellenverweis 4]. [00:26:56] Wir haben unter anderem Ethikrichtlinien für den Umgang mit Künstlicher Intelligenz in Europa entwickelt, von denen ich selbst als Ethiker, der da mitgearbeitet hat, sehr enttäuscht bin. [00:27:08] Und eines der Themen, das in dem Dokument stand, das schon letztes Jahr veröffentlicht worden ist, war künstliches Bewusstsein, genau das Thema, worüber wir hier reden. [00:27:17] Und das ist von einer Mehrheit der Vertreter komplett unterdrückt worden, also alle Passagen sind rausgelöscht worden aus dem Dokument, überwiegend auf Drängen der Wirtschaftslobby, die natürlich keine-, also die Leute nicht verunsichern will, die will ihre Märkte entwickeln, wir sollen alle diese Produkte kaufen und will nicht, dass da komische philosophische Diskussionen auftauchen, wachen die Dinger irgendwann auf oder so. [00:27:46] Und deswegen wurde dieses Thema komplett unterdrückt wie zwei, drei andere Themen auch. [00:27:51] Und da gab es aber eine interessante Diskussion, wo ich gesagt habe, na ja, wir sollten überhaupt auch nicht nur riskieren, dass mit europäischen Forschungsgeldern mal künstliches Bewusstsein erzeugt wird, solange wir nicht wissen, was wir da eigentlich tun und dass wir möglicherweise künstliches Leiden erzeugen. [00:28:11] Dann sagte der eine: Wieso? Wir quälen und bestrafen Tiere doch auch, damit sie schneller lernen und wenn ich den Dingern eine steilere Lernkurve verpassen kann, indem ich sie auch noch ein bisschen quäle und bestrafe, dann würde mich das interessieren. [00:28:26] Da könnten wir doch effektivere KI bauen, effektivere Roboter, wir bestrafen ja auch unsere Kinder und erziehen die, und wir quälen Nutztiere, und wir wissen es ganz genau, um bestimmte Ziele zu erreichen. [00:28:38] Warum sollten wir das mit Maschinen nicht machen? [00:28:41] Und das erzähle ich deswegen, damit man sieht, es gibt da auch eine ganz andere Einstellung zu. [00:28:47] Mir ist es sehr wichtig, dass wir überhaupt nicht erst einsteigen in so einen unkontrollierten Prozess einer

Evolution zweiter Stufe, der möglicherweise einen Ozean von Leiden erzeugen könnte, den es vorher nicht gegeben hat über sehr viele Kopien, oder der dazu führen könnte, dass wir Formen des bewussten Leidens in Maschinen erzeugen, die wir selbst überhaupt noch nicht verstanden haben, wo wir vielleicht überhaupt erst nach einiger Zeit entdecken, dass da schon ein Prozess am Laufen ist, der an uns vorbeigegangen ist. [00:29:22] Das sind ja alles Risiken, die bedacht werden müssen. [00:30:27]

Karsten Wendland: An der Stelle haben Sie auch eine klare Position und fordern ein Moratorium für synthetisches Bewusstsein und sagen, lasst mal für 30 Jahre die Finger von diesem Thema und entwickelt nicht in diese Richtung, um zu verhindern, dass nicht plötzlich doch aus irgendwelchen Gründen, weil die notwendigen Bedingungen erfüllt sind, die Sie erwähnt haben, plötzlich ein leidensfähiges synthetisches Wesen entsteht. [00:29:50] Da sind Sie ja sehr streng. [00:29:51]

Thomas Metzinger: Ja, ich habe das seit einigen Jahren schon gesagt, immer mal so angedeutet auch in diesem deutschen Buch „Der Ego-Tunnel“, dass wir jetzt uns nicht auf so eine schräge Bahn begeben sollten, einfach weil bestimmte Forscher ein Interesse daran haben, damit berühmt zu werden und das muss ich jetzt noch ein bisschen erläutern [Quellenverweis 5]. [00:30:16] Also ich bin selbst in der Bewusstseinsforschung ganz aktiv seit über 30 Jahren drin, ich bin einer, der die Association for the Scientific Study of Consciousness gegründet hat, und mich beunruhigt es etwas, dass es jetzt vier Labore auf der Welt schon gibt, Kollegen von mir, die ich sehr respektiere und hochschätze: [00:30:39] Stan Dehaene in Paris, Naotsugu Tsuchiya in Melbourne in Australien, Ryota Kanai in Tokio und Michael Grazianos Gruppe in Princeton, die ganz explizit sagen: Wir fänden das total cool, künstliches Bewusstsein zu machen, und wir würden es auch sofort tun, wenn wir wissen wie es geht. [00:30:56]

Karsten Wendland: Wie man das macht. [00:30:55]

Thomas Metzinger: Das möchte ich nicht, weil wir, glaube ich, die Folgen einer solcher historischen Entwicklung gegenwärtig überhaupt nicht übersehen können. [00:31:07] Ich muss dazu aber was sagen, um das zu verdeutlichen, ich selbst als jemand, der sich mit dem Problem des Bewusstseins seit über 30 Jahren sehr intensiv beschäftigt hat, glaube nicht, dass das morgen

passieren wird, und glaube auch nicht, dass das übermorgen passieren wird. [00:31:25] Also ich halte das künstliche Bewusstsein für etwas, was es in weit entfernter Zukunft geben wird oder vielleicht gar nicht, aber der Punkt ist, ich kann mich ja täuschen. [00:31:37] Es gibt aus der Wissenschaftsgeschichte einige Beispiele, zum Beispiel bei der Kernspaltung, wo die Experten, die den Durchbruch geschafft haben, sechs Monate vorher noch Medienvertretern gesagt haben, das schaffen wir entweder nie, aber frühestens in fünf Jahren. [00:31:54] Und dann sind plötzlich unerwartet ein paar Sachen zusammengekommen, Synergien nennt man das aus verschiedenen Wissenschaftszweigen. [00:32:02] Gerade die Pioniere, die geglaubt haben, wir kriegen das nicht hin, die sind dann die, die auf einmal schneller das schaffen. [00:32:11] Und so was könnte hier auch passieren, wir haben große Fortschritte in der mathematischen Modellierung des menschlichen Gehirns, es gibt gute theoretische Überlegungen zu Bewusstsein, es gibt Fortschritte in der KI. [00:32:24] Es könnte sein, dass ganz unerwartet unser kluger Doktorand in China das alles auf die richtige Weise zusammenfügt, und dann ist dieser historische Durchbruch plötzlich da, unerwartet für alle. [00:32:38] Deswegen ist es wichtig, dass wir uns jetzt schon über die Ethik Gedanken machen. [00:32:45] Die EU-Kommission in Brüssel wollte das nicht, die hat dieses Thema ganz absichtlich unterdrückt wie zwei, drei andere Themen auch. [00:32:54] Und ich fordere jetzt einfach, dass wir erst mal auf die Bremse treten, dass zum Beispiel in Europa keine Forschungsgelder ausgegeben werden sollten für jede Art von Forschung, die auch nur riskiert, also die nicht nur-, die nicht direkt künstliches Bewusstsein anstrebt, sondern die riskiert, dass es aus Versehen entsteht, und dass wir erst mal ganz in Ruhe darüber nachdenken sollten, was denn hier vielleicht riskant ist, was sicher ist, was wir alle wollen. [00:33:29] Und deswegen denke ich, bis 2050 sollten wir das einfach mal lassen. [00:33:34] Ein bisschen beruhigt mich, dass ich glaube, meine berühmten Kollegen, die alle darüber reden, machen eigentlich nur Wind, weil sie Forschungsgelder anziehen wollen. [00:33:45] Ich glaube nicht, dass sie das im Ernst bald hinkriegen, aber man weiß ja nie. [00:33:51]

Karsten Wendland: Angenommen, wir würden jetzt in Europa Regelungen finden und tatsächlich vermeiden, dass ein künstliches Bewusstsein aus Versehen entstehen könnte. [00:34:01] Was löst das denn in Ihnen als Ethiker aus, wenn Sie wissen, dass es andere Forschergruppen auf der Welt überhaupt nicht juckt, was wir hier in Europa uns an ethischen Vorgaben selbst

setzen, die machen selber weiter? [00:34:15] Sie erwähnten ja schon den Wettbewerb, den es in dieser Richtung gibt, wie ist das für Sie als Ethiker? [00:34:20]

Thomas Metzinger: Ein Risiko ist das, was man in der Ethik das „race to the bottom“ nennt, also ein Wettrennen um die niedrigsten ethischen Standards. [00:34:30] Das kann man sich so ähnlich vorstellen wie bei der Steuerflucht, dass reiche Leute aus Deutschland ihr Geld in Sicherheit bringen irgendwo in Panama oder sonst wo, weil sie das Gemeinwohl schädigen wollen damit. [00:34:46] Genauso ist es natürlich zum Beispiel auch in dem militärischen Problem eines KI-Wettrüstens, ganz großes Problem, was vielleicht damit auch zusammenhängt, wenn wir hier Forschung verbieten, das ist genauso wie mit Primatenexperimenten, Affenexperimenten, dann hat man das Risiko, dass die Forschung abwandert in Bereiche nach China oder in die USA, wo wesentlich laxere Standards gelten. [00:35:18] Ich selbst habe mich ja auch bemüht, weil ich glaube, dass es eigentlich nur was bringen würde, wenn wir eine globale Charta hätten mit Regeln für die Künstliche Intelligenz, aber ich erinnere mich noch sehr gut, ich habe letztes Jahr einen Vortrag in Washington gehalten und mit ein paar hochrangigen Juristen und anderen Leuten da geredet und die haben mir ganz offen ins Gesicht gesagt: „Wir wollen keine globalen Verhandlungen, wir wollen nur Regeln für amerikanische Firmen und wir wollen, das wird eh alles viel zu langsam und zäh auf der globalen Ebene, das funktioniert nicht, und wir wollen auch nicht Ethikquatsch, wir wollen Fairness und Transparenz für amerikanische Firmen und der ganze Rest, den Firlefanz, den könnt ihr da in Europa machen.“ [00:36:07] Es ist natürlich dann auch ein Unterschied, man sieht aktuell, wohin das führt. [00:36:13] Ganz aktuell kann man sich das anschauen in Amerika, aber muss natürlich auch sehen, dass wir hier in Europa sehr alleine sind zwischen den beiden großen Playern China und Amerika und dass die unsere Werte nicht teilen. [00:36:31] Und wenn denen irgendjemand sagt, um mal bei unserem Thema zu bleiben, künstliches Bewusstsein ist militärisch relevant oder man kann damit viel Geld verdienen, dann wird das natürlich gemacht, das ist überhaupt keine Frage. [00:36:47] Aber ich plädiere in der Ethik immer auch für ein Prinzip Selbstachtung. [00:36:55] Was meine ich damit – dass man das Richtige auch dann tut und ausspricht, wenn die Hoffnungen auf einen Erfolg und eine Einigung gering sind. [00:37:06] Ja, dass wir, auch wenn es im globalen Kontext oder sogar in Europa selbst nicht gelingt, ernsthafte ethische Regeln für die KI zu

entwickeln, trotzdem einfach immer weiter über das nachdenken, was richtig wäre und das auch aussprechen um unsere Forschungsergebnisse, einfach aus Respekt vor uns selbst, ja, sozusagen um keinen Schaden an unserer eigenen Seele zu nehmen, auch wenn wir damit nicht erfolgreich sind. [00:37:36] Und da sind wir jetzt, das fällt mir gerade ein, eigentlich bei Ihrer Frage von eben: Könnte eine Maschine auch so was wie diese Selbstachtung entwickeln? [00:37:47]

Karsten Wendland: So schließt sich der Kreis. [00:37:48] Könnte sie eine Selbstachtung dieser Art entwickeln oder wäre das dann nicht möglicherweise nur eine Simulation? [00:37:56] Man könnte das der Maschine ja auch beibringen, dass sie auf ihre eigene Simulation reagiert und dann so tut als ob und wir als Beobachter von außen wären damit vielleicht zufrieden und sagen okay, das hat jetzt funktioniert. [00:38:09] An welcher Stelle könnten wir denn erkennen, dass in einer Maschine plötzlich mehr los ist als zuvor, also wenn dieser Punkt des Bewusstseins erreicht wird? [00:38:21] Wie können wir das denn merken? [00:38:23]

Thomas Metzinger: Die Frage gefällt mir sehr gut, die sie eben gestellt haben, weil das ist natürlich die eigentliche philosophische Frage. [00:38:31] Die kann man übrigens noch einen Schritt weiter radikalisieren, also man könnte auch fragen, woran merken wir selbst eigentlich, dass wir bewusst sind und dass wir nicht nur Bioautomaten sind, die auf ihre eigene innere Simulation reagieren. [00:38:49] Zu der eigenen inneren Simulation könnte ja gehören, dass wir Bewusstsein haben, also wir könnten Bio-Roboter sein, die darauf reingefallen sind auf ihre eigene innere Simulation. [00:39:05]

Karsten Wendland: Das sagte der Kollege Dennett beispielsweise, „I am a Bio Robot“ und ist damit auch sehr erfolgreich (**Thomas Metzinger**: Ja, natürlich). [00:39:11] Also man kommt mit diesen Aussagen definitiv in die Medien und bekommt Aufmerksamkeit, aber es trägt nicht zur Beantwortung der Frage bei. [00:39:19] Meine Frage ist: Kann man denn diese Frage überhaupt beantworten? [00:39:23]

Thomas Metzinger: Also ich glaube schon, also man kann natürlich ein großes philosophisches Problem aufmachen. [00:39:28] Alles, was wir wissenschaftlich erforschen können, sind Funktionen, ja, so Input-Output-Relationen, auch feinkörnige Funktionsmuster im Gehirn. [00:39:39] Und es gibt so

eine standardphilosophische Diskussion, die sagt, dass phänomenale Qualitäten wie Röte oder vielleicht auch die Bewusstheit selbst nicht funktional analysierbar sind und dass man deswegen Simulationen eines bewussten Wesens niemals unterscheiden kann von einem bewussten Wesen. [00:39:57] Das halte ich aber für völlig falsch und eigentlich auch widerlegt, es gibt vielleicht logisch mögliche Welten, in denen das so ist, aber in dieser Welt, in der wir leben, ist es nicht so. [00:40:08] Und zwar wissen wir ja auch, dass mit dem Bewusstsein auch bei Tieren ganz bestimmte messbare Fähigkeiten einhergehen, zum Beispiel erhöht Bewusstsein die Selektivität des Verhaltensprofils. [00:40:27] Das heißt, wenn Sie schon mal einen Schlafwandler beobachtet haben, der ohne Bewusstsein nachts den Gang lang tappt-, das sieht man, dass da was fehlt. [00:40:37] Da fehlt in der motorischen Verarbeitung etwas, was Forscher auch Kontextsensitivität nennen. [00:40:44] Der ist sehr starr in seinen Reaktionen, wie Insekten, der ist nicht gerade flexibel, der Schlafwandler. [00:40:53] Auch die Adaptivität des Verhaltensprofils, das heißt, sich anpassen an Umgebung und so ist ziemlich schlecht. [00:41:03] Im Volksmund denkt man immer so, einem Schlafwandler passiert schon nix, besonders wenn der auf dem Dach ist, dann soll man ihn nicht wecken, aber das stimmt gar nicht. [00:41:11] Schlafwandler tun sich weh, Schlafwandler stürzen, Schlafwandler stoßen auch wo an. [00:41:18] Das heißt, dass bewusste Informationsverarbeitung ist eine bestimmte Form von Intelligenz, die erzeugt Flexibilität, Adaptivität, Kontextsensitivität, das kann man messen. [00:41:32] Das hängt wohl damit zusammen, dass das System ein integriertes Realitätsmodell hat, das auch zum Beispiel in der Zeit ein bisschen ausgedehnt ist, dass es so was wie einen gelebten Moment gibt. [00:41:45] Da gibt es schon sehr viele Arten, auch scheint es so zu sein, dass man Bewusstsein braucht für bestimmte, sehr schnelle Lernvorgänge, also wenn man, dass man eine Sache, die einem nur einmal wehgetan hat, dann wirklich für immer gelernt hat. [00:42:03] Das kann man alles sehr genau wissenschaftlich untersuchen und wir wissen ziemlich viel darüber, für welche der Fähigkeiten, die wir Menschen haben, Bewusstsein notwendig ist und die ohne Bewusstsein einfach nicht funktionieren. [00:42:19]

Karsten Wendland: Und könnten wir so weit gehen, wenn wir diese Fähigkeiten bei den Maschinen beobachten, dass wir ihnen dann das Bewusstsein nicht nur zuschreiben, sondern uns auch darauf verlassen können sollten, dass dieses Bewusstsein in den Maschinen jetzt da ist? [00:42:35]

Thomas Metzinger: Ja, da steckt natürlich eine zweite Frage drin, die auch ganz wichtig ist, nämlich die soziale Konstitution von Bewusstsein. [00:42:44] Wir fangen ja als sehr kleine Kinder wahrscheinlich an, indem wir erst mal zu verstehen versuchen, was die Mama will und dann auch merken, dass wir die Mama manipulieren können, indem wir schreien. [00:42:57] Das heißt, wahrscheinlich ist es so, dass wir erst anderen bewegten Objekten in unserer Umwelt innere Zustände zuschreiben, und dann entdecken wir auf einmal, dass wir selber so was haben, wenn die Mama sagt: „Du bist aber böse“, oder „Dir geht es doch-, es tut doch gar nicht weh“ oder so was. [00:43:20] Das heißt, in der frühkindlichen Entwicklung spiegeln auch Erwachsene das Bewusstsein in Kinder rein in einem bestimmten Ausmaß, das ist sehr interessant. [00:43:32] Und die Frage ist, wie viel unserer Bewusstheit auch von einer sozialen Zuschreibungspraxis abhängt, dass wir bestimmte Begriffe gebrauchen, unsere Kinder auf eine bestimmte Weise erziehen. [00:43:44] Und jetzt könnte es natürlich sein, dass wir bei der Maschine immer wieder sagen, sie sei robust oder sie sei wach, oder wenn Maschinen sich das gegenseitig sagen, dass da auch eine ganz interessante neue Ebene von Komplexität entsteht, die auch wichtig ist. [00:44:04] Also ein Ich-Gefühl ist nicht nur was, was im Gehirn entsteht, Ich-Selbstmodelle entstehen auch in Gesellschaften und aus interagierenden Wesen, die sich selbst auch anerkennen als bewusst. [00:44:21] Ich glaube aber, dass das für diese Grundwachheit der soziale Kontext nicht notwendig ist, also zum Aufwachen morgens und zum überhaupt sich orientieren und zu sich kommen, braucht man keine anderen Menschen, braucht man keinen intersubjektiven sozialen Kontext. [00:44:40] Und das kann man ja alles in Maschinen nachbauen, wir können natürlich Maschinen bauen, die zum Beispiel das haben, was wir beim Menschen den Orientierungsreflex nennen, also die sich aufsetzen und dann die Frage beantworten: Wo bin ich? Welche Zeit ist gerade? Und: Wer bin ich? [00:45:04] Das kann ich mir sehr gut vorstellen, dass die Systeme eine innere Zeitrepräsentation haben und dann eine Darstellung des Zeitpunkts: Ah, es ist wieder morgen, ich bin wieder hochgefahren worden und ich bin die KI 417D und mich gibt es seit sechs Jahren und ich habe diese Geschichte, ich habe Erinnerungen, ich habe eine Interaktionsgeschichte mit der Welt, an die ich mich erinnern kann. #00:454:30# All das könnte es in Maschinen geben und wir wissen es ja bei uns selbst gar nicht, wann der kritische Übergang stattfindet, weil wir ja diese frühkindliche Amnesie haben. [00:45:44] Also wir alle können

uns ja sehr schlecht an irgendwas erinnern, was vor dem Alter von zwei Jahren war und die meisten von uns gehen davon aus, wenn ein kleines Kind von anderthalb Jahren weint oder schreit, dass es was erlebt. [00:45:58] Das heißt, wir haben schon Sachen erlebt, bevor wir sprechen konnten, Sachen, die noch nicht zu unserer inneren Lebensgeschichte gehörten. [00:46:09] Und ich glaube, so eine einfache Art von Bewusstsein könnte natürlich auch in Maschinen entstehen, ja, Wissen, wo man ist, was man selbst im Unterschied zur Umgebung ist, welche Zeit jetzt gerade ist. [00:46:26] Das könnte passieren. [00:46:28]

Karsten Wendland: Und um das aufzugreifen, wäre dann vielleicht auch eine neue Fachdisziplin gefragt, eine Art Pädagogik für künstliche Systeme, die sich mit Bildungs- und Erziehungsfrage für synthetisches Bewusstsein beschäftigt. [00:46:40] Würden Sie soweit gehen? [00:46:42]

Thomas Metzinger: Na ja, vieles deutet ja darauf hin. [00:46:45] Also das Erste, ich bin ja schon einige Jahrzehnte dabei, das Erste, was wir sehr bald gelernt haben, ist, die Systeme brauchen einen Körper. [00:46:54] Na, diese ganze Robotik, Embodiment, bis man gemerkt hat, bestimmte Formen von Intelligenz in der wirklichen Welt können nur erlernt werden, wenn die Dinger einen Körper haben, mit der realen Welt interagieren. [00:47:07] Dann hat man sehr bald gemerkt, eigentlich brauchen die auch so was wie eine Kindheit, die müssen wie bestimmte Roboter in Japan, die ich gesehen habe, lernen auf dem Bauch zu krabbeln und sich selber aufzusetzen, also sozusagen aus dem Nichts heraus solche komplexen Motormuster zu evolvieren. [00:47:34] Dann gibt es natürlich spannende Experimente zum Spracherwerb, war ich selber mal am Wissenschaftskolleg in Berlin dran beteiligt, wie Systeme ihre eigene Sprache erzeugen können, all das. [00:47:49] Also, wir werden verkörperte Systeme haben, wir werden Systeme haben, die eine Kindheit haben und dann werden wir natürlich auch in einen Prozess reinkommen, wo wir nicht mehr kontrollieren können, in welche Richtung das geht oder selber nicht mehr verstehen. [00:48:05] Ist ja bei uns auch so, unsere Kinder überraschen uns, die werden auf einmal frech, die hauen auf einmal ab, die rauchen heimlich. [00:48:14] Die machen irgendwas, womit die Eltern nicht gerechnet haben. [00:48:19] Und je mehr Autonomie wir diesen Systemen geben, desto häufiger werden die uns auch überraschen oder vor Rätsel

stellen. [00:48:29] Und ein so ein Rätsel könnte sein, ja, tut das denn jetzt wirklich schon weh? [00:48:37]

Karsten Wendland: Oder tut er nicht nur so, als ob. [00:48:38] Ja. Okay. Jetzt beschäftigen Sie sich ja auch damit, wie Theorien entstehen und wie sie wieder vergehen und angenommen, man ist jetzt wissenschaftlich viele Jahre lang begeistert auf einem Holzweg unterwegs und bekommt dann irgendwann die Erkenntnis, dass man eigentlich falschlag. [00:48:57] Wie schwierig ist das in so einem aktiven Forscherleben plötzlich umzusteuern? [00:49:04]

Thomas Metzinger: Also, wer hat denn das gesagt, das war Max Planck, das ist eine Beerdigung, die nur vorwärts geht eigentlich. [00:49:10] Menschen sind so, dass sie ab einem gewissen Alter festhalten an – ist ja auch verständlich – an Theorien in die sie ihr ganzes Leben Tausende von Stunden von Arbeit investiert haben. [00:49:25] Das sieht man auch an der deutschen Philosophie, das sieht man an den wissenschaftlichen Disziplinen, es ist schwer, ein Kant-Forscher zu sein und alles zu wissen, über Kant und sein ganzes Leben investiert zu haben. [00:49:40] Und dann kommen junge Leute und behaupten frech, das sei nicht wirklich relevant mehr, obwohl sie es gar nicht kennen. [00:49:47] Und so geht das natürlich, Generation über Generation von Forschern definiert was Neues für sich, was sie wirklich relevant finden. [00:49:56] Und das Problem bei dem Prozess ist, dass manche nicht erkennen, Philosophen wissen das besonders gut, dass das Rad schon mal erfunden worden ist vor 2000 Jahren, oder dass es auch schon mal andere Leute gab, die da sehr tief darüber nachgedacht haben, obwohl die keine Computer und keine Naturwissenschaft im eigentlichen Sinne hatten. [00:50:22] Trotzdem sehe ich ja zum Beispiel jetzt im Moment so ein Hauptproblem. Ganz einfaches Problem ist, die Bundesregierung bräuchte Hunderte von erstklassig ausgebildeten Experten in Ethik und Künstlicher Intelligenz, die Firmen brauchen das. [00:50:38] Wir haben das nicht, weil eine verschwindend geringe Zahl von deutschen Philosophen sich für Künstliche Intelligenz interessiert hat und da kompetent geblieben ist in dem Bereich. [00:50:52] Deswegen haben wir jetzt eine riesige Lücke, die Industrie füllt die. Facebook macht an der TU München eine eigene Ethikausbildung. [00:51:00] Und warum? Das kann sich jeder denken. [00:51:04] Das heißt, uns fehlen jetzt, und das ist auch so ein bisschen die Verdrängung der älteren Generation der Geisteswissenschaftler, die immer gedacht haben, na, das ist was Böses. [00:51:18]

In meiner Studentenzeit, ich erzähle Ihnen jetzt noch mal eine Geschichte aus meinem Leben, als ich nach meiner Doktorarbeit 1987 mein erstes Seminar an der Universität Frankfurt halten durfte, da habe ich dem den Titel Künstliche Intelligenz und Philosophie gegeben. [00:51:35] Und dann, als ich mit klopfendem Herzen in diesen Raum, 104 war das, glaube ich, rein wollte, da konnte ich da nicht rein, weil da eine Traube von Menschen vor der Tür standen und alles war pechschwarz mit Leuten und ich konnte noch nicht mal zur Tafel vor mit meinem klopfenden Herzen, die wollten mich auch nicht reinlassen. [00:51:56] Was muss der jetzt noch mit seinen Turnschuhen noch ganz nach vorne. [00:52:00] Und im Grunde waren die alle gekommen, um das zu verhindern. [00:52:04]

Karsten Wendland: Um es zu verhindern. [00:52:05] Ich dachte, das wäre jetzt Ihr Publikum, die Geschichte hat so schön angefangen, ich dachte, das wäre das Publikum gewesen, (Thomas Metzinger: Nein, nein), das so an diesem Thema interessiert ist. [00:52:15]

Thomas Metzinger: Nein, Künstliche Intelligenz, das hatte man in der Frankfurter Schule noch nie gehört 1987 da, und Hilary Putnam, die amerikanischen Debatten waren da natürlich völlig unbekannt im angelsächsischen Bereich. [00:52:27] Und wir hatten als Studenten ein Wort, das hieß: Protofaschistisch, und ich glaube, viele hatten das Gefühl, das ist was, das ist: protofaschistisch. [00:52:39] Künstliche Intelligenz, im Ernst darüber nachdenken, ob geistige Funktionen von Menschen, von rationalen, freien Subjekten auf Maschinen dupliziert werden können, das ist politisch gefährlich. [00:52:51] Also da waren ganz viele Leute, die Gefühle hatten, glaube ich, hier soll was politisch Gefährliches lanciert werden. [00:52:57] Wir wissen zwar nicht, was es ist, aber sind schon mal dagegen. [00:53:02] Und das war so die Stimmung und das zeigt eigentlich was. [00:53:09] Es zeigt eine Abwehrhaltung in den Geisteswissenschaften gegen wissenschaftliche Trends, die sie verpasst haben aus Selbstgefälligkeit. [00:53:19] Das konnte man damals schon sehen und man sieht heute, dass es einfach sehr wenige Philosophen in Deutschland gibt, die überhaupt halbwegs was von diesem Bereich verstehen. [00:53:30] Und das ist jetzt auf einmal politisch auch ein Problem, weil die Bundesregierung natürlich gute Beratung bräuchte und hier viele solcher Leute ausbilden müsste. [00:53:40] Es hängt mit den Belohnungsstrukturen in akademischen Werdegängen in Institutionen zusammen. [00:53:48]

Wenn es eine Belohnung dafür gibt, einfach Kant, Fichte, Schelling, Hegel auswendig zu lernen, damit kann man bis 35 fertig sein und man dafür Professor werden kann, dann machen das viele Leute. [00:54:00] Wenn es Belohnungen dafür gibt, sich mit brandaktuellen wichtigen Themen wie Künstliche Intelligenz zu beschäftigen, tun das auch viele Leute, aber die Anreizstruktur und Belohnung waren bei uns so, dass wir jetzt ein bisschen Probleme haben. [00:54:16]

Karsten Wendland: Das ist der Schweinezyklus in der Philosophie (**Thomas Metzinger**: Ja, das haben Sie jetzt gesagt). [00:54:24] Ja, ich würde gern noch mal den Holzweg aufgreifen. [00:54:27] Möglicherweise sind wir ja zurzeit auf einem Holzweg, was die KI anbelangt. [00:54:33] Ich möchte einmal in die 80er-Jahre zurückgehen, da gab es auch solche Kampagnen. [00:54:37] Eine Kampagne bezog sich darauf, dass man mehr Milch trinken soll, da wurden T-Shirts verteilt auch an den Schulen, darauf stand: Die Milch macht‘s. [00:54:47] Und es wurde viel Milch produziert und irgendwann einige Jahre später hat man gemerkt, so viel Milch braucht man gar nicht. [00:54:54] Und den Viehwirten wurde dann Verschüttungsprämien dafür gezahlt, um die ganze produzierte Milch wieder irgendwie unauffällig los zu werden. [00:55:03]

Thomas Metzinger: Ich erinnere mich auch noch daran. Ja. Ja. [00:55:06]

Karsten Wendland: Und sagen wir heute vielleicht ganz unbedacht, die KI macht‘s, und sind auf einem ähnlichen Holzweg? [00:55:12]

Thomas Metzinger: Das ist auch ein ganz wichtiger Aspekt. [00:55:16] Man muss jetzt erst mal sehen, dass das, was im Moment den Boom nach mehreren KI-Wintern erzeugt hat, eigentlich nicht genuin theoretische Durchbrüche sind. [00:55:25] Also nicht das, was Philosophen interessieren würde, ein echt tieferes Verständnis, was Intelligenz ist, sondern es sind schon bekannte Verarbeitungsprinzipien, bloß zwei Sachen sind besser geworden. [00:55:42] Die Hardware ist viel schneller geworden und wir haben wesentlich größere Datenmengen zur Verfügung, mit denen wir die Systeme trainieren können, das heißt, die Dinger haben im Moment auf dem, was man die Performanzebene nennt, große Erfolge, die auch Fachleute überraschen. [00:56:00] Das heißt, es gibt bestimmte Dinge in der natürlichen Sprachverarbeitung, Gesichtserkennung und so weiter, Bildverarbeitung, die funktionieren jetzt auf einmal

praktisch gut. [00:56:10] Man darf das nicht verwechseln mit einer tiefen theoretischen Einsicht, sagen wir mal in das Wesen des menschlichen Geistes oder sogar in das, was Bewusstsein ist. [00:56:25] Es ist nur einfach so, dass wesentliche Dinge jetzt viel besser funktionieren, und allein das reicht ja auch schon aus, um diese ganzen ethisch relevanten Risiken zu erzeugen. [00:56:37] Also im militärischen Bereich, Massenarbeitslosigkeit, Automatisierung, das sind nicht immer solche Science-Fiction Szenarien, also das nächste Szenario ist große Arbeitslosigkeit, das könnte schon 2030 auf uns zukommen, und zwar nicht durch Cutting Edge KI, sondern durch ganz altmodische Automatisierung in der Produktion. [00:57:05] Und allein das kann schon dramatische gesellschaftliche Folgen haben, sollte zum Beispiel vielleicht ein Grundeinkommen notwendig machen, führt zu neuen Verteilungs- und Gerechtigkeitsproblemen in Gesellschaften. [00:57:20] Das heißt, es sind nicht immer diese Science-Fiction Themen, die für die Ethik wichtig sind, sondern ganz andere. [00:57:28] Zum Beispiel wird ja sehr viel Gutes auch auf uns zukommen im Bereich der Medizin und die Frage ist, wie verteilt man den medizinischen Fortschritt dann so, dass wirklich alle davon profitieren und nicht nur reiche Leute zum Beispiel oder Privatpatienten und so weiter und so fort. [00:57:47]

Karsten Wendland: Sie sagen, die Systeme sind nicht zwingend qualitativ besser geworden, sondern sie sind schneller geworden, sie sind stärker vernetzt, der Speicher ist günstiger geworden. [00:57:57] Und dementsprechend haben wir heute Systeme, die es vom Grundansatz auch schon vor zehn Jahren hätte geben können, heute eben in lauffähiger Form. [00:58:05] Der qualitative Durchbruch ist das nicht, die Wirkungen sind aber schon gegeben. [00:58:11]

Thomas Metzinger: Na ja, das ist halt eine Dialektik zwischen Quantität und Qualität, bestimmte Geschwindigkeiten und bestimmte Datenmengen erzeugen dann überraschend neue Qualitäten. [00:58:21] Ich nenne mal ein bisschen anderes Beispiel, das nicht direkt mit Bewusstsein zu tun hat. [00:58:26] Langsam beginnen auch hier in Europa Leute zu erkennen, dass die sozialen Medien die Grundlage der Demokratie untergraben und dass wir ein, zwei Generationen von Kindern und Jugendlichen verheizt haben. [00:58:39] Das erkennen zum Beispiel auch Hochschullehrer und das liegt natürlich daran, dass zum Beispiel nehmen wir mal an Facebook das Verhalten seiner drei

Milliarden Nutzer mit KI optimiert. [00:58:53] Das heißt, KI ist zum Beispiel auch immer an ein Geschäftsmodell gekoppelt, ja also an die Interessen der Werbekunden von Facebook oder Google, nicht an das Gemeinwohl, nicht an das Ideal der geistigen Gesundheit oder nicht an das Ideal einer Technologie, die den Menschen dazu hilft, ein besseres Leben zu leben, sondern es geht um die Profite der Werbekunden von Facebook. [00:59:18] Und da lernen Algorithmen jetzt einfach sehr effektiv und sehr schnell, und zwar 24 Stunden am Tag auf der ganzen Welt, was die Leute dazu bringt, länger am Bildschirm zu kleben als sie eigentlich vorhaben, wie man Leute dazu bringt, sich zu verlaufen im Internet, aus Versehen doch Werbung anzuschauen, die sie nicht anschauen wollten. [00:59:41] Und die KI lernt etwas, die lernt was über uns Menschen. [00:59:45] Die lernt zum Beispiel Empörung, Hass und Wut funktionieren sehr gut im Interesse unserer Werbekunden. [00:59:55] Wir lernen auf einmal, man kann das Benutzerverhalten steuern, indem man Zwietracht sät und den sozialen Neid stimuliert und so weiter und so fort. [01:00:08] Das hat die EU eigentlich noch gar nicht so richtig verstanden, aber wir sehen jetzt überall das Entstehen von Populismus, rechten Bewegungen, den Zerfall des sozialen Zusammenhalts. [01:00:19] Können Sie ja sehr deutlich zum Beispiel in USA beobachten, wo das mit den KI-gestützten sozialen Netzwerken schon tiefer eingedrungen war in die Gesellschaft als heute. [01:00:30] Das heißt, da mag kein wirklicher theoretischer Durchbruch dahinterstehen hinter dieser Anwendung von Künstlicher Intelligenz zur Optimierung und Steuerung und Manipulation von Benutzern sozialer Netze. [01:00:44] Aber der Effekt ist da und alle wundern sich: Woher kommen die Gewaltausbrüche denn? Woher kommen Studenten, die nichts mehr lesen wollen? Woher kommt all das, plötzliches Entstehen von immer mehr populistischen Machthabern oder Parteien? [01:01:03] Das sind KI-Auswirkungen und dazu braucht es nicht Science-Fiction KI. [01:01:09] Da braucht es nur, dass etwas sehr gut funktioniert, und zwar so gut, dass es alle überrascht. [01:01:14] Und das haben wir jetzt schon, wir sind selbst überrascht davon, dass wir alle ein kleines bisschen internetsüchtig sind. [01:01:23] Wir versuchen, dass alle darüber Witze zu machen und so ein bisschen wegzudrücken, aber jeder von uns merkt ja, wir gehen spazieren, ja, wir gehen in den Wald., aber wir merken genau wie der Raucher so eine kleine Sucht nach der nächsten Zigarette kriegt, dass irgendwas in uns diese Neuigkeitsreize wieder haben will, diese kleinen Dopamin-Ausschüttungen in unserem Gehirn.

[01.01:40] Und da gehen wir an diesen Spielautomaten, den wir unser Handy nennen und versuchen unser Glück und das hat alles mit KI zu tun. [01:01:58]

Karsten Wendland: Herr Metzinger, was wären denn Ihre Empfehlungen für die Hersteller von Systemen, die künstliche Intelligenz integriert haben und was wären Ihre Empfehlungen für die politischen Akteure? [01:02:11]

Thomas Metzinger: Für die Hersteller sowie für die Policy-Maker in Deutschland und Europa ist eine Sache wichtig: Unser Alleinstellungsmerkmal ist vertrauenswürdige KI in Europa und der strategische Vorsprung, den wir uns gegenüber diesen übermächtigen Akteuren China und USA erarbeiten können, ist, wenn wir das mit dem Green Deal von Frau von der Leyen verknüpfen [Quellenverweis 6]. [01:02:38] Das heißt, wenn wir hier vertrauenswürdige KI herstellen, die Umwelttechnologie insbesondere im Blick auf den Klimawandel und alles, was auf uns jetzt sehr stark zukommen wird, attraktiv ist, dann wird die ganze Welt nach Europa schauen und dann wird die ganze Welt europäische KI kaufen wollen, wenn uns das gelingt. [01:03:02] Erstens, weil wir dann die smarte Umwelttechnologie haben, die niemand hat und zweitens, weil man dann so wie es früher "Made in Germany" gab, "Trustworthy AI made in Europe" im Prinzip haben können. [01:03:17] Wo man weiß, da steckt nicht der chinesische Staat indirekt drin über Updates, da wird man nicht ausgehorcht, da sind wenigstens grundlegende demokratische Standards, Menschenrechtsstandards in der Technologie eingebaut, zum Beispiel über Industrienormen, über "Ethics by Design". [01:03:37] Das könnte auf dem Weltmarkt noch unsere Chance sein in dieser schwierigen Situation, weil KI aus Amerika und KI aus China wahrscheinlich genau diese Qualität nicht haben wird, der Vertrauenswürdigkeit. [01:03:57] Und das sollten auch deutsche und europäische Produzenten bedenken mit solchen Systemen, dass sie nicht ihren Ruf aufs Spiel setzen, weil ein Skandal, wenn sich das einmal wie bei der Autoindustrie rumgesprochen hat, die Technologie ist nicht sauber, die betrügen, dann ist der Rufschaden da und das zu vermeiden, das könnte strategisch sehr klug sein. [01:04:22]

Karsten Wendland: Sie würden also auf vertrauenswürdige KI setzen, hier in Europa, eine Künstliche Intelligenz, der wir den weißen Hut aufsetzen und nicht den schwarzen in dem Vertrauen, dass das Gute letztlich gewinnt. [01:04:37]

Thomas Metzinger: Nein, das ist nicht so, ich habe nicht das Vertrauen, dass das Gute letztlich gewinnt, das muss ich ganz deutlich sagen. [01:04:42] Also meine persönliche Lebenserfahrung deutet einfach nicht in diese Richtung, aber man kann ja versuchen, das Gute trotzdem anzustreben und zu tun. [01:04:53] Sehr gut, besser als das, was wir in Brüssel gemacht haben, war zum Beispiel der Bericht der Deutschen Datenethikkommission. [01:05:03] Frau Woopen hat da wirklich einen großen Erfolg gehabt. [01:05:07] Das ist differenzierter als unsere Ethikrichtlinien, die wir in Brüssel erarbeitet haben, das heißt, die Möglichkeit besteht, wenn der politische Wille dazu da ist, das umzusetzen. [01:05:18] Ich selbst bin skeptisch, aber wir sollten es zumindest probieren und wir sollten auch an ein anderes Stichwort wirklich noch mal denken, nämlich an die digitale Souveränität. [01:05:32] Wir haben die im Moment nicht, wir werden von amerikanischen Geheimdiensten bis auf die Knochen ausgehorcht. [01:05:41] Wir haben uns völlig unterworfen amerikanischen Software-Konzernen, Microsoft und Google, obwohl wir die Möglichkeit mit Open Source Software hatten, überall. [01:05:52] Wir müssen einfach auch unsere politischen Institutionen davor schützen, auf dem Weg über Technologien, über Software, über KI-Dienste übernommen zu werden aus Amerika oder aus China. [01:06:09] Das würde sich wirklich lohnen, da ein besonderes Augenmerk drauf zu richten. [01:06:13]

Karsten Wendland: Und da sind wir in unseren eigenen Wertebildern und unserem eigenen Bewusstsein gefragt. [01:06:17]

Thomas Metzinger: Ja, natürlich, es gehört aber auch dazu, dass wir zum Beispiel mit China einen ehrlichen und offenen Dialog über deren Ethik führen. [01:06:29] Die haben nämlich auch eine philosophische Tradition, die haben andere Werte als wir. [01:06:35] Unsere Werte sind zum Beispiel die individuelle Freiheit, Vernunft, die haben aber Werte wie Achtsamkeit und Mitgefühl zum Beispiel und sozialen Zusammenhalt, Familienwerte und die sind ja an sich nicht schlecht. [01:06:50] Das heißt, man muss auch immer einen Unterschied machen, sagen wir mal zwischen der chinesischen Regierung und ihrer Strategie und den tiefen geistigen Traditionen, die in anderen Kulturkreisen sind mit eigenen ethischen Grundwerten, und da mal ganz offen schauen, ob es da nicht vielleicht Dinge gibt, die auch gut für uns wären. [01:07:10]

Karsten Wendland: Herr Metzinger, wir sprachen jetzt in großem Bogen über Künstliche Intelligenz, Ethik, Verantwortung und auch über selbstbewusste Künstliche Intelligenz, die, wenn es sie denn gäbe, mehr wäre als eine bloße Maschine. [01:07:24] Frage zum Abschluss an Sie: Wie lange wird es noch dauern? [01:07:27]

Thomas Metzinger: Bis wir künstliches Bewusstsein haben? Oh, 2065. [01:07:36]

Karsten Wendland: Das war Thomas Metzinger, Philosoph in Mainz und seit vielen Jahren intensiv forschend zu Möglichkeiten von Bewusstsein bei Künstlichen Intelligenzen in unserer Podcast-Serie zu Selbstbewusster KI, Ihrem Forschungspodcast an der Grenze zwischen Mensch und Maschine. [01:07:53] Sind Ihnen beim Zuhören weitere Fragen eingefallen oder geniale Ideen gekommen? [01:07:58] Wir freuen uns über Ihre Gedanken. [01:08:00] Lassen Sie uns daran teilhaben und eine Nachricht über unsere Projekt-Website zukommen, die Sie im Internet unter www.ki-bewusstsein.de finden. [01:08:09] Oder schreiben und folgen Sie uns auf Twitter, dort finden Sie unser Projekt unter dem gleichen Namen @KIBewusstsein. [01:08:16] In der nächsten Folge sprechen wir mit Frauke Rostalski. [01:08:21] Sie ist Rechtswissenschaftlerin in Köln, Expertin für Biotechnologie und Künstliche Intelligenz und außerdem Mitglied des Deutschen Ethikrats. [01:08:30] Redaktion, Aufnahmeleitung und Produktion dieser Folge lagen in den guten Händen von Robert Sinitsyn. [01:08:36] Ich freue mich, wenn es Ihnen gefallen hat und auch diese Folge für Sie ein Betrag dazu war, KI-Bewusstsein etwas mehr zu entmystifizieren. [01:08:45] Bleiben Sie gesund, hoffnungsvoll und gestaltungsstark. [01:08:48] Das war Ihr und euer Karsten Wendland, bis bald!

Ende [01:08:52]

5 Erwähnte Quellen

Folgende weiterführende Quellen wurden in der Podcast-Folge genannt:

[1] René Descartes: Die Leidenschaft der Seele (Les passions de l'âme), L. Heimann 1870.

[2] The Minimal Phenomenal Experience (MPE) Project unter Leitung von Thomas Metzinger. https://www.philosophie.fb05.uni-mainz.de/arbeitsbereiche/theoretische/mpe/

[3] Artikel von Thomas Metzinger bei Spektrum.de zu maschinellem Bewusstsein mit Bezug auf den amerikanischen Philosophen Hilary Putnam und seine Einstellungen zu Rechten von Robotern. https://www.spektrum.de/pdf/gug-06-04-s068-pdf/832965

[4] Europäische High-Level Expert Group on Artificial Intelligence. https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

[5] Thomas Metzinger: Der Ego-Tunnel: eine neue Philosophie des Selbst. Von der Hirnforschung zur Bewusstseinsethik, Piper Verlag 2014. https://www.piper.de/buecher/der-ego-tunnel-isbn-978-3-492-30533-4

[6] Digitalisierung und der Green Deal der EU-Kommission. https://ec.europa.eu/germany/news/20200219digitale-zukunft-europas-eu-kommission-stellt-strategien-fuer-daten-und-kuenstliche-intelligenz_de

6 Kontakt

Zur Website des ITAS

Prof. Dr. Karsten Wendland
karsten.wendland@kit.edu

Karlsruher Institut für Technologie (KIT)
Institut für Technikfolgenabschätzung und Systemanalyse (ITAS)
Karlstraße 11
76133 Karlsruhe
GERMANY