Öffentliche Anhörung am 25. Februar 2021

Künstliche Intelligenz und Mensch-Maschine-Schnittstellen

Online-Veranstaltung

25. Februar 2021, 09:30 Uhr

Programm

Moderatorin der Veranstaltung:
Judith Simon · Deutscher Ethikrat

Hinweis: Es handelt sich bei dem folgenden Text nicht um eine wörtliche Transkription. Der Text wurde lektoriert, um eine gute Lesbarkeit herzustellen. Da eine Simultanübersetzung mitunter nicht alle Informationen aus dem gesprochenen Original wiedergibt, wurden im folgenden Transkript darüber hinaus an mehreren Stellen Ergänzungen oder Korrekturen vorgenommen. Der Videomitschnitt der Veranstaltung kann jedoch in voller Länger und im originalen Wortlaut unter folgendem Link auf unserer Website abgerufen werden: https://www.ethikrat.org/anhoerungen/kuenstliche-intelligenz-und-mensch-maschine-schnittstellen/.

Begrüßung

Alena Buyx · Vorsitzende des Deutschen Ethikrates

Guten Morgen, meine Damen und Herren, liebe Kolleginnen und Kollegen, ich begrüße Sie herzlich zum öffentlichen Teil der heutigen Plenarsitzung des Deutschen Ethikrates. Wir haben heute eine spannende Anhörung vor uns.

Bevor wir in die Tagesordnung einsteigen, habe ich eine freudige Pflicht: Ich darf unser neues Mitglied, Professor Armin Grunwald, sehr herzlich begrüßen, der heute seine erste Sitzung mit uns verbringt. Ich verzichte auf eine umfangreiche Vorstellung; alle Materialien finden Sie auf unserer Homepage. Herr Grunwald ist Professor für Philosophie und Ethik der Technik am Karlsruher Institut für Technologie. Viele werden ihn kennen in seiner Rolle als Leiter des Büros für Technikfolgen-Abschätzung beim Deutschen Bundestag, dem TAB. Seine vielen verschiedenen Funktionen, Ehrungen und Publikationen können Sie auf unserer Webseite nachlesen.

Wir freuen uns sehr, dass es Herrn Grunwald möglich war, jetzt so schnell in die heutige Sitzung zu kommen. Herr Grunwald, wir wissen das besonders zu schätzen, weil Sie tatsächlich im Zug sitzen. Wir freuen uns, dass Sie da sind, herzlich willkommen.

Armin Grunwald

Herzlichen Dank, liebe Frau Buyx, liebe Kolleginnen und Kollegen, ich freue mich sehr darüber, jetzt in dieser verantwortungsvollen Tätigkeit bei Ihnen mit an Bord sein zu dürfen.

Ich habe letzte Woche erst die Berufung erhalten und bitte deswegen um Verständnis, dass ich heute unter etwas ungewöhnlichen Umständen teilnehme und dass vielleicht auch in den nächsten Monaten noch keine volle Präsenz da sein wird. Es gibt meinerseits schon einen Kalender, in den ich die Termine Ihrerseits integrieren muss. I'll do my very best und ich bin optimistisch, dass das schnell gelingen wird.

Auf jeden Fall freue ich mich sehr auf die Arbeit, auf die Kooperation mit Ihnen und auf die spannenden Diskussionen – vor allen Dingen auf die nicht leichten Diskussionen. Das sind ja die spannendsten, weil man dort am meisten lernen kann. Ich freue mich darauf.

Alena Buyx

Vielen Dank, lieber Herr Grunwald. Wir freuen uns auch auf den heutigen Tag und gerade auf Ihre Expertise, die ja bei unserem heutigen Thema besonders einschlägig ist.

Damit steigen wir in die Tagesordnung ein. Tagesordnungspunkt 1: Begrüßung und Genehmigung der Tagesordnung. Ich bitte die Mitglieder, Kolleginnen und Kollegen im Rat, sich zu äußern, ob es Bedenken gegenüber der Tagesordnung gibt? – Das ist nicht der Fall. Damit ist die Tagesordnung genehmigt, und wir kommen zum Tagesordnungspunkt 2, der Anhörung der AG Mensch Maschine, heute mit dem Thema der technischen Expertise.

Meine Damen und Herren, wir sprechen in dieser Arbeitsgruppe – der ich ausdrücklich danke, stellvertretend der AG-Leiterin Professor Judith Simon für die Vorbereitung und Konzeption dieser Anhörung – über das ethische Verhältnis von Mensch und Maschine. Ganz besonders nehmen wir da die Technologien der sogenannten künstlichen Intelligenz in den Blick.

Um darüber nachdenken zu können, welche ethischen Herausforderungen, welche ethischen Implikationen sich ergeben daraus, dass diese Technologien unseren Alltag, viele zentrale Lebensvollzüge, viele Bereiche der Gesellschaft, von der Medizin über die Bildung bis hin dazu, wie wir

Demokratie leben und verstehen, durchdringen, brauchen wir eine gute Kenntnis der technischen Expertise und des Forschungsstandes. Ich freue mich sehr, dass es uns gelungen ist, heute vier kenntnisreiche Expertinnen und Experten zu gewinnen, die Ihnen Judith Simon gleich noch vorstellt und die uns hier Rede und Antwort stehen.

Das ist eine klassische Anhörung, die wir heute erleben werden. Wir haben einen Livestream, zwei Versionen davon: einen zweiten Stream mit Untertitelung für Hörgeschädigte und gehörlose Menschen. Das können Sie auf unserer Webseite wie immer frei wählen. Aber diesmal haben wir keinen Online-Chat, sondern diesmal befragen wir aus dem Rat die Expertinnen und Experten, die wir geladen haben.

Wie immer wird diese Veranstaltung dokumentiert und kann im Nachgang auf der Webseite angeschaut werden. Alle Präsentationen, Audio- und Videomitschnitt und die Mitschrift werden auf der Webseite zur Verfügung gestellt.

Zögern Sie nicht, sich dennoch auf Twitter zu beteiligen, und bringen Sie Ideen ein, denn einige von uns sind auf Twitter und werfen immer mal wieder einen Blick darauf. Aber die Anhörung heute ist vor allem dafür da, dass wir unsere Expertinnen und Experten löchern können. Deswegen mache ich keine weiteren Worte, sondern freue mich, dass alle da sind, und übergebe das Wort an Judith Simon, die diesen Teil unserer Sitzung heute moderieren wird. Vielen Dank.

Judith Simon

Vielen Dank, Alena, sehr geehrte Damen und Herren, herzlich willkommen auch von meiner Seite zu unserer öffentlichen Anhörung zum Thema „Künstliche Intelligenz und Mensch-Maschine-Schnittstellen“. Ich bin Mitglied des Deutschen Ethikrates und darf Sie als Sprecherin der AG Mensch und Maschine heute durch das Programm führen.

Wir sind sehr froh, dass sich heute ausgewiesene Expertinnen und Experten bereit erklärt haben, uns Rede und Antwort zu stehen. In den nächsten Stunden werden sie uns einen Einblick in ihre Forschung geben und hoffentlich viele unserer Fragen zum Thema künstliche Intelligenz, maschinelles Lernen und zu den verschiedenen Arten von Mensch-Maschine-Schnittstellen beantworten.

Beginnen wird Frau Professor Ulrike von Luxburg von der Eberhard Karls Universität in Tübingen mit einem Vortrag zum Thema: Ethik und KI: Was kann die Technik (nicht) leisten?

Danach folgt ein Vortrag ihres Kollegen Professor Matthias Bethge, ebenfalls von der Universität Tübingen, zum Thema: Adaptive Intelligenz: Aktuelle Entwicklungen und Disruptionspotenziale des maschinellen Lernens.

Danach haben wir eine 15-minütige Kaffeepause. Im Anschluss daran wird Frau Professor Tanja Schultz von der Universität Bremen über aktuelle Entwicklungen bei kognitiven Systemen im Bereich der Mensch-Maschine-Interaktion sprechen, und zum Abschluss wirft dann Professor Stefan Remy vom Leibniz-Institut für Neurobiologie in Magdeburg für uns noch einen Blick auf die aktuellen Entwicklungen der KI im Bereich der Neurowissenschaften.

Sie finden weitere Informationen zum Programm und zu allen Rednerinnen und Rednern bei uns in der Tagungsmappe.

Bevor ich das Wort an Frau Professorin von Luxburg übergebe, noch einige Hinweise zum Ablauf: Für die vier Vorträge sind jeweils 15 bis 20 Minuten vorgesehen und jeweils im Anschluss gibt es die Möglichkeit für Ratsmitglieder, Rückfragen zu den einzelnen Vorträgen zu stellen.

Nach den ersten beiden Vorträgen gibt es eine 15-minütige Kaffeepause und dann schließen wir an mit zwei weiteren Vorträgen. Danach gibt es erneut eine kurze Pause und dann haben wir noch 90 Minuten Zeit für eine Diskussion mit allen Referentinnen und Referenten. Die öffentliche Anhörung endet mit einem Schlusswort unserer Vorsitzenden Frau Professor Buyx.

Nun möchte ich das Wort Frau Professor Luxburg erteilen. Sie ist Mathematikerin, Informatikerin und Professorin für die Theorie des maschinellen Lernens an der Universität Tübingen. Sie eröffnet unsere Anhörung mit ihrem Vortrag „Ethik und KI – was kann die Technik (nicht) leisten?“.

Ethik und KI: Was kann die Technik (nicht) leisten?

Ulrike von Luxburg · Eberhard Karls Universität Tübingen

(Folie 1)

Guten Morgen, ich freue mich sehr, dass ich hier vortragen darf. Ich bin die erste Sprecherin und fange mit einer kurzen Bestandsaufnahme an.

(Folie 2)

Was kann KI? Was sind Anwendungen, die wir alle im Kopf haben? Es gibt einerseits diese tollen Anwendungen in der Medizin, wo medizinische Bilderkennungssysteme so etwas wie Hautkrebs besser oder zumindest gleich gut vorhersagen können wie die besten medizinischen Experten. Es gibt automatische Textübersetzungssysteme, die ich selber total faszinierend finde: Man […] in einen deutschen Text rein, drückt auf den Knopf und es kommt ein perfekter englischer Text dabei heraus.

Dann gibt es so verblüffende Dinge wie: Ein Computer kann Kunstwerke erschaffen. Wir füttern den mit irgendwelchen Bildern, die es schon gibt, mit alten Kunstwerken, trainieren den und sagen: „Mach ein neues Kunstwerk“, und er macht ein neues Kunstwerk und wir sind total verblüfft und denken uns: Es könnte tatsächlich sein, dass irgendein Künstler dieses Kunstwerk gemacht hat.

Das ist die eine Seite.

(Folie 3)

Die andere Seite sind die anderen Anwendungen, die wir auch alle im Kopf haben, die schlimmen Anwendungen. Wir können mit künstlicher Intelligenz autonome Waffen bauen. Wir können Überwachungssysteme in der größten Konfiguration bauen, die wir uns nur vorstellen können. Es gibt Anwendungen, wo künstliche Intelligenz dafür verwendet wird, vorherzusagen, ob bestimmte Leute ein Verbrechen begehen oder nicht, und darauf basierend werden Entscheidungen getroffen, ob jemand in Untersuchungshaft soll oder nicht. Das sind Anwendungen, wo sich mir persönlich der Magen umdreht.

(Folie 4)

Das waren die zwei Extreme. Dazwischen ist ein ganz großer Bereich und das ist eigentlich der fast wichtigste zu diskutieren. Es gibt viele Anwendungen: Beurteilungen von Kreditanträgen, Vorauswahl von Bewerbungen, Vergabe von Studienplätzen, aber auch Filtern von Nachrichten, selbstfahrende Autos. Da sind eigentlich keine Grenzen gegeben. Das ist ein wichtiger Bereich, wo auch viel Unsicherheit herrscht und wo nicht so klar in Gut und Böse unterteilt werden kann, sondern wo man bei jeder dieser Anwendungen gute und schlechte Aspekte findet und man versuchen muss, zu einer Meinung zu kommen, ob man diese Anwendungen so eigentlich haben will oder unter welchen Umständen.

(Folie 5)

Jetzt bin ich heute die Erste und ich dachte, ich erkläre kurz, wie maschinelles Lernen oder künstliche Intelligenz funktionieren. Jetzt habe ich insgesamt nur 15 Minuten und es dauert wahrscheinlich zwei Stunden, Ihnen das zu erklären. Deswegen möchte ich auf ein Video verweisen, das ich letztes Jahr aufgenommen habe: Wie funktioniert maschinelles Lernen? Das ist eine Erklärung, die sich an jedermann richtet, wo ich mit vielen Beispielen erkläre, wie es funktioniert und wie es aber auch nicht funktioniert. Den Link finden Sie am Ende der Folien und auf meiner Homepage.

Also: Wie funktioniert KI? Im Normalfall fängt es damit an, dass wir erst mal eine Aufgabe definieren. Im Normalfall sagen wir nicht, wir wollen ein intelligentes System bauen, sondern wir wollen ein System bauen, das zum Beispiel Hautkrebs erkennen kann. Das heißt, die Aufgabe ist klar definiert.

Dann haben wir Trainingsdaten, die wir dem System geben müssen. In dem Fall könnten das Bilder sein von Stellen mit Hautkrebs drauf und gesunder Haut, wo ein Mediziner vorher gesagt hat: „Das hier ist gesund und diese andere Stelle ist Hautkrebs."

Dann fängt das maschinelle Lernen an. Jetzt kommt der Lernalgorithmus. Was macht der? Der hat einen Raum von möglichen Beschreibungsfunktionen, und aus diesen Beschreibungen muss der Algorithmus eine aussuchen, die die Trainingsdaten gut beschreibt. Eine Beschreibung ist nichts anderes als eine Funktion, die zu jedem Bild sagt: Ist das jetzt Hautkrebs oder nicht?, also eine Ja-Nein-Antwort liefert. Was der Algorithmus macht, ist: Der hat diesen Raum aller oder vieler möglichen Funktionen und sucht eine, die das auf den Trainingsdaten gut hinkriegt.

Das heißt, an diesem Algorithmus ist eigentlich nichts, was besonders mysteriös wäre, zumindest wenn man die Technik dahinter versteht. Was da drinsteckt, ist im Wesentlichen mathematische Optimierung und Statistik. Die Algorithmen unterscheiden sich nur darin, welchen Raum wir durchsuchen und wie wir ihn durchsuchen.

Das ganze Geheimnis um die verschiedenen Architekturen oder ob man neuronale Netze oder Support Vector Machines oder was weiß ich verwendet, das Grundprinzip ist immer gleich: Ein Algorithmus durchsucht einen Raum nach einer passenden Funktion.

(Folie 6)

Das hört sich ganz einfach an. Was ist denn jetzt der Unterschied oder was ist anders an dieser Beschreibung als an vorherigen Vorhersagemethoden? Es gibt schon immer Statistik, die versucht, aus Daten irgendwelche Dinge zu schließen, und es gibt schon immer eine Wettervorhersage, wo wir Daten aus der Vergangenheit nehmen, um für morgen vorherzusagen, ob die Sonne scheint oder nicht. Was ist jetzt anders an KI?

Ich habe hier ein paar Punkte aufgeschrieben. Das sind vielleicht nicht die einzigen, aber die, die mir eingefallen sind, was ich am wichtigsten finde.

Das Erste ist: KI operiert oft auf Daten, die nicht für diesen Zweck geschaffen worden sind. Wenn eine Firma einen Algorithmus baut, der Gesichtserkennung machen soll, können die sich einfach Bilder vom Internet runterladen und auf diesen Bildern Gesichtserkennung trainieren. Niemals wurden diese Bilder dafür generiert, eine repräsentative Stichprobe oder irgendwas zu sein, und vielleicht sind auf den Bildern auch nicht nur Menschen drauf, sondern auch Bäume und Häuser. Das heißt, das sind keine perfekten Daten, sondern fast Zufallsdaten, aus denen man aber hofft, trotzdem etwas Gutes rausfinden zu

können. Das führt oft zu dem Problem des Bias, das ich nachher ansprechen werde.

Das Zweite: Maschinelles Lernen – oder viele Algorithmen zumindest – macht üblicherweise Vorhersagen, ohne Modelle zu entwickeln. Im Gegensatz zur Wettervorhersage brauchen wir nicht erst ein physikalisches Modell, wie die Winde blasen und wie sich die Temperatur verändern wird, sondern es wird einfach eine Funktion gesucht, die gut klappt, ohne ein Modell zu entwickeln. Das führt oft dazu, dass man das Gefühl hat, gar nicht so recht nachvollziehen zu können, was die Algorithmen eigentlich tun.

Außerdem gibt es ein riesiges Spektrum an Anwendungen, das haben wir schon gesehen.

Und ganz wichtig finde ich diesen letzten Satz auf der Folie: „Jeder kann es." Im Gegensatz zu anderen Technologien – Biotechnologie, genetische Dinge, nukleare Technologie und so weiter – brauchen wir kein großes Labor oder spezielle Experten, die das anwenden können, sondern jeder, der Informatik studiert hat, kann im Prinzip maschinelles Lernen anwenden mit all den guten und schlechten Auswirkungen, die es hat. Das ist, glaube ich, ein Punkt, der es sehr anders macht als andere Technologien.

(Folie 7)

Dem Vorgespräch habe ich entnommen, dass Sie sich dafür interessieren, ob es starke KI geben wird: Ist das wirklich intelligent, was da passiert? Meine sehr klare Antwort auf diese Frage ist: Ich sehe weit und breit keine Intelligenz. Das möchte ich Ihnen erläutern.

Die Durchbrüche, die wir im Moment sehen, diese verblüffenden Sachen – ein Algorithmus komponiert ein Musikstück oder schlägt den besten Go-Spieler durch wahnsinnig kreative Züge – das funktioniert alles auf diesem Verfahren, das ich zu Beginn erklärt habe: Ein Algorithmus durchsucht einen Raum von Funktionen auf schlaue Art und Weise.

Da kommen natürlich oft andere Lösungen raus, als wenn es ein Mensch machen würde. Wenn ich jetzt Go-Expertin wäre und sehe mir diese Situation an und der Computer macht einen Zug, den ich total verblüffend finde, dann ist oft die Situation, dass ich denke: Oh, der muss ja intelligent gewesen sein. Ich kann seit 30 Jahren Go spielen und jetzt kommt der Computer und kann einen tollen Zug, auf den ich nie gekommen wäre. Das Problem an dieser Stelle ist: In dem Moment, wo wir als Menschen verblüfft sind, schreiben wir dem Verfahren Intelligenz zu. Das ist aber nicht das, was hier zum Erfolg führt, und auch nicht das, was dahintersteht.

Die Computer suchen einfach auf eine andere Art und Weise und finden andere Lösungen, und wir finden die dann verblüffend. Aber das heißt nicht, dass die intelligent sind. Ich glaube nicht, dass bei dem, was im Moment passiert, Intelligenz im Spiel ist. Vielleicht sehen das einige der Sprecher anders, aber das ist meine persönliche Meinung.

Was mir aber wichtig ist: Wir können uns stundenlang über dieses Problem unterhalten und in Filmen und Literatur wird das in allen Details ausgebreitet. Es gibt aber sehr viele drängende Probleme im Bereich von KI, die direkt vor der Tür stehen. Ich finde es sehr wichtig, dass wir uns auf die Probleme konzentrieren, die direkt hier sind, und nicht auf das, was vielleicht irgendwann mal passieren wird oder auch nicht.

(Folie 8)

Auf einige dieser Probleme möchte ich kurz eingehen.

Das erste Problem ist der Bias von KI-Systemen. Ich möchte es anhand eines Beispiels erklären. Stellen Sie sich eine Quiz-Situation vor. Wir haben den Algorithmus, der guckt sich Daten an

(Textdaten, von Journalisten verfasst) und soll eine Art Quiz lösen, und das Quiz sieht so aus: Wir sagen: „Frankreich – Paris, Deutschland –“, was ist die passende Ergänzung? Also wie in einer Quiz-Show: Man sieht „Frankreich – Paris, Deutschland“, und dann sollte der Algorithmus sagen: „Berlin“. Wir versuchen, solche Ähnlichkeitsfragen zu finden. So funktioniert dieses Spiel.

Jetzt kann man das Spiel spielen und zum Beispiel sagen: „Mann – Chirurg, Frau“. Was ist die Ergänzung für Frau? Was sagt dann der Algorithmus? Der sagt: „Krankenschwester“. Er sagt nicht „Ärztin“ oder „Expertin“ oder irgendwas, sondern „Krankenschwester“.

Wir können den Algorithmus fragen: „Mann – Programmierer, Frau“, was ist die passende Ergänzung? Da kommt dann nicht „Ingenieurin“ oder so was, sondern da kommt „Hausfrau“.

Wir können das Spiel auch mit Hautfarben machen, also „weißer Mann – Anwalt, schwarz“, was sagt der Algorithmus? Der sagt dann nicht „Jurist“ oder „Richter“, sondern „Fußballer“.

Das ist ein Effekt, der einfach in den Daten steckt. Der Algorithmus wurde trainiert auf Daten, auf Texten, in denen natürlich viele Vorurteile stecken. Diese Vorurteile schnappt der Algorithmus auf, und die kann man nicht wieder rausholen. Man kann sagen: Das ist vielleicht an dieser Stelle nicht so schlimm oder vielleicht kann man das korrigieren. Aber ich möchte ein Beispiel anführen, wo es mir ein bisschen den Magen umgedreht hat:

Vor einem Jahr hatte ich in der Zeitung gelesen, dass ein Arbeitsamt darüber nachdenkt, die Sachbearbeiterinnen und Sachbearbeiter dadurch zu unterstützen, dass man Arbeitslose in Kategorien einteilt: leicht vermittelbar, schwer vermittelbar, um gezielter mit denen arbeiten zu können. Wie soll das passieren? Auf Basis von Daten der Vergangenheit.

Was passiert? In der Vergangenheit waren die Frauen tendenziell teilzeitbeschäftigt; Migranten haben Jobs mit geringer Qualifikation. Wenn wir jetzt ein System auf diesen Daten trainieren, tragen wir diese Vorurteile weiter fort. Das ist das Problem des Bias von KI-Systemen.

(Folie 9)

Was ist die technische Sicht auf diesen Bias? Viele Leute arbeiten daran, ob man den Bias korrigieren kann. Manchmal kann man das ein bisschen. Aber das funktioniert nicht perfekt, und insbesondere gibt es keine technische Lösung, die diese Biases automatisch verhindern kann.

Das führt dazu, dass, wenn man unbedachte Weise Machine Learning anwendet, da oft problematische Dinge herauskommen.

Was man sich an der Stelle immer bewusst machen muss: KI-Systeme sind nicht neutral und können es prinzipiell nicht werden. Die neutrale oder objektive KI, die keine Vorurteile hat, wird es nicht geben.

(Folie 10)

Ein weiteres Problem ist die Fairness. Das ist vielleicht eine spezielle Spielart von dem Problem, das wir vorher hatten. Stellen wir uns ein Beispiel vor: Vergaben von Krediten. Die Bank hat Daten aus der Vergangenheit, bekommt einen Antrag von jemandem und fragt sich: Soll diese Person den Kredit kriegen oder nicht? Und man kann beobachten, dass Leute aus Minderheiten als weniger kreditwürdig eingestuft werden. Der Grund ist ein ähnlicher: Das steckt in den Daten der Vergangenheit drin.

(Folie 11)

Jetzt ist wieder die Frage: Kann man das technisch lösen? So, und jetzt fangen wir an, in die

Eingeweide der Informatik, Statistik und Mathematik einzutauchen. Wenn wir einen Algorithmus trainieren wollen, der fair ist, dann müssen wir ihm sagen, welche Sorte von Fairness er anwenden soll. Jetzt gibt es viele verschiedene Definitionen von Fairness, und das Problem ist: Die schließen sich gegenseitig aus, beweisbar. Das heißt: Es ist nicht offensichtlich, und das ist eines der Probleme in dieser großen Diskussion um das COMPAS [Correctional Offender Management Profiling for Alternative Sanctions]-System in Amerika, wo genau das beurteilt wird, ob Straffällige in Untersuchungshaft bleiben sollen oder nicht. Man kann jetzt verschiedene Begriffe von Fairness anwenden, und die schließen sich aber gegenseitig aus und kommen zu unterschiedlichen Schlüssen.

Was man technisch probieren kann, ist, dass man versucht, dem Algorithmus eine Sorte von Fairness beizubringen, aber nicht alle auf einmal. Das heißt, hier ist von demjenigen, der das System baut, eine Entscheidung nötig.

Das Zweite, was man immer wissen muss, ist, dass Fairness einen Preis hat. Wenn ich den Machine-Learning-Algorithmus auf das beste Ergebnis trainiere und der ist dann nicht fair, hat er die beste „accuracy“. In dem Moment, wo ich aber Fairness mit einbaue, wird die „accuracy“ (oder das erste Kriterium, was ich eigentlich angewendet habe) nach unten gehen. Das heißt, um diese Fairness einzubauen, muss man einen Preis zahlen. Hier ist wieder die Frage, die im Kontext der Anwendung geklärt werden muss: Welchen Preis wollen wir für Fairness zahlen?

Das ist ein sehr komplexes Thema. Wenn Sie mehr dazu wissen wollen: Ich habe im letzten Sommer viele Vorlesungen zu diesem Thema gemacht. Die Videos liegen auf der Homepage und sind – zumindest die ersten – gut verständlich für viele Leute. Sie sind nur auf Englisch.

(Folie 12)

Das letzte Problem, was ich noch ansprechen will, ist Erklärbarkeit. Es wird oft gesagt: Maschinelles Lernen ist intransparent, da kommen diese Deep-Learning-Algorithmen und spucken irgendwas aus, das ist Blackbox, darunter können wir uns überhaupt nichts vorstellen und darauf können wir uns auch nicht verlassen.

Ein Arzt will vielleicht eine Erklärung, warum das System sagt: Hier ist Hautkrebs. Wenn der Kunde bei der Bank keinen Kredit kriegt, würde er vielleicht gerne wissen, woran das lag, oder wenn das selbstfahrende Auto gegen einen Baum fährt und es zum Gerichtsprozess kommt, muss man vielleicht eine Erklärung haben, was das Auto eigentlich gemacht hat, ob es sinnvolle Dinge getan hat oder nicht.

Das ist eine große Forderung, die auch aus der Datenschutzrichtlinie kommt, dass automatische Entscheidungen erklärbar sein sollen.

(Folie 13)

Die technische Sicht: Das ist ein großes Forschungsgebiet und es gibt viele Ansätze, wie man KI erklärbar machen kann. Ich habe aber wieder ein kleines Aber: Diese Erklärungen sind immer Vereinfachungen, und immer, wenn ich eine Vereinfachung mache, muss ich zwangsläufig einen Bias einführen. Das heißt, auch diese Erklärungen sind nicht neutral.

Außerdem können auch automatische Erklärungen manipuliert werden. Also selbst wenn ich eine Firma dazu verdonnere, mir zu erklären, ob sie – was weiß ich – das Geschlecht von Leuten bei der Bewerbung berücksichtigt hat oder nicht, kann man da viele Algorithmen anwenden, und bisher konnte man alle diese Algorithmen auf subtile Art und Weise manipulieren. Das heißt, es kommt eine Erklärung raus, die sich toll anhört,

die aber vielleicht nicht das ist, was der Algorithmus macht.

Das heißt auch aus technischer Sicht: Die Erklärungen können uns im besten Fall unterstützen und es gibt bei medizinischen Anwendungen auch Beispiele, wo das gut klappt. Die sind aber nicht gerichtsfest oder wir können uns nicht auf die verlassen, denn sie können manipuliert werden.

(Folie 14)

Das ist alles zusammen genommen ein Konglomerat, was nicht so ganz einfach ist. Es gibt tolle KI-Anwendungen. Ich forsche in dem Bereich und ich bin begeistert davon, was KI kann. Man muss sich aber bewusst sein, dass KI nicht perfekt werden kann. Wir kriegen das Problem des Bias nicht gelöst, wir kriegen das Fairness-Problem nicht gelöst, wir kriegen die Erklärbarkeit nicht gelöst, zumindest nicht in dem perfekten Sinne, wie man sich das wünschen würde.

Was mir sehr wichtig ist in der Diskussion heute und auch in Zukunft, dass wir vor dem Hintergrund diskutieren, was technisch möglich ist. Denn nicht alles, was wünschenswert ist, ist technisch umsetzbar. Der Grund ist nicht, dass wir das vielleicht noch nicht können und noch fünf Jahre forschen müssen und dann können wir das, sondern das ist prinzipiell nicht umsetzbar. Bestimmte Dinge gehen nicht.

Deswegen ist mir das so wichtig, wenn wir jetzt diese Diskussionen führen, was kann KI, was kann sie nicht, wo wollen wir sie anwenden, wo wollen wir sie nicht anwenden?, dass wir nicht immer – oft wird dann diskutiert: Wie kommt die Ethik in die KI? Und dann werden so Fantasieforderungen gestellt, was die KI alles tun soll, und das kann man halt alles nicht. Man muss einfach vor dem Hintergrund diskutieren, *was* man kann, und das ist sehr beschränkt.

Das ist meine technische Sicht auf diese Ethikdebatte im Bereich KI.

Den letzten Satz konnte ich mir nicht verkneifen. Zu dem Thema bin ich nicht eingeladen worden, aber ich finde es wahnsinnig wichtig, dass wir Regulierung und viel mehr Transparenz brauchen. Wir haben in Tübingen ständig Podiumsdiskussionen zum Thema Ethik und KI. Die Bevölkerung ist total verunsichert (zu Recht, würde ich sagen), weil das undurchschaubar ist, wo KI drinsteckt, wo nicht, wo sie angewendet wird, zu welchem Zweck. Wenn wir da das Vertrauen der Bevölkerung wiedergewinnen wollen, dann brauchen wir dringend Regulierungen in dem Bereich.

Damit bin ich am Ende. Vielen Dank.

Judith Simon

Herzlichen Dank für den Vortrag und die Einführung in das Thema. Ich gebe jetzt das Wort an meine Ratskolleginnen und -kollegen. Gibt es Fragen oder Kommentare zum Vortrag von Frau von Luxburg? –

Armin, ich sehe deine Hand gehoben, und danach Herr Bormann.

Armin Grunwald

Vielen Dank für den Vortrag. Es ist eine für Philosophen beruhigende Sicht der Dinge, die Sie aus dem technischen Feld dort einlegen, die Frage nach der Transparenz, Nachvollziehbarkeit, Verständlichkeit.

Eine meiner großen Sorgen (aber ich bin kein Informatiker, kein Ingenieur, kein Techniker) ist, dass in dieser zunehmenden Komplexität von Algorithmen, die sich selbst verändern, von Big Data, von Riesenmodellen eine opake Struktur entsteht, wo man am Ende nicht mehr weiß und nicht mehr rekonstruieren kann, wie das, was hinten rauskommt, mit dem zusammenhängt, was man vorne reingesteckt hat. In Vorträgen sage ich

gelegentlich: Dann sind wir wieder wie die alten Griechen vor dem Orakel von Delphi, wir können glauben oder nicht glauben, und das ist für eine aufgeklärte moderne Gesellschaft kein guter Zustand. Wie sehen Sie da den Stand der Technik zurzeit?

Franz-Josef Bormann

Vielen Dank, Frau Luxburg, ich wollte in die gleiche Richtung fragen, im Blick auf die Forderung nach Transparenz. Ist es sinnvoll, da zwei verschiedene Formen von Intransparenz oder Opazität zu unterscheiden: eine kontingente, an der man auch was machen kann, und eine prinzipielle? Also hat die Aufklärung sozusagen – sehen Sie irgendwo prinzipielle Grenzen der Einhegung dieser Opazität?

Eine zweite Frage: Das, was Sie zur Fairness gesagt haben, fand ich sehr interessant. Da könnte man natürlich sagen: Wieso soll die Maschine besser sein als der philosophische Diskurs? Wir haben auch in der Philosophie unterschiedliche Vorstellungen von Gerechtigkeit, von Fairness etc. und können nicht erwarten, dass uns die Maschine hier eine Eindeutigkeit schafft, die wir selber im philosophischen Diskurs nicht haben. Vielleicht können Sie dazu etwas haben.

Ulrike von Luxburg

Ich fange vielleicht von hinten an. Sie haben recht, das ist ein Problem, was ist fair und was ist nicht fair, worüber man ewig diskutieren kann, wozu es viele Ansätze und verschiedene Standpunkte gibt, über die man sich oft nicht einigen kann. Aber wenn wir fordern, dass ein Algorithmus fair sein soll, müssen wir ihm irgendwie sagen, was er machen soll. Das heißt, wir müssen an dieser Stelle eine Entscheidung treffen. Das ist der Unterschied dazu, ob ich mir grundsätzlich überlege, was für Formen von Fairness es geben könnte oder ob ich diesen Algorithmus habe und sage, der soll fair sein. In dem Moment, wo ich implementieren will, muss ich mich entscheiden. Das ist dieses Dilemma, das in vielen technischen Systemen steckt, aber was wir an dieser Stelle nicht lösen können.

Mein Plädoyer sagt auch nur: Wir Techniker können das nicht lösen. Jemand anders muss das entscheiden, und dann können wir versuchen, das so einzubauen. Aber wir sind nicht diejenigen, die das entscheiden wollen und können. Dafür sind andere besser geeignet.

Die andere Frage ging in Richtung Transparenz. Man kann immer noch Algorithmen testen. Es ist nicht so, dass ein Algorithmus irgendwie auf dem Mond entsteht und dann eingesetzt wird, und er ist 30 Jahre auf irgendwelchen Daten trainiert, von denen ich nichts weiß, und der tut irgendwas im Geheimnisvollen, sondern wenn so ein Algorithmus angewendet wird, wird er natürlich sehr stark getestet. Den Algorithmus an sich transparent zu machen wird nicht funktionieren. Was man aber transparent machen kann, ist die Prozedur, die dem vorausgehen muss, wann wir einen Algorithmus einsetzen dürfen und wann wir sagen, der ist jetzt ausreichend getestet worden.

Das ist ja beim Flugzeug auch so: Ich muss nicht wissen, wie der Flugzeugmotor funktioniert. Ich muss mich aber darauf verlassen können, dass der TÜV oder wer auch immer das gut genug getestet hat, dass in dem Sinne dessen, was möglich ist, ich mich darauf verlassen kann, dass dies gemacht worden ist. Das ist genau das, was im Moment im Bereich Machine Learning nicht funktioniert: Jeder darf seinen Algorithmus irgendwie – jede Firma darf irgendwas machen und das irgendwie einsetzen. Es gibt keine Prozeduren, wie getestet werden muss, auf welchen Daten, in welchen Szenarien und wann man sagen würde, ja, die haben jetzt genug Arbeit reingesteckt und das ist gut

genug gemacht worden, sodass ich damit leben kann.

Der letzte Punkt, zu dem ich noch etwas sagen will: Es ist auch immer die Frage, was der Vergleichsmaßstab ist. Menschliche Entscheidungen sind auch immer nicht perfekt transparent, und wenn jetzt der Richter entscheidet, ob die Person in Untersuchungshaft kommt oder nicht, hat der natürlich auch interne Bias, da stecken auch komplexe – vielleicht hat er gerade Hunger und ist schlecht gelaunt oder gestern ist was Tolles passiert und er ist gut gelaunt. Da stecken auch viele Einflüsse drin, das heißt, da ist immer auch die Frage: Was ist der Vergleichsmaßstab? Wir können kaum erwarten, dass die KI-Systeme besser werden als das, was die Menschen auch machen würden, und vielleicht auch nicht transparenter. Aber ich glaube, die Transparenz ist nicht so sehr der Mechanismus, sondern eher die Prozedur, mit der wir bei einem Algorithmus anfangen. Das ist das, was ich wichtig finde.

Judith Simon

Herzlichen Dank. Ich sehe im Moment noch zwei gehobene Hände, von Herrn Demuth und von Herrn Gethmann.

Hans-Ulrich Demuth

Vielen Dank. Frau von Luxburg, Ihr Vortrag hat mir gut gefallen und es war in der Kürze, die Sie zur Verfügung hatten, ein schöner Überblick. Was mir besonders gefallen hat, weil es auch teilweise meine Meinung trifft, war der Bezug zur Fehlerhaftigkeit von KI heutzutage. Das haben Sie sehr gut dargestellt.

Ich habe eine Frage. Gibt es KI-Systeme, die schon in der Lage sind, menschliche Empathie, sagen wir mal Trauer zu simulieren und umzusetzen und einen entsprechenden Output zu generieren?

Ulrike von Luxburg

– Soll ich direkt antworten? Ich glaube, man könnte sicher solche Chatbots machen oder was auch immer. Es gibt sehr ausgefeilte Textanalysen, die auch versuchen, die Gefühle aus Texten herauszulesen, und die man wahrscheinlich auch dahin trainieren könnte, dass sie etwas besonders Trauriges schreiben oder so. Ich bin keine Sprachwissenschaftlerin, aber ich glaube, dass das nicht so schwierig ist. Man muss halt den entsprechenden Punkt treffen. So was geht sicher bis zu einem gewissen Grad.

Judith Simon

Vielen Dank. Dann habe ich eine Frage von Herrn Gethmann und danach von Frau Schultz.

Carl Friedrich Gethmann

Ich wollte auf Ihre doch ziemlich pessimistische Aussage hinsichtlich der Formulierung eines Fairnessverständnisses eingehen. Wenn man in dieser globalen Form die Frage stellt: Was ist Fairness?, da haben Sie natürlich recht, das ist aussichtslos. Aber wenn man das Thema stärker kontextualisiert, sieht die Sache nicht so schlecht aus.

Ich stelle zwei Beispiele gegeneinander, die Sie selbst vorgebracht haben. Gerechtigkeit bei der Kreditvergabe: Dass da Bonität und bisherige Rückzahlungsmoral eine Rolle spielen, wird man gut plausibel machen können; da ist eher das Differenzprinzip plausibel. Bei der Studienplatzvergabe wird man die Bonität des Studienplatzsuchenden nicht für ein gutes Kriterium halten und vielleicht eher zu einer Gleichheitsstrategie wie Losvergabe übergehen. Das heißt, je nach Kontext, in dem man die Algorithmen einsetzt, halte ich es für nicht so aussichtslos, zu einem Konsens hinsichtlich solcher normativen Orientierungen zu kommen.

Ulrike von Luxburg

Ja, Sie haben recht, es gibt manche Anwendungen, wo relativ klar ist, in welche Richtung der Fairnessbegriff gehen sollte, und in diesen Anwendungen kann man versuchen, diese Fairness umzusetzen. Technisch geht es bis zu einem gewissen Grad. Man kriegt es nicht perfekt hin, aber man kann zumindest versuchen, das in die Richtung zu schieben.

Trotzdem gibt es viele Anwendungen, wo der Fairnessbegriff genau der Knackpunkt ist. Ich möchte noch mal diese COMPAS-Debatte – ich weiß nicht, ob das allen präsent ist: Da ging es darum, dass man Straffällige in den USA im System hatte, was diese straffälligen Leute dahingehend bewertet, ob sie in Untersuchungshaft bleiben sollen oder nicht, dadurch, dass man ihnen einen Score gibt, der sagt: Wie stark glaubt der Algorithmus, dass diese Person, wen ich sie jetzt freilasse, einen weiteren kriminellen Akt begeht oder nicht?

Da ist es schon sehr subtil (ich habe Folien dazu in meinen Vorlesungen, wenn Sie sich das angucken wollen), je nachdem, welchen Fairnessbegriff ich verwende. Das kommt immer darauf an, welches sind die Bedingungen – wie soll ich das erklären? Also jemand kriegt in diesem System den Score 8 und man kann dann fragen: Macht es einen Unterschied, ob er schwarz oder weiß ist? Nein. Das könnte man als Fairnessbegriff ansetzen. Und das ist das, was die Firma gemacht hat. Das sieht perfekt fair aus und hört sich super an. Man kann umgekehrt aber fragen: Unter denen, die jetzt als unfair betrachtet sind, sind da eigentlich unproportional viele Schwarze drin? Dann kommt raus: Ja.

Das sind sehr subtile Unterschiede, wo ich glaube, dass es in der öffentlichen Diskussion sehr schwierig sein wird, dort überhaupt auf den Begriff zu kommen. Ich weiß nicht, was die Lösung ist, ob die Lösung sein sollte, dass man in diesem Fall vielleicht einfach keine KI anwendet und sagt, man lebt mit der Unfairness, die in dem Richter steckt, oder ob man sagt, man vergleicht die KI mit der Unfairness, die im Richter steckt, aber auch da muss man sich wieder einigen, auf welches Kriterium. Ich habe dafür keine besondere Lösung, sondern ich kann hier nur die technische Sicht darstellen. Das ist oft nicht so einfach. Aber Sie haben recht, es gibt Anwendungen, wo es einfacher ist.

Judith Simon

Vielen Dank, ich habe noch eine letzte Frage von Frau Schultz.

Tanja Schultz

Das ist keine Frage, sondern eher eine Bemerkung oder Antwort auf die Frage von Herrn Demuth zu Empathie erzeugen und erkennen.

Es gibt eine ganze Menge am Markt erhältliche Systeme und Forschungssysteme, die Emotionen von Menschen sehr genau erkennen und auch generieren können. Es gibt beispielsweise Avatare, die in Tränen ausbrechen können. Solche Systeme werden zum Beispiel verwendet, um Autisten zu helfen, die reduzierte Möglichkeiten haben, Emotionen wahrzunehmen, oder auch als Bewerbungstrainer. Da gibt es schon viele Anwendungen. Kürzlich kam von Amazon das Halo auf den Markt. Das ist ein Armband, das man ständig tragen kann und das anhand von Hautleitwert und Stimme relativ gut Auskunft gibt, ob man gut oder schlecht drauf ist. Da gibt es tatsächlich schon marktreife Systeme.

Judith Simon

Vielen Dank für die Ergänzung. Wir haben jetzt noch zwei Minuten Zeit, und da erlaube ich mir selber noch eine Frage zu stellen. Frau von Luxburg, gibt es – Sie haben das vorhin schon angedeutet mit der COMPAS-Software – einen

Bereich, in dem Sie sich gegen den Einsatz von maschinellem Lernen aussprechen würden? Und aus welchen Gründen?

Ulrike von Luxburg

Ja, es gibt viele Bereiche. Ein offensichtlicher Bereich, wo ich sagen würde, das geht zu weit, ist, dass ein Algorithmus nicht darüber entscheiden kann, ob jemand ins Gefängnis geht oder nicht. Es gibt viele Aspekte, die ein Richter beurteilen kann, die ein Algorithmus nicht beurteilen kann, einfach dadurch, dass er die Person vor sich sitzen hat. In dieser Hinsicht würde ich sagen – und ich glaube, ich könnte schwer damit leben, dass meine Tochter im Knast landet, weil ein Algorithmus gesagt hat, sie soll im Knast landen.

Das Argument, das dann immer gesagt wird, ist: Na ja, das ist ja nur ein Assistenzsystem und das macht nur Vorschläge oder nur einen Score, und der Richter entscheidet am Schluss. Auch da bin ich sehr pessimistisch, weil ich denke, dass der Richter – natürlich ist das ein Assistenzsystem, aber hier muss man sich angucken, wie der Richter das benutzen wird. Jetzt kriegt er vorm Mittagessen gesagt: Das System sagt: Dieser Mensch begeht ein Verbrechen, und dann sagt er, klar, macht er und fertig ist die Entscheidung.

Ich glaube, es gibt Systeme, und zwar speziell solche, wo es um Leben oder Tod geht, autonome Waffen und so weiter, wo man sich nicht erlauben kann, einen Fehler zu machen. Dort sollte keine KI eingesetzt werden. Also entweder wo man sich nicht erlauben kann, einen Fehler zu machen, oder wo es unbegründbar, unverhältnismäßig wäre, wenn Fehler gemacht werden.

Es gibt andere Sachen: Kreditvergabe. Ich meine, die Banken machen das seit 50 Jahren, dass sie irgendwelche Scores haben – gut, ob man das jetzt künstliche Intelligenz nennt oder einfach nur schlaues Auf-die-Daten-Gucken. Ich denke, das ist mehr ein Graubereich. Davon geht die Welt nicht unter. Das ist eine Diskussion, die man führen muss. Aber ja, es gibt Anwendungen, wo ich keine KI sehen will.

Judith Simon

Herzlichen Dank. Ich übergebe jetzt an unseren nächsten Redner, Herrn Professor Matthias Bethge, auch von der Universität Tübingen. Er wird sprechen zum Thema adaptive Intelligenz, aktuelle Entwicklungen und Disruptionspotenzial des maschinellen Lernens.

Adaptive Intelligenz: Aktuelle Entwicklungen und Disruptionspotenzial des Maschinellen Lernens

Matthias Bethge · Eberhard Karls Universität Tübingen

(Folie 1)

Herzlichen Dank. Ich bin froh, dass Ulrike von Luxburg einen so fundierten Vortrag zu den Problemen mit der künstlichen Intelligenz gehalten hat. Mein Vortrag wird sich etwas mehr auf einem etwas höheren Level bewegen.

Ein Kommentar noch zu der Frage, wo man es sich nicht erlauben kann, Fehler zu machen: Da würde ich eine andere Nuance setzen. Denn Menschen sind sehr fehlerhaft, und oft wird Ingenieurswesen dazu genutzt, diese Fehlerhaftigkeit zu verhindern. Wobei die Technologie bisher immer sehr eng gestaltet ist und oft nicht das Verständnis vom Ganzen hat. Das ist auch der Stand, auf dem wir uns in der künstlichen Intelligenz und der Regelungstechnik befinden, dass wir bisher immer sehr enge Lösungen entwickelt haben. Darum wird es auch in meinem Vortrag gehen: dass die ersten Möglichkeiten am Horizont sind, dass man nicht nur mit vorgefassten, sehr engen Umwelten

umgehen kann, sondern durch das Selbstlernen der Maschinen mit großen Datenmengen die Möglichkeit besteht, immer mehr mit einer variablen Umwelt umzugehen. Das hat mit dem Thema Adaption zu tun, deshalb der Titel „Adaptive Intelligenz: Aktuelle Entwicklungen und Disruptionspotenzial des maschinellen Lernens“.

(Folie 2–6)

Kurz zu mir: Ich forsche an der Schnittstelle zwischen Neurowissenschaft und künstlicher Intelligenz, bin Leiter des Kompetenzzentrums für maschinelles Lernen in Tübingen, Mitbegründer der ellis[European Laboratory for Learning and Intelligent Systems]-Initiative, wo wir versuchen, die Methodik des maschinellen Lernens auch in Europa voranzutreiben, bin Mitbegründer des Schülerwettbewerbs Künstliche Intelligenz und habe drei Start-ups mitgegründet.

(Folie 7–12)

Eines der ersten Start-ups davon ist direkt aus einer wissenschaftlichen Publikation heraus entstanden, wo wir eine Methode entwickelt haben, wie man Bilder tatsächlich in Kunstwerke umwandeln kann, indem man in diesem Fall von Künstlern den Stil klaut, also die Texturen von einem anderen Bild extrahiert und mit beliebigen Fotos kombinieren kann. Das können Sie auch selbst ausprobieren. Als das herauskam, gab es viel öffentliche Aufmerksamkeit und viele Fragen.

(Folie 13–15)

Hier ist ein Beispiel, wie Sie das ausprobieren können.

(Folie 16)

Dann kam die Frage: Können Maschinen kreativ sein? Die Tatsache, dass das Gestalterische von Maschinen gemacht werden kann, was man früher nicht für möglich gehalten hat, wirft solche Fragen auf, wobei in diesem Fall der Algorithmus definitiv nicht kreativ war. Dennoch denke ich, dass Kreativität durchaus etwas ist, was Maschinen bewerkstelligen können.

(Folie 17)

Ich habe mir überlegt, die Frage des maschinellen Lernens an bekannten Beispielen darzustellen. Es gab vor zwei Jahren einen Vortrag von Demis Hassabis von DeepMind zu den Fähigkeiten von selbstlernenden Systemen.

(Folie 18–19)

Deep Mind hat eine Reihe von großen Herausforderungen gesucht, um zu zeigen, ob man das mit maschinellem Lernen lösen kann. Das erste Beispiel war, ob man Atari-Spiele lernen kann zu spielen. Danach kam AlphaGo, das hat eine große Aufmerksamkeit erzeugt. Was in der breiten Öffentlichkeit vielleicht ein bisschen weniger wahrgenommen wurde, war Alpha Zero, die Weiterentwicklung von AlphaGo, die im Gegensatz zu AlphaGo kein Vorwissen mehr über Bewertungen von Spielsituationen von Menschen berücksichtigt hat, sondern wirklich von Grund auf gelernt hat.

(Folie 20)

Darin ist auch ein wesentlicher Unterschied zu sehen. Wenn man sich erinnert: 1997, als zum ersten Mal der Schachweltmeister von einem Computer geschlagen wurde, hat das nicht zu einer Revolution der künstlichen Intelligenz geführt.

(Folie 21–22)

Bei Alpha Zero ist es jedoch so, dass er nicht nur Go spielen kann, sondern der gleiche Algorithmus, weil er eben kein Vorwissen, kein Expertenwissen brauchte, durch Selbsttraining andere Spiele wie zum Beispiel Schach spielen konnte und dadurch sogar die besten Schachcomputer, die mit Expertenwissen angereichert worden sind,

nach Deep Blue über die letzten zehn Jahre Stockfish, von Alpha Zero geschlagen wurde.

(Folie 23)

Das ist eine sehr starke Demonstration, dass vier Stunden maschinelles Lernen natürlich auf einem sehr großen Computerrahmen die jahrzehntelange Software-Entwicklung der Experten übertrifft.

(Folie 24)

Damals wurde auch in der Süddeutschen Zeitung berichtet. Der abschließende Absatz war wieder eine kritische Hinterfragung:

> „Und auch über den Nutzen von Systemen wie Alpha Go Zero jenseits der Brettspiele wird noch zu diskutieren sein. Denn sein Erfolg beruht auf Kenntnis der Spielregeln".

Insofern wurde der Erfolg ein bisschen in Frage gestellt.

(Folie 25/26)

Interessant ist, dass zwei Jahre später das Spiel Starcraft II gelöst worden ist in dem Sinne, dass es besserbewerkstelligt wird als – dass der beste Spieler jetzt ein Computer ist und nicht mehr ein Mensch. Da ist es nicht mehr so einfach, die Regeln vorzugeben, sondern es muss gelernt werden und ein Modell von der Umgebung geschaffen werden.

(Folie 27)

Ein Schritt weiter ist es so, dass im November 2019 MuZero rausgekommen ist, das die Spielregeln selbst gelernt hat. Das ist eine Weiterentwicklung von Alpha Zero, wo auch die Spielregeln nicht mehr vorgegeben werden müssen, sondern dieses Spiel (wie viele andere Spiele, auch die Atari-Spiele) von einem Algorithmus gelernt werden kann, ohne dass man irgendetwas vorgeben muss.

(Folie 28)

Das hat in den letzten Jahren zu viel Aufmerksamkeit geführt, als nicht nur Spiele, sondern bei einem wissenschaftlich relevanten Problem ein großer Fortschritt erzielt werden konnte: bei der Proteinfaltung, was auch anerkannte Forscher in dem Gebiet als einen Durchbruch ansehen, dass die Vorhersage der Proteinfaltungen von dem Algorithmus sehr gut mit den tatsächlich gemessenen übereinstimmt.

Insofern wird die Generalität, die dem maschinellen Lernen zugrunde liegt (genau wie Frau von Luxburg das schon beschrieben hat), die Suche aus der Auswahl von Möglichkeiten, eine richtige Lösung zu finden, dieses Prinzip, das dem zugrunde gelegt wird, immer effizienter und dadurch auf immer mehr Anwendungen, auch sinnvolle Anwendungen anwendbar.

(Folie 29)

Insofern wollte ich auch die Frage nach der industriellen Revolution stellen, also was für einen Impact diese Technologie hat. Sie kennen sicherlich alle die politische Perspektive, auch die industrielle Revolution, die Einteilung in die Industrie 1.0 bis 4.0.

(Folie 30)

Als Physiker und Wissenschaftler habe ich darauf eine wissenschaftliche Perspektive, die hinter diesen Entwicklungen auch immer wissenschaftliche Entwicklungen sieht. Die industrielle Revolution am Anfang wurde hervorgerufen durch ein besseres oder aufkommendes Verständnis der Thermodynamik, also der Frage, wie viel Energie sich umwandeln und speichern lässt.

Als Nächstes kam, wie sich das über große Distanzen effizient verteilen lässt. Die Stromnetzwerke sind für unsere heutige Zivilisation wesentlich gewesen.

Das wiederum hat es ermöglicht, nicht nur Energie, sondern auch Informationen über lange Strecken zu verteilen, was die Datennetzwerke und das Internet hervorgebracht hat. Das wiederum war eine wesentliche Grundlage für das maschinelle Lernen, weil es jetzt einfach ist, sehr große Datenmengen global zusammenzusammeln.

Gerade am Anfang des maschinellen Lernens spielt es eine große Rolle, dass wir nach wie vor viele Daten brauchen, um Maschinen Dinge lernen zu lassen. Das muss aber nicht unbedingt so sein. Die Erforschung des maschinellen Lernens geht eigentlich dahin, dass man versucht, mit immer weniger Daten robuste Vorhersagen machen zu können. Insofern sehe ich da eine wesentliche Schlüsselrolle in der Entwicklung des maschinellen Lernens.

(Folie 31)

Das ist auch wieder eine Entwicklung. Auch die Dampfmaschine wurde nicht erst in der industriellen Revolution erfunden, sondern auch die Griechen hatten schon darüber nachgedacht, und auch beim maschinellen Lernen gibt es viele Denker in der Historie, die darüber nachgedacht haben.

Ein Denker, den ich ein hervorheben möchte, ist Alan Turing, weil er sich sehr spannende Gedanken auch zur Frage von Mensch und Maschine gemacht hat.

(Folie 32)

Diese schöne Arbeit [Folie: *Computing Machinery and Intelligence*] fängt an mit der Frage: Can machines think? Er hat sich erst mal die Frage gestellt, wie man diese Frage überhaupt sinnvoll beantworten kann. Daher kam der Turing-Test, dass man sagt: Okay, wenn sich eine Maschine in der Interaktion mit Menschen ununterscheidbar verhält wie Menschen, dann kann man davon ausgehen – das ist eine pragmatische Definition davon, dass diese Maschine in der Lage sein muss, die sprechenden Informationsverarbeitungsprozesse zu bewerkstelligen.

Dann hat er eine Abschätzung gemacht, wie viel Informationen man eigentlich in so eine Maschine reinbringen muss, damit die sich so wie Menschen verhalten kann. Da hat er so was abgeschätzt wie 10 hoch 10 bis 10 hoch 15 Bits, und hat vorhergesagt, dass die Entwicklung in der Computerwissenschaft erst mal dahin gehen wird, den Speicher immer stärker zu vergrößern, und dass wir tatsächlich jetzt ungefähr in der Lage sein werden, genügend Speicher zur Verfügung zu haben.

Aber dann bleibt immer noch das Problem: Wie füllt, programmiert man eigentlich diesen Speicher? Da hat er darauf verwiesen, dass es wahrscheinlich wichtig sein wird, Maschinen – also das Programmieren zu ersetzen durch Lernen, so wie Kinder auch aus Erfahrung lernen.

(Folie)
[nicht in der Präsentation enthalten]

Zur Frage: Was ist Intelligenz? Das ist ein schwieriger Begriff, aber ich wollte auf ein Papier hinweisen, das mit meinem Bauchgefühl gut übereinstimmt, was grob gesprochen sagt: die Fähigkeit, mit Situationen umzugehen, die sich von vorhergehenden Situationen unterscheiden, also ganz stark die Frage der Generalisierung aus meinen Erfahrungen: Wie kann ich aus Erfahrungen, die ich in der Vergangenheit gemacht habe, erfolgreich mit neuen Situationen umgehen?

Da ist es so, dass wir trotz der Erfolge, die wir jetzt sehen (ich hab es schon angesprochen), bei den meisten Maschinelles-Lernen-Problemen sehr große Datenmengen brauchen. Wenn man das vergleicht mit Menschen: Wir können oft nach ein, zwei wenigen Erfahrungen unsere gesamte Sicht, unsere Entscheidungen verändern. Dies ist nach wie vor eine Herausforderung.

(Folie 33)

Man kann solche Unterschiede zwischen Mensch und Maschine gut auch beim Sehen illustrieren. Sehen ist jetzt erst mal die Eigenschaft, Objekte erkennen, die uns leicht fällt und die in der Ingenieurswissenschaft schon lange erforscht wird. Am Anfang dachte man: Das Sehen haben wir schnell gelöst. Aber es hat sich herausgestellt, dass Sehen und auch Handeln (also Sachen, die eigentlich jeder Mensch ohne Probleme kann) für die künstliche Intelligenz erst mal besonders schwer gewesen sind.

Trotz der Erfolge des maschinellen Lernens gibt es nach wie vor große Unterschiede. Zum Beispiel sehen Sie hier links die beiden Fotos oben und unten, die sind für uns ununterscheidbar. Maschinen würden das untere Bild als einen Vogel Strauß erkennen oder jede andere beliebige Kategorie, die Sie wünschen. Also man kann minimale Änderungen in den Bildern generieren, die die Entscheidung von Maschinen verändern. Das ist das Problem der adversalen Beispiele, aber auch andere Empfindlichkeiten in der Entscheidungsfindung, zum Beispiel rechts in der ersten Box die Kuh: Wenn die nicht auf einem grasigen Hintergrund ist, sondern in einer Umgebung wie an einem Strand, dann kann das die Standardmaschinen von heute noch verwirren.

Das sind aktuelle Forschungsgebiete, wie man Entscheidungen in neuronalen Netzen robuster machen kann.

(Folie 34)

Eine Herausforderung für die nächste Zeit wird darin gesehen, dass das Verständnis von Kausalität bei der Entscheidungsfindung ganz wichtig ist.

(Folie 35)

Es ist auch für uns im Kompetenzzentrum ein wichtiges Thema.

(Folie 36)

Bei uns geht es auch um die Frage, wie wir die Zukunft gestalten werden. Hier ist ein Bild aus einem Science-Fiction-Film. Mich spricht das nicht so besonders an. Das sieht sehr danach aus, dass sich die Menschen immer mehr an die Technologie angepasst haben, und das ist eigentlich auch der Trend, den wir im Beginn der industriellen Revolution sehen:

(Folie 37)

dass sich der Mensch immer mehr an die Technologie anpassen musste.

(Folie 38)

Auch in der Landwirtschaft: Wenn wir die Felder angucken, sind die sehr darauf ausgerichtet, dass große Maschinen damit gut umgehen können. Hingegen ist Permakultur viel variabler.

(Folie 39)

Die Frage oder die Hoffnung ist – das ist die Hypothese –, dass wir mit intelligenteren Maschinen, die mit der größeren Variabilität der Umwelt umgehen können, eine Trendwende in der industriellen Revolution ermöglichen können: dass sich in Zukunft nicht mehr Mensch, Natur und Umwelt an die Maschinen anpassen müssen, sondern sich die Maschinen an die Mensch und Umwelt anpassen.

Abgeändert von Niels Bohr, der damals gesagt hat: „Wir hängen in der Sprache“, denke ich, dass wir sehr stark in den aktuellen technologischen Möglichkeiten hängen und dass es gut ist, diese Annahmen zu hinterfragen. Zum Beispiel beim Thema Privatsphäre ist – also auch wenn ich zustimme, dass sich viele Dinge nicht lösen lassen, sind doch deutliche Verbesserungen möglich, zum Beispiel durch technologische Möglichkeiten im Federated Machine Learning. Wir brauchen Daten nicht zentral zu sammeln, sondern

können dem Algorithmus auch die Maschinen bringen und dort ein Update machen lassen. Insofern denke ich, dass dieser Austausch, der hier gesucht wird, um die Frage, was die technologischen Möglichkeiten sind, ganz wesentlich ist.

Ich danke Ihnen für Ihre Aufmerksamkeit.

Judith Simon

Herzlichen Dank, Herr Bethge, für den Vortrag. Ich sehe schon die ersten Wortmeldungen und habe als Erstes Herrn Nida-Rümelin auf der Liste und dann Kerstin Schlögl-Flierl.

Julian Nida-Rümelin

Herr Bethge, Sie sprachen am Rande einen sehr interessanten Punkt an. Dazu würde mich Ihre Meinung interessieren. Sie sprachen davon: Wir müssen das eigentlich hinbekommen, dass die Maschinen Kausalität erkennen. Nun besteht in der allgemeinen Wissenschaftstheorie keine Einigkeit darüber, was eigentlich Kausalität ist, aber doch eine Einigkeit, dass das hochgradig theorieabhängig ist. Deswegen meine Frage: Ist das eigentlich ein vernünftiges Ziel? Ist nicht das Feststellen von Kausalrelationen etwas, was letztlich der Wissenschaft im Sinne von Hypothesenbildung usw. überlassen bleibt, weil die Hypothesenbildung selbst etwas ist, was Maschinen nicht können?

Kerstin Schlögl-Flierl

Vielen Dank, Herr Bethge, für den informativen und instruktiven Vortrag. Ich habe vier Nachfragen. Sie haben gesagt, auch die Maschine bewerkstellige Kreativität, und es würde mich interessieren, dass Sie das weiter ausführen.

Die zweite Frage geht zu MuZero. Was ist genau der Unterschied zu Alpha Zero? Ist der Unterschied, dass MuZero die Regeln erschafft? Da würde mich der Unterschied interessieren, denn das habe ich nicht ganz verstanden.

Auch an Sie die Frage (als dritte Frage hier), wo sollte KI eingesetzt werden und wo nicht? Haben Sie auch so klare Entscheidungen wie Ihre Vorrednerin Frau von Luxburg?

Die vierte Frage: Die Schlussthese von Ihnen war, dass Mensch und Umwelt sich an Maschinen anpassen sollen. Ist das wünschenswert in Ihrer Meinung?

Judith Simon

Vielen Dank, und jetzt das Wort an Herrn Bethge bitte.

Matthias Bethge

Ich fange mit der Kausalität an. Ich möchte es nicht auf die Frage der Kausalität, auf die theoretisch exakte Definition von Kausalität versteifen, sondern eher fragen: Was ist robust in der Generalisierung? Ich kann bei der Frage, ob ich ein Objekt erkenne, entweder die Eigenschaft der Form oder der Textur nehmen. Und dann stellt sich heraus, dass sich Texturen vielleicht schneller ändern in einer Umgebung: Wenn sich die Lichtbedingungen ändern, verändern sich die Farben; wenn es schneit, krieg ich irgendein Rauschen drüber; wenn die Sensorik verrauscht ist, ist das ein deutlich unzuverlässigeres Signal. Hingegen ist die Form in der Regel deutlich robuster.

In diesem Sinne kommt es darauf an, Maschinen beizubringen, die möglichst robusten Strukturen in der Welt zu benutzen, um Vorhersagen zu machen. Ich glaube, da gibt es einen engen Zusammenhang mit der Kausalität. Ich würde es im Zweifelsfall zurück auf die Robustheit, auf die pragmatische Robustheit bezüglich der Generalisierung zurückführen.

Dann zu den vier Fragen. Die eine Frage war Kreativität. Kreativität ist eine Mischung aus – also das Neue wird geschaffen, aber nicht dadurch, dass ich komplett zufällig irgendwas Neues würfele, sondern aus Erfahrung die Struktur der Welt

erfasst habe, insbesondere auch die Kompositionseigenschaften der Welt. Das ist genau das, worum es auch beim maschinellen Lernen geht, wenn ich generalisiere. Ich kann nur dann erfolgreich generalisieren, wenn ich die Kombinatorik der Welt verstehe. Und durch die Kombinatorik brauche ich nicht jedes mögliche Beispiel in der Vergangenheit gesehen haben, sondern kann eine exponentielle Menge von Möglichkeiten richtig vorhersagen. Wenn ich diese Kombinationseigenschaften gut verstanden habe, kann ich im Sinne der Kreativität neue Dinge machen, also erdenken oder vorschlagen, die einen Sinn ergeben innerhalb der Kompositionsregeln, die ich erfasst habe.

In dem Sinne sehe ich da keinen prinzipiellen Unterschied zwischen menschlicher und maschineller Kreativität, im Prinzip. Aber im Faktischen, also Stand heute sehe ich da große Unterschiede. –

Judith Simon

Es gab noch die Bereiche, wo KI eingesetzt werden soll und wo KI nicht eingesetzt werden sollte Ihrer Meinung nach.

Matthias Bethge

Ja. Was mir sehr gut gefallen hat bei Frau von Luxburg, war der Hinweis, dass – also ein prinzipieller Unterschied zwischen Mensch und Maschine, also bei Fairness usw., das wissen wir alle, fehlerhaft auch, aber wenn wir jetzt neue Systeme bauen, müssen wir als Ingenieure – also wir bereichern die Welt mit lauter Systemen, die vorher nicht in der Welt waren und die sich irgendwie verhalten und unsere Welt beeinflussen. Insofern ist es eine entscheidende Frage, was für Systeme wir in die Welt setzen und wie die unsere Welt beeinflussen.

Ich glaube, das ist weniger die prinzipielle Frage, ob etwas – also wenn wir beweisen, dass Fairness nicht möglich ist, das trifft ja auch Menschen wie auf Maschine genauso zu, aber jede Anwendung, die wir machen, hat eine Auswirkung auf unsere Lebenswirklichkeit, und deshalb ist es wichtig, da mit Ruhe und Bedachtsamkeit vorzugehen. Ich habe jetzt keine ganz einfachen Regeln dafür, aber ich diskutiere zum Beispiel mit meinen Doktoranden diese Frage: Was sind die möglichen Anwendungen? Und ich finde es gut, sich darüber Gedanken zu machen, sich bei dem Dualitätsproblem, am Anfang, auch bei Grundlagenforschung (bei maschinellem Lernen liegt das ja nahe beieinander) zu überlegen: Hat das einen positiven Effekt oder sind negative Effekte wahrscheinlicher? Wenn wir uns darauf konzentrieren, möglichst viele positive Effekte zu haben, dann hilft das vielleicht, dass wir ein paar negative Effekte vermeiden können.

Judith Simon

Die zwei anderen Fragen von Frau Schlögl-Flierl waren der Unterschied zwischen Alpha Zero und MuZero und die Frage, ob die Anpassung von Mensch und Umwelt an die Maschine wünschenswert ist.

Matthias Bethge

Bei Alpha Zero wurden Simulatoren für das Spiel schon vorprogrammiert, und bei MuZero muss das Spiel die Regeln des Spiels selbst erlernen.

Dass sich die Maschine an den Menschen anpasst, finde ich extrem wünschenswert. Viele technologische Sachen haben die Eigenschaft, dass wir in unserem Verhaltensfreiraum eingeengt werden. Wir gucken alle immer auf Monitoren, benutzen Tastaturen usw. Unser Verhalten verengt sich sehr stark dadurch, dass sich die Interaktion mit Technologie bisher noch auf sehr begrenzte Möglichkeiten reduziert. Wenn wir unsere Wahrnehmung von Technologie und unsere Möglichkeiten, das als etwas Positives zu empfinden, auch zum Beispiel bei der Landwirtschaft, wenn wir

eher eine Multikultur machen können, die weniger CO_2 ausstößt etc. und bessere Böden erlaubt, ist das extrem wünschenswert.

Judith Simon

Vielen Dank, jetzt Herr Gethmann und dann Herr Kruse.

Carl Friedrich Gethmann

Ich komme auf die Frage von Herrn Nida-Rümelin zurück. Der entscheidende Punkt ist ja nicht die Frage, ob der Algorithmus uns irgendwann ein Kausalitätsverständnis bringt, sondern wie es mit der Theoriegebundenheit menschlicher Erkenntnis ist. Sie haben den Begriff der Intelligenz dahingehend expliziert, dass Sie gesagt haben: Die ist gebunden an Informationen aus früher wahrgenommenen Situationen, also diese Situationsgebundenheit, und haben von daher den Turing-Test relativ milde beurteilt, also pragmatisch.

Jetzt muss ich an Sie appellieren als gelernten Physiker: Wissen ist doch eigentlich etwas anderes als situationsgebundene Intelligenz. Also wenn wir wissen, dass Kupfer Elektrizität leitet, dann ja nicht deswegen, weil wir es zehnmal in Situationen erfahren haben, sondern weil wir eine Theorie haben, die uns einen Überschuss über die tatsächlich gemachte situativ gebundene Erfahrung liefert, also uns berechtigt, verallgemeinerbare Aussagen zu machen. Dann spricht man von Gesetzen usw. Ich will diese Metaphern gar nicht aufgreifen, sondern einfach auf diesen Überschuss hinaus. Und so weit und so lange ein System nicht in dieser Weise auch Theorien fabriziert, wird es uns nie diesen Überschuss liefern, und das ist die prinzipielle Grenze der Turing-Test-artigen Ansätze.

Andreas Kruse

Herzlichen Dank für den inspirierenden Vortrag. Sie haben der Intelligenz eine zentrale Position zugeordnet, und wenn Sie sich jetzt die klassische Intelligenz-Definition anschauen, sehen Sie zwei zentrale Unterschiede im Zugang zu unserer Problemlösung. Sie haben das konvergente Denken und Sie haben das divergente Denken. Sie haben sehr schöne Beispiele für das konvergente Denken gegeben, dass Sie gesagt haben: Wenn ich sehr viele Informationen sammle, bin ich in der Lage, auch eine neuartige Situation zu bewältigen. Dieses Neuartige widerspricht aber nicht dem konvergenten Denken. Sie können eine neuartige Situation durchaus mit den Ihnen geläufigen Denkoperationen bewältigen.

Das divergente Denken, sozusagen eine Metapher zu verwenden, das Querständige, das nicht unmittelbar naheliegende Problemlösungskonzept, das hat (jetzt gehe ich auf die Überlegung von Herrn Gethmann ein) viel mit Wissen zu tun. Das heißt, Sie müssen die Welt in einer Weise verstanden, durchdrungen haben, dass es Ihnen möglich ist, einen ganz neuen Bewältigungsansatz, Denkansatz oder ein ganz neues strategisches Konzept einzusetzen. Das wäre eine zentrale Grundlage für Kreativität, damit etwas Innovatives, vorher so noch nicht Gedachtes entstehen kann.

Als Zweites wäre hinzuzufügen, inwiefern Sie auch so etwas wie ein Motiv haben, innovativ zu denken. Zur Kreativität gehört ja nicht nur die kognitive Komponente, sondern auch die motivationale, etwas ganz neu versuchen zu wollen. Wie würden Sie diese beiden Aspekte in Ihre Überlegungen integrieren?

Judith Simon

Herr Bethge, mit der Bitte um eine kurze Antwort. Wir werden das alles später noch in der Diskussion vertiefen können.

Matthias Bethge

Was mir in dem Zusammenhang am wichtigsten erscheint: Maschinelles Lernen ist im Moment noch sehr stark wie eine Regression: Input,

Output, Vorhersage, viele der Sachen funktionieren so, und nicht so sehr Modellbildung. Wir sind gerade dabei, auch Unsupervised Learning, dass Modelle von der Welt gelernt werden. Denen kann man dann alle möglichen Fragen stellen, und da kann man auch mögliche Zukünfte simulieren. Das ist eine spannende Frage, wie wir maschinelles Lernen weiterentwickeln können. Da gibt es viele philosophische Fragen, auf die ich jetzt nicht mehr eingehen kann.

Judith Simon

Das holen wir in der Diskussion später nach. Wir haben jetzt eine kurze Kaffeepause. Bis später! –

Herzlich willkommen zurück zu unserer Anhörung Künstliche Intelligenz und Mensch-Maschinen-Schnittstellen. Wir kommen jetzt zum dritten Vortrag. Ich freue mich auf den Vortrag von Frau Professorin Tanja Schultz von der Universität Bremen. Sie wird sprechen zu aktuellen Entwicklungen bei kognitiven Systemen und im Bereich der Mensch-Maschine-Interaktion.

Aktuelle Entwicklungen bei Kognitiven Systemen und im Bereich der Mensch-Maschine-Interaktion

Tanja Schultz · Universität Bremen

(Folie 1)

Herzlichen Dank, Frau Simon. Ich freue mich sehr, dass ich heute in diesem Kreis über aktuelle Entwicklungen von kognitiven Systemen und Mensch-Maschine-Interaktion sprechen darf.

(Folie 2)

Ich springe gleich hinein, indem ich zunächst eine Definition präsentieren möchte über kognitive Systeme. Das tue ich aus der Perspektive einer Informatikerin. Für mich ist ein technisches kognitives System ein digitales System, was Schnittstellen hat zwischen der digitalen und der realen Welt, und dieses kognitive System kann damit wahrnehmen, verstehen, Schlüsse ziehen und auch lernen.

Zur Erfassung dieser realen Welt, also von Mensch und Umwelt, sind die Schnittstellen mit Sensoren ausgestattet. Ich möchte Ihnen in meinem Vortrag einige praktische Beispiele vorstellen, um die jüngsten Entwicklungen in diesem Bereich herauszuarbeiten. Dabei konzentriere ich mich auf Mensch-Maschine-Schnittstellen, die zur Sprachkommunikation entwickelt worden sind, denn das ist der Bereich, aus dem ich komme und mit dem ich mich bereits seit 20 Jahren befasse.

(Folie 3)

Sie kennen sie vermutlich alle, die digitalen Assistenten und Gadgets wie Handys, Smartphones, Navigationssysteme. Die werden immer kleiner, sind ständig auf Empfang und tragbar, und weil sie so klein sind, hat man inzwischen keinen Platz mehr für Tastaturen. Das heißt, da ist jetzt ein Mikrofon drin und die Interaktion dieser Systeme läuft überwiegend mit Sprache. Das ist praktisch, denn wenn man unterwegs ist, hat man Augen und Hände frei, denn die braucht man zum Sprechen nicht.

Damit diese Interaktion funktioniert, benötigt man Algorithmen, insbesondere die der Spracherkennung und Sprachbearbeitung. Es sind im Prinzip Softwaresysteme. Die nehmen das Sprachsignal, das vom Mikrofon aufgefangen wird, als Eingabe und erzeugen daraus eine textuelle Repräsentation, die möglichst nahe an dem ist, was die Person tatsächlich gesagt hat.

In der Spracherkennung ging es uns so, wie Herr Bethge vorhin schon sagte, mit der Bildverarbeitung und vielen anderen Tasks: Wir haben uns 30,

40 Jahre lang mit viel Ingenieurswissen und Ingenieurinnenfachwissen an eine möglichst gute Erkennungsleistung herangekämpft, und plötzlich, 2017, war man endlich so weit, dass in dem Fall Microsoft als Erstes verkünden konnte, dass Spracherkennungssysteme genauso gut funktionieren und genauso wenig Fehler machen wie Menschen.

(Folie 4)

Das hat insbesondere drei Gründe (zwei davon haben wir schon von Herrn Bethge gehört):

Das Erste sind die Algorithmen, das heißt das maschinelle Lernen, mit denen das möglich war; das Zweite sind enorme Rechenkapazitäten, die heute in riesigen Serverfarmen konzentriert sind, und das Dritte ist der sprunghafte Anstieg von verfügbaren Daten, insbesondere von persönlichen, vielfältigen, komplexen Daten, die in einer rasanten Geschwindigkeit generiert und transferiert werden können.

Aus meiner Sicht kommt aber noch ein weiterer Faktor dazu, der nicht so oft im Fokus steht, aber meiner Ansicht nach wesentlich ist: und zwar ist das die Miniaturisierung von Sensoren, die stattgefunden hat, und die Möglichkeit, diese Sensoren in Mensch-Maschine-Systeme zu integrieren, die Millionen oder Milliarden von Nutzern mit sich herumtragen.

Damit haben nun die kognitiven Systeme praktisch milliardenfache digitale Fenster bekommen, das heißt Fenster von der digitalen Welt in die reale Welt, das heißt ein Blick auf die Nutzer und auf die Umwelt, um damit zu lernen und sich anzupassen.

(Folie 5)

Ich habe Ihnen hier die Entwicklungen von nur den letzten 18 Monaten aufgelistet, was sich also alles getan hat, welche Mensch-Maschine-Schnittstellen auf dem Markt erschienen sind, mit denen Daten aufgezeichnet werden können. Da sind Ringe, die man am Finger tragen kann. Die sind ausgestattet mit Mikrofonen, und wenn man in diese Ringe reinspricht, kann man sich mit Alexa verbinden (das ist ein digitaler Service von Amazon) und sich alle möglichen Informationen geben lassen: Abfahrtspläne von der U-Bahn, Wetter usw.

Ähnliches kann man mit Brillen machen. Da gibt es die sogenannten Echo-Frames. Man kann das auch mit Kopfhörern machen, das sind die Echo-Buds, oder mit neuartigen Kopfhörern, in denen praktisch alle Sensoren, die man sich vorstellen kann, verbaut sind, um beispielsweise den Fitnesstracker im Ohr zu tragen. Man kann sogar Sprachübersetzungen im Ohr machen, man kann Sturzdetektion im Ohr machen. Alle diese Daten werden gesammelt, kontinuierlich 24/7 und ständig vernetzt. Diese Daten und auch die persönlichen Daten sind im Prinzip der Preis, die ein Nutzer oder eine Nutzerin zahlt, wenn sie diese Informationen von Alexa oder so haben möchte. Das heißt, persönliche Daten werden zur Währung der Nutzer, der Serviceleistungen.

Diese Daten enthalten aber auch Informationen über individuelle Vorlieben, über Kaufverhalten, Bewegungsprofile, die Social Bubble. Leider wird dem Nutzer nicht immer ausführlich erklärt, welche Daten tatsächlich aufgezeichnet werden. Daher hat auch der Durchschnittsnutzer keine Ahnung davon, dass er oder sie das Leben gläsern macht, und nicht nur das eigene, sondern auch das seiner Mitmenschen, mit denen er interagiert. Denn die Mikrofone, die hier integriert sind, die hören alles und zeichnen alles auf, was in Hörweite ist.

(Folie 6)

Die aktuellen Sensoren, die wir gerade gesehen haben, sind in Geräte integriert, die wir abends ablegen oder ausziehen können. Die nächste Generation der Sensoren lässt sich allerdings schon auf die Haut drucken, wie wir hier auf einer Hand sehen (das ist eine Entwicklung vom MIT [Massachusetts Institute of Technology]) oder auch unter die Hand injizieren oder implantieren.

Damit sind natürlich die Diskussionen zu Chancen und Risiken in vollem Gang. Aber sie zeigen uns auch, dass wir in Zukunft eine sehr große Bandbreite an Körpersignalen (ich nenne sie Biosignale) aufzeichnen und nutzbar machen können für neuartige Anwendungen.

Damit bin ich wieder bei der Sprachkommunikation, denn wie Sie hier auf der linken Seite sehen: Sprache manifestiert sich nicht nur darin, dass wir ein akustisches Signal haben, was wir mit Mikrofonen aufzeichnen können, sondern Sprache wird produziert, indem zunächst im Gehirn geplant wird, Muskeln innerviert werden und diese Muskeln im Artikulationsapparat sich bewegen und dadurch das Sprachsignal produzieren. Damit bieten sich ganz neue Möglichkeiten, sprachliche Kommunikation zu bewerkstelligen.

(Folie 7)

Wir haben bereits vor 15 Jahren mit einer Idee begonnen, dass wir gesagt haben, wir konzentrieren uns nicht auf das akustische Signal, sondern auf das Abgreifen von Muskelaktivitätssignalen, die entstehen, wenn wir unseren Artikulationsapparat bewegen. Sie sehen hier auf der rechten Seite einen unserer Doktoranden, der diese Elektroden trägt. Das nennt sich Oberflächen-ElektroMyoGraphy [EMG], und diese Potenziale werden nun gemessen, während man spricht. Das Interessante an dieser Idee ist, dass die Muskeln auch dann aktiv und bewegt werden, wenn man die Sprache nicht hörbar produziert, sondern nur lautlos. Also anstatt „Hallo“ zu sagen, könnte man sagen: „–“. Die Bewegungen sind identisch gleich.

(Folie 8)

Damit ergibt sich nun die Möglichkeit einer lautlosen Kommunikation. Wir nennen das Silent Speach Interfaces, denn das EMG-Signal zeichnet die Bewegung auf und nicht das akustische Signal.

Damit bekommen wir ganz neue Möglichkeiten: Wir könnten zum Beispiel lautlos telefonieren, wenn wir im Zug sitzen. Wir können vertrauliche Informationen abhörsicher übermitteln, aber wir können auch die Chance nutzen und Menschen eine Stimme geben, die verstummt sind, beispielsweise durch Unfall oder Erkrankungen wie Kehlkopfkrebs. Damit kann man also neue Möglichkeiten der Sprachkommunikation eröffnen.

(Folie 9)

Wenn man schon mal bei der Muskelinnervierung ist, dann ist der Weg nicht mehr weit bis hin zu den Gehirnaktivitäten. Ein zweites Korrelat der Sprache, nämlich die im Gehirn, ist die, wenn man sich nur noch vorstellt, Sprache zu produzieren. Dann gibt es im Gehirn Aktivitäten, die man hier abgreifen kann.

(Folie 10)

Auch diese Möglichkeit haben wir uns zunutze gemacht, und zwar haben wir mit Patientendaten gearbeitet, die wir von unseren amerikanischen Kooperationspartnern erhalten haben. Das sind Patienten, die aus medizinischen Notwendigkeiten Elektroden implantiert bekommen, um Aufzeichnungen zu machen. Es geht darum herauszufinden, wo, an welchen Stellen im Gehirn starke Epilepsieanfälle ausgelöst werden.

Die Patienten verbringen in der Regel zwei Wochen mit diesen implantierten Elektroden, und wir

haben sie in der Zeit gebeten, mit uns zu kooperieren. Wir haben sie gebeten, Texte vorzulesen, die wir ihnen auf dem Bildschirm gezeigt haben. Und wir haben parallel dazu akustische Sprache aufgezeichnet. Nun haben wir auf der einen Seite das akustische Signal, aus dem wir mit traditioneller Spracherkennung sagen können, wann welche Laute produziert werden, und wir haben parallel zeitsynchron dazu die charakteristischen Hirnaktivitätssignale, die dabei produziert wurden. Das heißt, wir können mit maschinellem Lernen diese Korrespondenz herstellen und als nächsten Schritt praktisch diese Muster verwenden, um kontinuierlich gesprochene Sprache allein aus Hirnaktivitätssignalen zu erzeugen.

(Folie 11)

Das haben wir bereits 2015 gezeigt, und ich möchte Ihnen dazu einen kleinen Ausschnitt aus einem Video zeigen. [ca. 15 Sekunden Video]

Das heißt, wir können dem Gehirn jetzt praktisch zeitsynchron beim Sprechen zuschauen, diese Signale in ein traditionelles Spracherkennungssystem einfüttern und damit Sprache erkennen.

Falls Ihnen dieses Video bekannt vorgekommen ist, dann vielleicht von der Facebook-Ankündigung zu Brain-to-Text im Mai 2017. Die haben nicht nur unseren Begriff übernommen, sondern auch unser Video, haben dem eine professionelle Note gegeben, was ja an sich sehr schmeichelhaft ist, aber wir hätten uns doch gefreut, wenn sie vorher erwähnt hätten, dass das System von uns ist. Es war auch nicht so schön, dass sie dieses Video verwendet haben, um ihre nichtinvasive Technologie anzukündigen für 2017, dass sie aus der Distanz gerne Facebook-Kunden auslesen möchten.

(Folie 12)

Wir haben unser System inzwischen weiterentwickelt, und zwar um Menschen zu helfen, die nicht mehr mit der Außenwelt kommunizieren können, zum Beispiel Locked-in-Patienten. Im Zentrum dieses neuen Systems steht die Idee, dass wir den Menschen in den Loop mit einbringen. Wir wollten, dass das System ein schnelles Feedback, also ohne Zeitverzögerung, an den Nutzer oder die Nutzerin liefert. Damit kann praktisch die Person das hören, was sie sich gerade vorstellt. Also sie hört das, was das System daraus macht und was das System verstanden hat.

Das ist unserer Arbeitsgruppe kürzlich weltweit das erste Mal gelungen, dass so was geht. Sie sind heute die Ersten, die dieses Video sehen; ich zeige Ihnen mal kurz, wie das aussieht [ca. 20 Sekunden Video]

Das heißt, die Person spricht oder stellt sich nur noch vor, zu sprechen, Sprache zu produzieren, und das System bietet unmittelbar die Ausgabe des Gesagten.

(Folie 13)

Die jüngsten Entwicklungen auf dem Gebiet der Interpretation von Hirnaktivitäten bringen auch viele Herausforderungen mit sich. Zum einen muss man sich fragen, wie man solche kognitiven Systeme robust und zuverlässig gestalten kann, wie man sie so gestaltet, dass sie fördern, aber nicht schaden, wie man sie so gestaltet, dass sie Grundrechte wahren, insbesondere auch den Nutzern und Nutzerinnen die Hoheit über ihre Daten überlässt, und transparent vermittelt, was hier aufgezeichnet wird und was das System damit tut, dass man Vertrauen aufbaut, anstatt Vertrauen zu missbrauchen, und Suchtpotenziale vermeidet.

Der zweite große Themenkomplex bezüglich der Gestaltung von kognitiven Systemen ist, *wer* diese kognitiven Systeme gestaltet, das heißt, welchen Bias die Gruppe der Entwickler, die momentan noch überwiegend aus Silicon Valley und neuerdings aus China stammen, hat und vor allem, welches Interesse die Auftraggeber haben.

Für uns als Professoren und Professorinnen ist es natürlich besonders wichtig, dass wir diejenigen sind, die die nächste Generation an Forschern und Entwicklern ausbilden, und wir müssen uns auch Gedanken darüber machen, mit welchen Werten und Vorstellungen wir unsere Doktoranden in die Welt entlassen.

(Folie 14)

Sie hatten mich gefragt, ob ich der Meinung bin, dass kognitive Systeme bereits künstlich intelligent sind. Aus meiner Sicht hängt das davon ab, was man als künstliche Intelligenz versteht. Man unterteilt ja gemeinhin in schwache und starke KI. Aus meiner Sicht erfüllen heutige kognitive Systeme bereits alle wesentlichen Kriterien der schwachen KI: Sie interagieren und kommunizieren mit Menschen, mit anderen Geräten und anderen Services. Sie lernen und optimieren sich selbst, und sie bilden einzelne Aspekte menschlicher Intelligenz bereits erfolgreich und zum Teil sogar schon besser nach, als wir Menschen in der Lage sind, es zu tun. Das heißt, aus der Perspektive würde ich sagen: Kognitive Systeme entsprechen der schwachen künstlichen Intelligenz.

(Folie 15)

Jetzt kann man sich fragen: Ist das schon starke künstliche Intelligenz? Und da bin ich der Meinung, dass das nicht so ist, und vermutlich muss man sagen, dass es *noch* nicht so ist.

(Folie 16)

Damit komme ich zum Ende. Ich möchte mich ganz herzlich bei Ihnen bedanken für Ihre Aufmerksamkeit und natürlich auch bei meinen Mitstreitern, die diese Forschung erst möglich gemacht haben.

Judith Simon

Herzlichen Dank, Frau Schultz, für diesen wunderbaren Einblick in Ihre Forschung. Ich kann mir vorstellen, dass sich da viele Fragen ergeben, auch ethischer Natur. Sie hatten ja auf der letzten Folie schon einige Fragen skizziert, die wir vielleicht später in der Diskussion aufgreifen können. Gibt es Rückmeldungen oder Fragen der Ratsmitglieder? Alena Buyx.

Alena Buyx

Herzlichen Dank für den spannenden Vortrag. Ich muss Sie das jetzt fragen, das haben Sie jetzt einfach so hingelegt mit dem abschließenden angenommenen Unterschied zwischen starker und schwacher KI. Wie ist denn Ihre Prognose? Das geht ein bisschen auf das zurück, was Herr Kruse beim vorherigen Vortrag angesprochen hat. Wie ist Ihre Prognose mit Blick auf diesen Sprung Richtung starker KI? Ist der möglich? Oder ist er grundsätzlich ausgeschlossen? Ist das eine Frage technischer Machbarkeit? Oder sehen Sie da eine grundlegende theoriebasierte Unmöglichkeit, dort jemals hinzukommen?

Judith Simon

Ich würde gerne die Frage von Herrn Kruse noch dazunehmen.

Andreas Kruse

Tanja, herzlichen Dank für den wunderbaren Vortrag. Ich habe zwei Punkte. Du hast dich zur Elektromyographie geäußert. Meine Frage ist: Wie seht ihr das Forschungspotenzial mit Blick auf die differenzierte Abbildung von Emotionen und Affekten? Das ist eine wichtige Thematik mit Blick nicht nur auf Ethik, sondern auf Mimik. Die Frage ist: Können wir diese Emotionen und Affekte durch eine derartige KI dechiffrieren? Das wäre beispielsweise für die Kommunikation mit Parkinson- oder Schlaganfall-Patienten von großer Bedeutung.

Das Zweite geht in die Richtung von Frau Buyx. Ich glaube, wenn man deinen Vortrag anwenden würde auf die Erfassung kognitiver Operationen

bei Patientinnen und Patienten mit einer Demenzerkrankung, dann müsste möglicherweise der Sprung von einer schwachen zu einer starken KI vorgenommen werden, weil man ja die Pathologie der Demenzerkrankten metaphorisch auch als einen unglaublich chaotischen, aber in diesem Chaos vielleicht auch kreativen Prozess interpretieren kann, und das würde bedeuten, wenn wir von einer wirklichen Mensch-Maschine-Interaktion sprechen, dass die Maschine in der Lage ist, in diesem Chaos, das ja in Teilen noch etwas geordnet sein kann, so etwas wie eine gewisse Ordnung zu finden, eine Ordnung vielleicht auch herzustellen und entsprechend auf demenzkranke Menschen zu antworten.

Tanja Schultz

Zur ersten Frage: Ist starke KI wirklich möglich? Grundsätzlich wissen wir momentan, dass mit den maschinellen Lernverfahren so, wie wir sie jetzt betreiben, irgendwas nicht stimmen kann. Das merkt man daran, dass Google beispielsweise – um diese Human Parity, die Spracherkennung zu erreichen, trainieren die aus 100.000 Stunden Daten. Wenn man das mal runterbricht, wie lange ein Mensch bei 100.000 Stunden zuhören müsste, bis er oder sie in der Lage ist, Sprache zu erkennen, da würden wir 20 Jahre alt werden. Also das kann nicht sein, dass der Mensch das so macht wie die neuronalen Netze. Von daher würde ich sagen: Davon sind wir noch ein Stück weit entfernt.

Aber, und das hat Herr Bethge auch gesagt, es gibt inzwischen viele neue Verfahren, bei denen man mit wenigen Daten versucht, Probleme zu lösen mit maschinellen Lernverfahren. Man versucht zu transferieren, das Wissen, was man in einer Domäne erlangt hat, in eine andere Domäne zu übertragen, und das ist letztendlich auch ein Konzept, was der Mensch verfolgt: dass man Erfahrungen, die man auf der einen Seite gesammelt hat, auf der anderen Seite anwendet.

Lange Rede, kurzer Sinn: Ich glaube und manchmal auch fürchte, dass die starke KI möglich ist. Ich würde mich ungern festlegen wollen, jetzt in Jahreszahlen zu gehen, aber ich weiß, dass Wissenschaftler im Schnitt von 20 bis 40 Jahren Zeit ausgehen, bis wir da angelangt sein könnten. Aber ich halte das nicht für ausgeschlossen.

Die zweite Frage von Andreas Kruse zu den Emotionen mit EMG. Vielen Dank, dass du das gefragt hast. Tatsächlich haben wir solche Studien schon gemacht und haben rausfinden können, dass wir auch feinste mimische Nuancen, die ja auf Emotionen hinweisen, mit EMG-Sensoren gut auslesen können. Momentan sieht es noch so ein bisschen ungeschickt aus, denn man braucht alle Sensoren um die Augen und um den Mund. Das ganze Gesicht war verpflastert. Aber ich denke, wenn man die irgendwann injizieren oder aufdrucken würde, könnte man sehr differenzierte Aussagen über emotionale und affektive Zustände machen. Ja, definitiv.

Zu den Kognitionen und Menschen mit Demenz: In der Tat, und das bringt mich noch mal zurück zur Sprache: Sprache ist ein sehr gutes Maß für kognitive Veränderungen, denn Sprechen und Sprachverstehen erfordert sehr hohe kognitive Leistungen, und man merkt an der Sprache, ob sich kognitiv etwas verändert. Wir nutzen auch Sprache und sprachliche Informationen, um herauszufinden, ob die kognitive Leistungsfähigkeit beim Menschen abnimmt. Wir nutzen die beispielsweise auch, um spezielle Aktivierungsprogramme einzusetzen. Dazu haben wir ein sehr erfolgreiches Projekt gemacht, das sogenannte I-CARE, ein System, das sich speziell an vulnerable Gruppen und Menschen mit Demenz richtet, um über Sprache zu kommunizieren und Aktivierungsinhalte anzubieten, die auf individuelle

Möglichkeiten und Eigenschaften angepasst sind, mittels schwacher KI.

Judith Simon

Ich habe noch zwei Wortmeldungen, Herr Bormann und Frau Schlögl-Flierl.

Franz-Josef Bormann

Frau Schultz, ich habe eine Frage, die an den Anfang Ihres Vortrags geht. Ich fand es sehr faszinierend, als Sie sagten, es gibt Sprachprothesen für Locked-in-Patienten etc. Da geht es darum, dass bei einem menschlichen Gehirn, das durch eine Behinderung, eine Krankheit nicht mehr in der Lage ist, zu sprechen, praktisch die Gehirnaktivität über einen Sprachcomputer ausgelesen wird.

Am Anfang Ihres Vortrags sprachen Sie davon, dass diese kognitiven Systeme Dinge verstehen könnten. Offenbar spielt die Sprache, die Spracherkennung, die Sprachsysteme eine Schlüsselrolle in Ihren Forschungen. Meine simple Frage ist: Was heißt es genau für Sie, wenn Sie sagen würden, ein kognitives System kann etwas verstehen? Ist das für Sie primär eine Frage der korrekten sprachlichen Syntax, der Semantik oder auch der Einbettung von Sprachformen in pragmatische Kontexte oder was für ein Verstehensbegriff ist das, mit dem Sie da operieren?

Kerstin Schlögl-Flierl

Frau Schultz, vielen Dank für den instruktiven Vortrag. Ich habe zwei Fragen an Sie. In welchen Bereichen würden Sie sagen, dass KI eingesetzt werden soll? Und gibt es Ihrer Meinung nach Bereiche, wo die KI nicht zur Anwendung kommen sollte?

Die zweite Frage hat sich aus Ihrem Vortrag ergeben. Sie haben gesagt, Sie würden die Mitarbeiter und Mitarbeiterinnen nicht ohne gewisse Wertvorstellungen oder Werte oder Vorstellungen aus dem Labor oder aus ihrem Arbeitsbereich entlassen worden. Ich würde gern wissen, was Sie hier genau meinen.

Tanja Schultz

Die erste Frage zu Verstehen versus Erkennen: Vielen Dank, Herr Bormann, das hatte ich tatsächlich nicht gut rausgearbeitet. Die Spracherkennung ist zunächst mal die reine Transformation von Sprachsignalen in Text. Was da rauskommt, hat der Erkenner aber nicht verstanden, sondern zunächst nur textuell repräsentiert.

Der nächste Prozess, und der ist tatsächlich mit KI in gewissem Sinne schon lösbar, ist das Verständnis. Da gibt es zwei Strategien: Zum einen benutzt man stark domäneneingeschränkte Komponenten, die helfen zu interpretieren, was die Person in einem bestimmten Kontext gemeint haben könnte. Das funktioniert tatsächlich schon. Das heißt, das System versteht, was der Nutzer oder die Nutzerin möchte. Das sind Dialogsysteme, Dialogkomponenten oder bei Alexa wird das beispielsweise mit den Alexa Skills gelöst. Das wäre sozusagen der Gang vom Menschen zur Maschine.

Das, was ich vorhin vorgestellt hatte, hat aber auch eine andere Komponente, nämlich die zwischenmenschliche Kommunikation. Da genügt es, wenn das System eine Sprachausgabe liefert, die tatsächlich vom Menschen verstanden und interpretiert wird. Das heißt, an der Stelle, wenn es um zwischenmenschliche Kommunikation geht, genügt es, wenn das System eine hörbare Ausgabe liefert, die der Mensch, der Empfänger, der Zuhörer, interpretieren kann.

Die zweite Frage war, welche Einsatzgebiete ich für die KI als sinnvoll empfinde. Ehrlich gesagt fürchte ich, diese Frage stellt sich gar nicht. Wir können überhaupt nicht verhindern, wofür KI alles eingesetzt wird. Wir werden ja momentan

schon rechts und links überholt von allem, was auf den Markt kommt und an Services angeboten wird. Ich glaube, unsere Aufgabe als Wissenschaftler, Wissenschaftlerinnen und auch Mitglieder demokratischer Regierungen ist es, uns bewusst zu werden, dass wir dafür sorgen sollten, dass Transparenz, Open Source, Open Data in diesem Bereich KI gepflegt und vorangetrieben wird (damit bin ich schon gleich bei den Werten), dass man auch vermittelt (auch denen, die wir jetzt ausbilden und auf den Arbeitsmarkt entlassen), dass Transparenz, Datenhoheit, Grundrechte wahren ganz wesentliche Aspekte sind, auf die wir als Entwickler und Entwicklerinnen unmittelbaren Einfluss haben, wenn wir diese KI haben.

Was aber die großen Firmen tun, das können wir nicht kontrollieren. Wir können nur dafür sorgen, dass die Erforschung dieser Systeme sehr transparent und offen geführt wird in den wissenschaftlichen Einrichtungen.

Judith Simon

Herzlichen Dank, Frau Schultz.

Wir kommen jetzt zum letzten Vortrag dieser Anhörung von Professor Stephan Remy vom Leibniz-Institut für Neurobiologie in Magdeburg. Er wird uns aktuelle Entwicklungen in der KI im Bereich der Neurowissenschaften vorstellen. Er ist Professor für molekulare und zelluläre Neurobiologie an der Otto-von-Guericke-Universität Magdeburg und Geschäftsführender Direktor des Leibniz-Instituts für Neurobiologie in Magdeburg. Damit gebe ich das Wort an Professor Remy.

Aktuelle Entwicklungen in der KI im Bereich der Neurowissenschaften

Stefan Remy · Leibniz-Institut für Neurobiologie Magdeburg

(Folie 1)

Bevor ich auf aktuelle neurowissenschaftliche Aspekte der Mensch-Maschine-Interaktion eingehe, möchte ich kurz meinen Hintergrund vorstellen. Ich bin von Haus aus Mediziner und Neurobiologe und hatte die Möglichkeit, Erfahrungen in der Epilepsieforschung zu sammeln, auch in der invasiven Diagnostik, und beschäftige mich nunmehr wie das Institut, das ich leiten darf, das Leibniz-Institut für Neurobiologie in Magdeburg, sowohl mit erkenntnisorientierter mechanistischer Grundlagenforschung als auch der anwendungsorientierten Erforschung von Lernen und Gedächtnis.

Das Institut richtet sich derzeit – wie viele Institute – mehr und mehr in Richtung Datenwissenschaften, Neuroinformatik, künstliche Intelligenz und Computational Neuroscience aus.

(Folie 2)

Dieser grundsätzliche Strategiewechsel beruht auf der Erkenntnis, dass Methoden der künstlichen Intelligenz und der Computational Neuroscience die Zukunft der Hirnforschung bestimmen werden. Die Neurobiologie befindet sich gerade in einem Wandel, denn der Erkenntnisgewinn ist heutzutage nicht mehr durch die Verfügbarkeit von Daten limitiert. Die Interpretation der Daten stellt den Flaschenhals dar.

Die Daten, die wir heute generieren, sind komplex, hochdimensional und spiegeln die Komplexität des Organs, mit dem wir uns beschäftigen dürfen und wollen, wider: des menschlichen Gehirns. Und das Feld greift mehr und mehr auf

Methoden der Datenwissenschaften, der Informatik und KI zurück, um verborgene Datenstrukturen zu identifizieren und dann in einer hypothesenbasierten Art und Weise Hirn und Verhalten idealerweise mit dem Ziel in eine kausale Beziehung zu setzen.

(Folie 3)

Invasive Messverfahren, die Aussagen über die Aktivität einzelner Nervenzellen während des Verhaltens zulassen, haben sich kontinuierlich weiterentwickelt, und wir betrachten hier die einzelnen Nervenzellen als die kleinsten Korrelate des Gedächtnisses und unseres Verhaltens allgemein. Und wir lernen, dass Nervenzellen recht selten selektiv an einer Verhaltensäußerung, einem Prozess oder einer Repräsentation beteiligt sind, sondern als Population betrachtet werden müssen. Die Entwicklung der Technologien geht in Richtung höherer Elektrodenzahl, höherer Auflösung, höherer Übertragungsbreite und entsprechend großer übertragener Datenmengen.

Sie sehen das hier illustriert: Die invasiven Systeme, die benutzt wurden, zum Beispiel vom Nobelpreisträger John O'Keefe, um bestimmte Ortszellen zu entdecken, die unsere Position im Raum widerspiegeln, repräsentieren, ließen die Beobachtung von bis zu acht Nervenzellen gleichzeitig zu. Und aktuelle Systeme wie dieses aktuell entwickelte Neuropixel-System erlauben die simultane Aufnahme von bis zu 10.000 Nervenzellen. Sie sehen das hier bildlich gegenübergestellt. Das Erkennen von Zusammenhängen und Strukturen in diesen Daten, also in neuronaler Populationsaktivität, ist gar nicht mehr denkbar ohne den Einsatz von maschinellem Lernen und artifiziellen neuronalen Netzwerken zur Mustererkennung.

Aber setzt man diese Zahl an Nervenzellen, die wir untersuchen können (also Zehntausende), in Relation zur Gesamtzahl der Nervenzellen des menschlichen Gehirns (86 Milliarden), stellt man fest, dass wir an der Oberfläche kratzen. Auch hier arbeitet das Feld an KI-basierten Methoden, die uns erlauben, die Aktivität in anderen Systemen vorherzusagen, ohne dass wir sie messen.

(Folie 4)

Die Humananwendung dieser invasiven Verfahren hat nach wie vor eine exklusiv klinisch-diagnostische Indikation, hauptsächlich in der Epileptologie oder bei Bewegungsstörungen. Über Jahre waren für den Einsatz am Patienten zugelassene invasive Systeme technologisch auf einem deutlich niedrigeren Entwicklungsstand und wurden und werden in Fällen eingesetzt, in denen ein klarer diagnostischer Vorteil für den Patienten erkennbar ist und der Eingriff damit ethisch vertretbar wird.

Die Operationsrisiken sind teilweise erheblich. Es kann bei Implantationen und Anwendungen, gerade bei Langzeitanwendungen, zu Infektionen, Blutungen und Funktionsverlust durch Gewebeschädigung und Narbenbildung kommen.

Sie erkennen die Invasivität hier am Beispiel einer Computertomografie. Der Einsatz von wenigen invasiven Mikrokabeln hat bereits geholfen, Risiken zu reduzieren und erkenntnisorientierte Forschung mit der Auflösung einzelner Nervenzellen an klinisch implantierten Patienten durchzuführen.

(Folie 5)

Das Marktpotenzial dieser invasiven Systeme war bisher recht gering, gerade wegen des beträchtlichen Risikos und der sehr engen Indikationsstellung. Eine Zulassung in nichtmedizinischen Anwendungen erschien da eher unwahrscheinlich.

Aktuell positionieren sich einige Unternehmen aus der Tech-Branche in dieser Hinsicht neu. Ich

führe hier stellvertretend für andere Neurotech-Unternehmen einen der Protagonisten an, Sie kennen ihn wahrscheinlich: Elon Musk, der mit seinem Unternehmen Neuralink Anwendungen in der Allgemeinbevölkerung anstrebt. Genannt werden hier kognitives Enhancement, Computerspiele, telepathische Kommunikation und auch das Schreiben und Lesen von Gedächtnisinhalten.

Der Weg dorthin soll über die Anwendung bei Erkrankungen des Nervensystems wie Epilepsie, Bewegungsstörungen, Parkinson, Alzheimer, Angststörungen und Querschnittslähmung beschritten werden. Auf der Engineeringseite treiben hier professionelle Entwicklerteams die Technologie schnell voran. Die Ziele sind klar: eine kabellose Übertragung, Langzeitverträglichkeit und Austauschbarkeit. Auch an Risikoreduktion wird gedacht bei Implantation durch die Entwicklung von OP-Robotiksystemen (Sie sehen das hier auf der rechten Seite), die eine komplikationslose Implantation, aber auch den Austausch ermöglichen.

Hier muss man bedenken, das muss man klar sagen, dass jede invasive Maßnahme, die das Hirn mit der Außenwelt verbindet, prinzipiell ein Infektionsrisiko darstellt. Aus meiner persönlichen Sicht hat diese Entwicklung positive und negative Auswirkungen. Was die Verbesserung von Krankheitssymptomen angeht, sehe ich ein großes Potenzial. Das sieht man heutzutage schon bei technologisch einfachen Verfahren wie der Tiefen Hirnstimulation bei Parkinson, wo Tremor, also der Schüttelanteil der Erkrankung ganz klar verbessert werden kann.

Auch die intensive Technologieentwicklung wird das wissenschaftliche Feld voranbringen, wobei fraglich ist, ob kommerzielle Produkte die wissenschaftliche Nutzung überhaupt zulassen werden. Den mit dem Marketing verbundenen Erwartungsaufbau, diese Heilsversprechen explizit, implizit, diesen Hype, das sehe ich sehr kritisch. Sie sollen aus meiner Sicht dazu dienen, eine Entwicklung von Technologien zu rechtfertigen, die in der Zukunft privilegierten Bevölkerungsgruppen Zugang zu kognitiven Ressourcen und Informationen bieten, Stichwort ist hier Human Enhancement.

Ich finde, die gesellschaftlichen Auswirkungen sind hier schwer vorherzusehen, und umgekehrt: Wer interpretiert, wer verarbeitet die Information? Wer hat den Zugang? Wer kann lesen, wer kann schreiben? Und wie werden Datensicherheit und Schutz sichergestellt?

(Folie 6)

Ich möchte gern etwas spezifischer auf die Möglichkeiten und Risiken der KI in Bezug auf das Auslesen und Manipulieren von Gedächtnisinhalten eingehen. Als Neurowissenschaftler streben wir nach Erkenntnisgewinn. Wir wollen Mechanismen verstehen. Wir bilden Arbeitsmodelle, die wir dann durch Testung von Vorhersagen systematisch verifizieren.

Aus meiner Sicht verdeutlicht diese Veranschaulichung, die ich hier mache, den Stand der Gedächtnisforschung sehr gut. Wenn der blaue Kreis, den Sie sehen, die Erkenntnisse darstellt, die wir über das menschliche Gehirn gewinnen müssen, um Modelle zu generieren, die uns erlauben, gezielt mechanistisch in unsere Gedächtnisprozesse einzugreifen, also Gedächtnisinhalte auszulesen und gezielt zu manipulieren, dann spiegelt

(Folie 7)

dieser kleine weiße Punkt hier den jetzigen Stand der Wissenschaft wider.

Wir Grundlagenwissenschaftler tun uns oft schwer, uns einzugestehen, dass Erkenntnisgewinn auch ohne mechanistisches Verständnis

möglich ist. Und hier begegnen wir, auch verstärkt durch Neurotech-Marketing, aber auch an der Schnittstelle von Informatik und Neurobiologie häufig einer neuen Argumentation, die ich hier mal in drei Aussagen zusammengefasst habe, bewusst ein bisschen überspitzt.

(Folie 8)

[1] Man muss das Hirn nicht verstehen, um Informationen auszulesen.

[2] Technologieentwicklung allein reicht aus, um Erinnerungen zu dekodieren und zu manipulieren. Wir brauchen kein mechanistisches Verständnis. Sobald eine dauerhafte Schnittstelle mit ausreichender Bandbreite etabliert ist, erledigt maschinelles Lernen den Rest.

[3] Interessanterweise werden wir auf diese Weise neue mechanistische Erkenntnisse über die Funktion des Gehirns erhalten.

Hier wird oft (das haben wir auch schon heute gehört) ein konkretes Beispiel bemüht: Alpha Go, der Algorithmus, der, wie Sie wissen, die weltbesten Go-Spieler in die Schranken wies, zeigte überlegene Spielstrategien auf und führte zum Erkenntnisgewinn.

Aber in gewissem Maße werden die grundlegenden wissenschaftlichen Prinzipien in Frage gestellt. Wir sind an einem Punkt (so sehe ich das), wo wir uns eingestehen müssen, dass die Argumentation berechtigt ist. Es wird bestätigt durch aktuelle Erfolge (gerade von Frau Schultz haben wir das eindrucksvoll gesehen) im Bereich des Einsatzes von KI in motorischen Systemen der Prothetik. Ich führe hier zwei auf: ein Durchbruch in der Steuerung von Handprothesen, ein weiterer ist sicher im Zusammenhang mit den Ergebnissen von Frau Schultz zu sehen, in der Generierung von verständlicher Sprache aus Hirnaktivitätsmustern.

Wenn wir davon ausgehen (und das tun wir), dass die Aktivität Tausender einzelner Nervenzellen Gedächtnisinhalte, Bewegungausführung, Emotionen und sensorische Wahrnehmung repräsentiert, und wir diese Aktivitätsmuster wiederholt aufrufen können, dann ist eine Dekodierung, also das Auslesen von Gedächtnisinhalten, durch Technologie-Entwicklung und natürlich den operativen Zugang zu Gedächtniszentren möglich und ein in absehbarer Zeit lösbares Problem. Wie bald, ist schwer zu sagen. Wissenschaftliche Durchbrüche haben sich historisch als schwer zeitlich vorhersagbar erwiesen.

(Folie 9)

Aus meiner Sicht stellt sich die Situation bei der Intervention, also beim Schreiben in das Gedächtnis etwas anders dar. Ich will hier den Punkt machen, dass es aus meiner Sicht neurobiologische Grundlagenforschung braucht und Modelle, denn wir wissen bereits, dass jegliche Manipulation das System, das wir dekodieren wollen, verändert. Der Schlüsselprozess, den ich hier ganz wichtig finde, ist synaptische Plastizität, also die aktivitätsabhängige Stärkung neuraler Verbindungen, die dann zu einer zellulären Repräsentation von Gedächtnisinhalten führt.

Ich zeige hier eine Kette von Nervenzellen, die einen Gedächtnisinhalt (in Grün) repräsentieren sollen. Wir nennen diese Ketten Engramme. Und die Grundlagenforschung zeigt, dass bereits ein einziger Reiz, der die Aktivität von Nervenzellen erhöht, unweigerlich plastische Prozesse auslöst, also Engramme verändern kann.

Das heißt: Jede Stimulation verändert das System in einer Art und Weise, die wir ohne Modelle und grundlagenwissenschaftliche Erkenntnisse nicht vorhersagen können und die sich mit den klassischen Methoden des maschinellen Lernens auch nicht verlässlich dekodieren lässt.

Wir haben schon gehört, dass sich das Feld weiterentwickelt. Zu bedenken ist auch, dass eine einzelne Nervenzelle an multiplen Prozessen, multiplen Engrammen, multiplen Gedächtnisinhalten beteiligt ist und dass Vorhersagen über die Wirkung der Stimulation einiger Zellen in der Relation zum Gesamtnetzwerk auf ein spezifisches menschliches Verhalten verlässliche Modelle brauchen, auf denen wir den Erkenntnisgewinn basieren.

Ein Beispiel, das ich anführen will, ist das autobiografische Gedächtnis, eine Form des episodischen Gedächtnisses. Das ist eine Grundlage unserer Persönlichkeit, unserer Individualität. Wir Menschen lernen, unser Verhalten aufgrund von Erfahrungen anzupassen, und ohne ein mechanistisches Verständnis, also eine hypothesenbasierte Intervention, müssten wir eine Veränderung von Persönlichkeitsmerkmalen als Wirkung in Kauf nehmen.

Bei Gedanken über den Nutzen der Stimulation zur gezielten Manipulation des Gedächtnisses sollte man bereits jetzt Bedenken äußern und zunächst aus meiner Sicht den Erkenntnisgewinn vorantreiben, denn der mögliche Nutzen wird nicht klar definiert, und kommerzielle wie therapeutische Ansätze sind natürlich denkbar.

Der Schaden, der Menschen zugefügt werden könnte durch fahrlässigen Gebrauch, und auch die Grenzen des Gebrauchs sind noch nicht klar definiert. Das ist der Gedanke, mit dem ich enden möchte.

Ich danke Ihnen für Ihre Aufmerksamkeit.

Judith Simon

Herzlichen Dank, Herr Remy, für den tollen Vortrag, der sicherlich wieder viele Fragen aufgemacht hat. Herr Demuth hat sich gemeldet.

Hans-Ulrich Demuth

Vielen Dank für Ihren Vortrag. Es war spannend zu sehen, wie sich momentan die molekularen Neurowissenschaften mit der Frage der technologischen Entwicklung auseinandersetzen. Ich habe zu den letzten zwei Folien eine Frage.

Um Sie richtig verstanden zu haben: Bringt unter Umständen unter den jetzigen Bedingungen und dem, was wir heute wissen und können, Input in das menschliche System die übereinanderliegenden Schichten, die sich natürlich mit vielen verschiedenen Gedächtnisprozessen beschäftigen, durcheinander?

Judith Simon

Ich würde noch eine zweite Frage hinzunehmen, Frau Buyx hat sich auch gemeldet.

Alena Buyx

Vielen Dank, Herr Remy, das war sehr spannend. Ich wollte Sie bitten, einen Punkt, den Sie aufgeführt haben, noch auszuführen: die Frage von Instrumentalisierung, von Hype, die Sie erwähnt haben, also Ihre Sorge, dass das irgendwann dazu führen könnte, dass ein Zugang zu Technologie des Human Enhancement für besonders Privilegierte möglich wäre.

Das wirft viele Fragen auf, unter anderem gegebenenfalls die Frage der gesellschaftlichen Regulierung. Würden Sie versuchen, diesem Hype aus wissenschaftlicher Perspektive über die eben schon besprochenen, auch von Frau Schultz eingebrachten selbstregulativen Mechanismen und auch in die Gesellschaft wirkenden Mechanismen aus der Wissenschaft zu begegnen? Oder würden Sie dafür plädieren, durchaus auch vorgreifend, da gesellschaftliche Wege von Regulierung, eine härtere Regulierung einzubringen? Was ist da Ihre Perspektive?

Stefan Remy

Ich beginne mit der Frage von Herrn Demuth, ob die vorliegenden Schichten der Informationsverarbeitung des Gehirns durcheinandergeraten. Ich will das mal so beantworten: Wir haben ein anatomisches Verständnis des Gehirns: ein Aufbau, wir haben die Spezialisierung von Hirnregionen, bestimmte Dinge, aber letztendlich ist das Gehirn ein plastisches Organ und wir wissen, dass Hirnregionen problemlos Aufgaben anderer Hirnregionen übernehmen können, gerade relativ früh in der Entwicklung des Gehirns, und weil die Vernetzung viel besser ist und viel weitreichender und wir das im Augenblick immer tiefer lernen, bringt jede Manipulation das System in einem gewissen Maße durcheinander.

Natürlich ist die Hirnaktivität sehr robust und es gibt viel Redundanz, sodass ich mir sicher bin, dass von so einem biologisch hochdifferenzierten Organ Kompensationsmechanismen entwickelt werden. Aber ein Eingriff, der aus nichtmedizinischen Indikationen gemacht wird (also nicht, um Symptome zu verbessern), der verändert. Der kann je nachdem, in welcher Region man eingreift, sehr wohl das Verhalten dauerhaft verändern und auch Persönlichkeitszüge, individuelle Erfahrungen auslöschen oder verändern.

Die zweite Frage finde ich auch total wichtig: Ich sehe die Rolle der Wissenschaft, der staatlich geförderten Wissenschaft, die wir in Deutschland haben, die Informationen über den Stand der Technik – also Veranstaltungen wie diese hier sehe ich als ganz wichtig an, dass die Grundlagenwissenschaft an die Bevölkerung geht, die richtigen Foren findet, um diesem Hype entgegenzusteuern, um die Machbarkeit zu zeigen, um die ethischen Bedenken frühzeitig zu äußern und es der Bevölkerung zu ermöglichen, eine möglichst objektive Sicht der Dinge zu erhalten.

Ich sehe den Hype als Gefahr. Ich sehe die besondere Gefahr in dem Fall darin, dass Krankheiten als Vorwand – also dass das Heilen von Krankheiten nicht das endgültige Ziel der Entwicklung ist, sondern dass an die Folgeanwendungen gedacht wird und dass die positiven Effekte, die bei Erkrankungen sicher erzielt werden können, dass die Weiternutzung in Richtung Human Enhancement ermöglicht wird, und das muss man sich von vornherein bewusst machen.

Judith Simon

Vielen Dank. Wir haben noch vier Fragen. Zunächst Frau Schreiber und dann Herr Kruse.

Susanne Schreiber

Herzlichen Dank für den schönen Vortrag. Ich möchte eine Frage aufgreifen, die wir schon vorher hatten: Mich würde deine Ansicht als Neurobiologe im Grenzbereich zu Computational Neuroscience zur starken KI interessieren. Wir wissen alle nicht, ob sie kommt und wann sie kommt. Aber siehst du grundsätzliche Dinge, die dagegensprechen, also grundsätzliche Probleme, warum so was nicht möglich sein sollte? Wenn wir zum Beispiel in der Lage wären, Mechanismen zu erkennen und das Gehirn wirklich in einem Computerprogramm nachzubilden. Könntest du dir grundsätzlich vorstellen, dass man, wenn man das Gehirn genau kennen würde, auch die menschlichen Fähigkeiten dort abbilden könnte? Oder würdest du als Neurobiologe sagen: Nein, das ist nicht möglich, selbst wenn wir alles wüssten, inklusive der Neuromodulatoren und zellintrinsischen Prozesse, selbst wenn wir das alles nachbilden könnten, wäre das nicht möglich. Diese grundsätzliche Einschätzung würde mich interessieren.

Andreas Kruse

Herr Remy, vielen Dank für Ihren schönen Vortrag. Was mich besonders angesprochen hat, dass

Sie neben Ihrer medizinischen Kompetenz auch die zentralen ethischen Fragen adressiert haben, die wir ja gerade im Bereich der Neuroscience und der Entwicklung der Neuroscience adressieren müssen.

An zwei Punkten würde ich gerne eine wichtige Frage aufhängen. Sie haben beispielsweise von der Tiefen Hirnstimulation gesprochen und sagten, das ist heute, wenn wir Parkinson-Patienten behandeln, fast schon ein gängiges Verfahren. Es wäre noch vor einigen Jahren auch unter ethischen Fragestellungen fast undenkbar gewesen zu sagen: Das wird irgendwann mal „marktgängig" sein.

Vieles von dem, was wir heute im Kontext von Krankheit adressieren, wird möglicherweise in Zukunft auch in einem erweiterten Kontext stehen, dass man sagt: Na ja, der Übergang von Gesundheit und Krankheit ist ein fließender. Beim Enhancement haben Sie das ja.

Meine Frage geht dahin: Welche Aufgabe würden Sie aus einer ethischen Perspektive an den Neurowissenschaftler, an die Neurowissenschaftlerin stellen, um vor solchen möglichen Gefahren, immer abgehoben zu den Potenzialen, zu warnen bzw. auf diese aufmerksam zu machen?

Vielleicht noch ein besseres Beispiel: Sie haben über das episodische Gedächtnis und die neuronale Plastizität in ihrer Möglichkeit, das episodische Gedächtnis in irgendeiner Form zu wandeln oder zu verändern, gesprochen. Jetzt kann ich sagen, wenn ich an dieses episodische Gedächtnis herankommen, kann ich einem Patienten nach Schlaganfall, mit einer Demenz oder mit Parkinson vielleicht helfen, wieder den Zugang zum episodischen Gedächtnis zu finden.

Aber auf der anderen Seite habe ich eine unglaubliche manipulative Potenz, von der Sie ja gesprochen haben. Auch da stellt sich wieder die Frage: Als der Ethiker in Ihnen, wie würden Sie Ihrer Fachgesellschaft, Ihrer Fachwelt kommunizieren, wo hier so etwas wie ein ethischer Scheideweg ist?

Stefan Remy

Ich bin fasziniert von den Fragen. Die erste Frage, Susanne: Ich sehe das so (ich hatte es ja kurz dargestellt), wir kratzen im Verständnis absolut an der Oberfläche. Du hast Prozesse genannt wie Neuromodulation oder Dinge, die wir in der Hirnforschung überhaupt nicht verstehen. Diese Analogie mit dem großen blauen Kreis und dem kleinen weißen Kreis: Ich würde sagen, dass der weiße Kreis eigentlich noch viel kleiner ist. Ich habe ein bisschen übertrieben.

Trotzdem glaube ich, dass das möglich ist, und ich glaube, dass wir Modelle generieren können, die die Hirnaktivität eins zu eins widerspiegeln. Ja. Aber ich sehe die Frage für jetzt noch nicht als so wichtig an, weil ich mir nicht vorstellen kann, dass das eine schnelle Entwicklung ist.

Aber wir müssen uns bewusst werden, aus meiner Sicht, dass das möglich *wird.* Ich bin überzeugt davon, dass, wenn wir das System so nachbilden, wie es ist, und auch die erfahrungsabhängigen Informationen so in das System geben, also eine menschliche Entwicklung und die entsprechenden sensorischen Inputs bis hin zur Informationsvermittlung simulieren können, dass wir am Ende bei einer starken KI ankommen können.

Die zweite Frage: Herr Kruse, die Tiefe Hirnstimulation ist ein gutes Beispiel. Die Erfolge, die man sieht, wenn man die verbesserte Symptomatik bei Patienten sieht, lenken den Blick auf den Nutzen, und es ist völlig richtig, dass man aus dem Auge verliert, was denn der mögliche Schaden wäre oder die möglichen Gefahren sind.

Ich hatte versucht, klarzumachen, dass die Information, die aus der Wissenschaft nach außen, in

die Bevölkerung getragen wird, sehr wichtig ist. Was jetzt Demenzerkrankungen angeht, damit habe ich mich beruflich in den letzten zehn Jahren im Deutschen Zentrum für neurodegenerative Erkrankungen viel beschäftigt. Das ist genau ein Ziel, was ich zum Beispiel in meiner Forschung habe, bei Demenz, wo Nervenzellen zugrunde gehen, aber Gedächtnisinhalte vorhanden sind, aber nicht mehr abrufbar sind, mechanistisch zu verstehen, wie man den Zugriff wieder ermöglichen kann.

Das ist für den Wissenschaftler die Motivation, also gerade medizinisch orientierten Wissenschaftlern, die Möglichkeit der Heilung oder Verbesserung zu bringen. Ich halte es für wichtig, dass man die Debatte in der Wissenschaft führt, denn mir und anderen geht es sicher so, dass man mit dem Ziel vor Augen, zu helfen, wichtige Dinge aus den Augen verliert.

Judith Simon

Ich habe noch zwei Wortmeldungen, aber die würde ich nach der Kaffeepause gleich als Erstes drannehmen. Herzlichen Dank an alle vier Rednerinnen und Redner, wir sehen uns gleich nach der Pause.

Diskussion

Judith Simon

Herzlich willkommen zurück zu unserer Anhörung zu künstlicher Intelligenz und Mensch-Maschine-Schnittstellen. Vor der letzten Pause hatten wir vier Vorträge, die sich diesen Fragen sowohl aus der Perspektive der Informatik als auch der Neurowissenschaften und der Schnittstelle gewidmet haben. Jetzt haben wir ausreichend Zeit, Fragen zu stellen und mit den Expertinnen und Experten zu diskutieren.

Ich möchte beginnen mit der ersten Frage von Herrn Gethmann und direkt im Anschluss die Frage von Frau Klingmüller.

Carl Friedrich Gethmann

Meine Frage ist im Laufe des Vortrags von Herrn Remy aufgekommen, deswegen richte ich die Frage zunächst an ihn. Sie hat aber eine generellere Tragweite. Herr Remy, Sie haben ein wissenschaftsphilosophisches Problem gestreift, und deswegen will ich dahin meine Frage richten, nachdem die ethischen Fragen ja schon angesprochen wurden.

Sie haben gesagt (wenn ich das richtig verstanden habe), dass es aussichtslos sei, die Mikromechanismen aufzuklären und aus denen dann ein Gesamtverständnis des Gehirns zusammenzubauen, sondern dass Sie den anderen Weg gehen über den Versuch, also Modelle aufzustellen und die dann zu falsifizieren. Das ist wissenschaftsphilosophisch in bester Popper'scher Tradition und leuchtet mir auch ein. Ich sehe auch eine Affinität. Wir haben vor Jahren schon mal eine Herbsttagung zum Thema Verständnis des Gehirns gehabt, in Düsseldorf, da hat Herr Jäncke aus Zürich uns gesagt: Wir haben keinen Mangel an Daten (im Gegenteil, durch die bildgebenden Verfahren haben wir Datenfriedhöfe, mit denen wir kaum etwas anfangen können), sondern uns fehlen die Modelle zum Verständnis der Daten. Das konvergiert ja durchaus.

Aber jetzt kommt die Frage: Woher haben Sie die Modelle? Auch Popper sagt ja zu der Frage, woher die Hypothesen kommen, das sei eigentlich egal, die könnten aus Märchen, aus der Bibel oder aus intuitiven Einfällen kommen, Hauptsache, wir haben welche, und die fangen wir dann an zu falsifizieren.

Das scheint mir ein unter Umständen sehr umwegiges und teures Verfahren zu sein. Die Modelle

müssen ja eine gewisse Anfangsplausibilität haben, damit man sie überhaupt als Kandidaten ernst nimmt und sie dann in einen Falsifikationsprozess gehen. Wobei falsifizieren ja vieles heißen kann; das kann heißen, sagen wir mal – klinischer Erfolg wäre ja auch etwas, was in einem Falsifikationsverfahren interessant ist, weil Sie die Tiefe Hirnstimulation und so was ansprechen. Also wenn es funktioniert, dass der Zittereffekt runtermoderiert wird oder sogar zur Ruhe kommt, dann kann man sagen: Das ist eine gute Bestätigung meiner Modellannahme. Es können aber auch rein theoretische Prognosen sein. Da ist vielerlei denkbar.

Mich würde also genauer interessieren, wie Sie die Modelle aufstellen. Wie kommen Sie zu denen und was sind die anthropologischen Hintergrundannahmen, die dazu führen, dass Sie Modelle für plausibel halten?

Zweitens: Wie muss man sich den Falsifikationsprozess genauer vorstellen?

Warum ist das für uns wichtig? Weil das zu der Frage der Belastbarkeit und der Verlässlichkeit der Hypothesen führt, mit denen wir in ethischen Kontexten wiederum anfangen, normative Überlegungen anzustellen.

Judith Simon

Wir müssen gar nicht mehr bündeln, das heißt, ich kann Ihnen gleich das Wort alleine erteilen, Herr Remy. Wir machen jetzt eine Frage nach der anderen, denn wir haben ein bisschen mehr Zeit.

Stefan Remy

Herr Gethmann, vielen Dank. Sie haben viele Antworten schon während Ihrer Frage gegeben, aber ich möchte zu den Modellen etwas sagen. Ich nehme Bezug zu dem Anfang Ihrer Frage, dass der Eindruck, wenn ich den erweckt habe, dass die Mikrountersuchung von Mechanismen nicht wichtig ist. Den wollte ich nicht machen, weil das eigentlich meine Grundüberzeugung ist, dass man datengetrieben, mit genauem Verständnis von Teilsystemen anfängt und dann sieht, wie man generelle Prinzipien daraus ableiten kann, also wie man über andere Hirnareale, Hirnregionen grundsätzliche Prozesse darauf runterbrechen kann und die Feinheiten, die Spezialisierung von neuraler Kommunikation dadurch identifizieren kann und die generellen Prinzipien erfasst. Und dann, wenn man ein realistisches – also man geht in kultivierte Nervenzellen, auch im Tiermodell, testet die grundsätzlichen Prinzipien und setzt die in das Gesamtsystem ein. Man muss also ein Modell haben, das die Komplexität vielleicht auch im ersten Schritt eines Hirnareals wiedergibt.

Wichtig ist, dass man die Eingänge in ein Modell, also die Variablen kontrollieren kann und dass man die Ausgänge kennt. Dann ist die wissenschaftliche Arbeitsweise der Wissenschaftler, Vorhersagen aus dem Modell zu machen, die testbar sind. Das ist wichtig, dass die im Experiment testbar sind.

Je komplexer das Modell wird (das ist im Augenblick unser Dilemma oder unser Problem, also wenn wir jetzt mit 86 Milliarden menschlichen Nervenzellen arbeiten), desto schwieriger ist es, durch unsere Technologie, die wir im Augenblick haben, wirklich testbare Hypothesen zu generieren.

Judith Simon

Vielen Dank. Die nächste Frage kommt von Frau Klingmüller.

Ursula Klingmüller

Ja, und die schließt sich perfekt an und geht auch an Herrn Remy. Ich wollte eben auch – dieser blaue Kreis hat viele beeindruckt, und ich als Grundlagenwissenschaftlerin bin besonders angesprochen worden in dem Moment, wo Sie gesagt haben, dass wir jetzt schon viele Erkenntnisse

erzielen, ohne wirklich den Mechanismus zu verstehen. Wenn uns das Ziel unserer Bemühungen klar ist, also im Zusammenhang mit Demenz, mit Erkrankungen, dann ist das sicherlich verständlich. Aber wenn es um den Bereich Enhancement geht, wo es relativ diffus wird, wo wir gar nicht so genau wissen, welches Ziel wir verfolgen und vor allem welche Gefahren da sind, braucht man da nicht doch Mechanismen, um besser abschätzen zu können, welche Folgen man erzielt?

Stefan Remy

Ich hoffe, das ist in meinem Vortrag klar geworden, dass ich das für ganz wichtig erachte, gerade im letzten Teil, wenn wir sagen, wir manipulieren das System, dass man das nicht hypothesenfrei macht und ausprobiert, wie denn das menschliche Gehirn auf den einen oder anderen systematisch veränderten Stimulus reagiert. Das geht aus meiner Sicht auch gar nicht, diese Art von iterativem Vorgehen.

Ich halte es für ganz wichtig, dass wir mit diesem hypothesenbasierten Erkenntnisgewinn weitermachen und aus den einzelnen Mechanismen, die wir identifizieren, hoffentlich generell gültige Prinzipien der Gehirnfunktionen ableiten können, indem wir schrittweise viele Einzelmechanismen erforschen.

Noch mal zu diesem Kreis: Ich halte es für wichtig, dass wir Grundlagenwissenschaftler bescheiden sind. Wir lernen ja immer, nach außen zu gehen und so Durchbrüche zu kommunizieren. Und da sollten wir wahrscheinlich – also ich denke, das Gehirn ist erschlagend in der Komplexität, und da darf man die Ziele nicht zu hoch ansetzen.

Judith Simon

Vielen Dank. Die nächste Frage kommt von Julian Nida-Rümelin.

Julian Nida-Rümelin

Ich fand das sehr sympathisch, Herr Remy, dass Sie auch die Grenzen des aktuellen Wissensstandes und der Möglichkeiten so deutlich gemacht haben. Sie sind, wie auch Frau Schultz, auf unsere Frage eingegangen, wie Sie das mit der starken und der schwachen künstlichen Intelligenz sehen. Und nun haben wir natürlich nicht begleitend zu dieser Frage noch mal den erreichten Erkenntnisfortschritt im Deutschen Ethikrat vorwegschicken können, und insofern ist die Beantwortung mit Ja oder Nein natürlich extrem unterkomplex.

Die Frage ist: Was versteht man unter starker künstlicher Intelligenz? Das geht innerhalb der Disziplinen oder zwischen den Disziplinen ziemlich durcheinander, da herrscht große Konfusion. Deswegen würde ich diese Frage noch mal präzisieren, damit Sie darauf antworten können.

Also: Wenn man unter starker künstlicher Intelligenz die These versteht, dass aktuelle Softwaresysteme über mentale Eigenschaften verfügen, also erkennen im Wortsinn, empfinden, Emotionen, Einfühlungsvermögen usw. haben, dann kann man eine Antwort geben und sagen: Nein, das glaube ich nicht. Das hatten wir zum Beispiel heute bei Frau von Luxburg als deutliche Antwort.

Und dann kann man die Frage stellen: Ist das grundsätzlich ausgeschlossen? Auch darauf kann man eine Antwort geben und sagen: Ja oder nein.

Den Optimismus, den Sie beide haben, den teile ich, dass sich die verschiedenen Funktionalitäten im Bereich der sogenannten künstlichen Intelligenz zunehmend integrieren lassen und dass wir immer komplexere Systeme entwickeln können, die sehr viel simulieren können, vielleicht sogar so gut wie alles simulieren können, das teile ich. Aber das ist eine Antwort auf eine andere Frage. Die ist auch legitim; so was verstehen manche

auch unter starker künstlicher Intelligenz. Also deswegen noch mal meine Nachfrage an Sie beide zu dieser Fragestellung.

Tanja Schultz

Ich finde Ihre letzte Frage sehr wichtig und interessant, nämlich: Kann man sich vorstellen, dass ein System, ein kaltes, metallisches Konstrukt, Emotionen empfindet, und vor allem: Warum sollte es das tun? Ich glaube, da ist man an einem Punkt, wo man sich fragen muss, mit welchen Zielen optimieren wir denn momentan Maschinen? Im maschinellen Lernen ist das Optimierungskriterium im Wesentlichen, mit möglichst hoher Präzision ungesehene Daten vorherzusehen, mit einer Wahrscheinlichkeit, und je höher die Wahrscheinlichkeit ist, mit der man eine bestimmte vorgegebene Kategorie trifft, umso besser ist es, und danach werden die Modelle optimiert. Das ist nur eines von vielen möglichen Kriterien.

Man könnte sich beispielsweise überlegen, dass man einem System Werte vorgibt, nach denen es optimieren soll, beispielsweise den Wert, dass sich der Nutzer oder die Nutzerin, die mit dem maschinellen System zu tun hat, möglichst wohlfühlt. Das fände ich einen sehr guten Wert, den wir diesem System mitgeben, und wenn dann ein System versucht, diesen Wert zu optimieren, ist das nicht schon nahe an der Empathie?

Julian Nida-Rümelin

Es könnte ja nur eine Simulation sein von Empathie. Dann ist es keine Empathie, sondern eine Simulation.

Tanja Schultz

Ja. – Das stimmt. –

Judith Simon

Das ist eine Frage, die wir vielleicht an alle Rednerinnen und Redner stellen sollten. Ich spitze es noch einmal zu: Gibt es ab einem gewissen Punkt der perfekten Simulation für Sie noch den Sprung zwischen der Simulation und der Realität? Oder ist das für Sie dasselbe?

Denn da gibt es in der Philosophie und in der Informatik ganz unterschiedliche Konzeptionen davon, ob diese Differenz besteht, und die Antwort auf diese Frage, ob da eine Differenz besteht, erklärt zumindest teilweise auch die Antwort auf die Frage, wie man zu starker KI steht. Ich habe mir erlaubt, das ein bisschen zuzuspitzen, aber würde die Frage gern an alle Rednerinnen und Redner ausweiten, bevor ich wieder auf die offizielle Rednerliste zurückgehe.

Herr Remy, wollen Sie vielleicht, weil Sie auch direkt angesprochen waren, noch reagieren? Und dann Herr Bethge und Frau Luxburg, in der Reihenfolge des Erscheinens jetzt gerankt.

Stefan Remy

Ich finde es gut, dass wir das ausweiten. Ich habe mich mit der Frage nur im Zusammenhang mit dieser Veranstaltung auseinandergesetzt und längst nicht so differenziert, wie Sie das gemacht haben, Herr Nida-Rümelin.

Mein Denken basiert darauf, dass ich in meiner eigenen Arbeit zum Beispiel realistische Modelle von einzelnen Nervenzellen in Software generiere, die durch Parametersuche, Experiment möglichst realistisch werden und sich dann kaum von einer realen Nervenzelle unterscheiden. Das geht, aber ist natürlich immer eine Vereinfachung, aber die unterscheiden sich kaum in der Funktion. Jetzt weite ich das in Gedanken aus und denke: Okay, wenn ich die Eingänge auf diese Nervenzelle und die Ausgänge verstehe, simulieren kann, dann kann ich hier anfangen, diese Nervenzellen zu verknüpfen, kann sensorische Eingänge, kognitive Prozesse simulieren, und habe am Ergebnis ein menschliches Gehirn, das die

Funktionsprinzipien eines biologischen Gehirns hat. Aber es ist deswegen immer noch eine Simulation.

Judith Simon

Danke. Herr Bethge?

Matthias Bethge

Ich habe im Wesentlichen eine Frage auch vielleicht als Laie. Ihre Frage hat mich an die Unterscheidung zwischen Intelligenz und Bewusstsein erinnert: Wer empfindet da eigentlich was? Vielleicht können Sie da noch mal nachjustieren, so wie Sie die Frage gestellt haben. Auch Ihr Kommentar zu Frau Schultz, ob das eine Simulation ist oder die Maschine wirklich was empfindet – das hat ja viel mit dem Explanatory Gap von Bewusstseinsfrage zu tun oder der Zombie-Hypothese: Wissen Sie, ob ich irgendwas empfinde, oder verhalte ich mich da nur? Insofern die Rückfrage von mir, wie Sie das auffassen würden.

Julian Nida-Rümelin

Das ist ein wichtiger Punkt. In der Zeit, in der Turing den berühmten Aufsatz in *Mind* schrieb, der den Turing-Test hervorgerufen hat, war die Stimmung stark behavioristisch, das heißt, man sagte sich, in der Sprachphilosophie, aber auch in den Sozialwissenschaften: Wir müssen dieses ganze mentalistische Gerede über mentale Eigenschaften und so loswerden. Wir müssen das ersetzen durch etwas anderes, was öffentlich zugänglich ist, was auch wissenschaftlich verwertet werden kann usw.

Darüber sind wir in der Philosophie jedenfalls längst hinaus, denn man kommt in ganz verrückte Situationen: Dann hat der Super-Spartaner keine Schmerzen, ja? Irgendwie klingt das komisch. Natürlich kann der Super-Spartaner, der keine Schmerzäußerungen hat, immer noch Schmerzen haben. Die Zugänglichkeit ist natürlich ein Problem, und in der Wissenschaft muss alles intersubjektiv nachprüfbar sein. Deswegen sind mentale Zustände erst mal nur indirekt Gegenstand der Wissenschaft.

Wenn wir uns allerdings kritisch prüfen, wissen wir, dass die ganze wissenschaftliche Praxis auf einer Selbstverständlichkeit der interpersonellen Zuschreibung von mentalen Zuständen, Erwartungen, Überzeugungen, Entscheidungen usw. beruht. Sonst könnte der wissenschaftliche Austauschprozess gar nicht in Gang kommen. Das heißt, wir präsupponieren, dass wir Zugang haben zu dieser Realität mentaler Zustände.

Davon hängt natürlich viel ab, denn Entscheidungen, Einstellungen, Bewertungen, Prognosen – das ist alles aufgeladen. Da ist jemand, der bewertet, der oder die eine Einstellung hat, eine Person oder so was. Deswegen hängt da relativ viel dran.

Und in einer Lesart ist halt […] die künstliche Intelligenz die These, ja, da sind wir dran. Wir sind auf dem Wege, ein Gegenüber zu schaffen, mit dem wir dann interagieren, kommunizieren, *echt* kommunizieren, nicht nur simuliert kommunizieren. Und andere, diejenigen, die einem behavioristischen Paradigma verhaftet sind, sagen: Ach, das können wir doch gar nicht unterscheiden. Wenn das so aussieht *wie*, dann sollten wir auch.

Aber wenn wir diesen behavioristischen Standpunkt einnehmen, kommen wir in des Teufels Küche. Dann könnte man sagen: Virenpopulationen sind intelligent, weil sie sich intelligent verhalten auf bestimmte Herausforderungen. Aber das würden wir natürlich nicht akzeptieren; da ist niemand, der über Intelligenz verfügt. Das ist ein Prozess der Evolution, der dazu geführt hat, dass sich Virenpopulationen so verhalten, als wären sie intelligent.

Meine Sorge ist ein bisschen, dass wir durch die KI-Entwicklung diese erreichten gedanklich-begrifflichen Klärungen wieder aufgeben.

Judith Simon

Vielen Dank, Herr Bethge, das war eine Rückfrage von Ihnen. Dann gebe ich das Wort noch mal an Sie zurück und dann an Frau von Luxburg.

Matthias Bethge

Ich glaube, dass die Frage, ob mein Computer Bewusstsein hat oder nicht, natürlich für die Ethik sehr relevant sein kann, zum Beispiel wenn wir es einem Menschen nicht zumuten, dass er die ganze Zeit von mir versklavt wird – also dem Computer muten wir es auf jeden Fall zu. Insofern verstehe ich das.

Auf der anderen Seite glaube ich, ist es trotzdem wichtig, oder ich würde es stark von einem Intelligenzbegriff trennen. Ich würde mir nicht erlauben oder nicht wagen, eine Definition davon zu geben, aber für mich reserviert sich das sehr stark für die Fähigkeit, Informationen zu verarbeiten, Modelle von der Welt zu bauen, und da glaube ich, das unterscheidet sich auch von einem Virus, das eben kein inneres Modell von der Welt baut und dadurch Handlungsoptionen planen kann.

Insofern würde ich sagen für die Frage Intelligenz: Für mich, mit meiner Terminologie, ist diese Bewusstseinsfrage nicht so relevant. Bewusstsein per se kommt mir nicht besonders intelligent vor, aber es ist natürlich eine wichtige Realität von jeden von uns, die auch bei den ethischen Bewertungen eine wichtige Rolle spielt.

Ulrike von Luxburg

Ich wollte an einen Kommentar anschließen, den Frau Schultz vorhin gemacht hat, denn sie hat gesagt: Na ja, jetzt könnten wir versuchen, die Systeme darauf zu optimieren, dass der Benutzer sich zum Beispiel damit wohlfühlt, und sie hat dann etwas provokant gefragt, ob das nicht Empathie ist.

Darauf würde ich provokant antworten: Nein, denn das passiert doch schon die ganze Zeit. Facebook versucht uns die Nachrichten so darzustellen, dass wir uns besonders wohlfühlen, dass wir uns in unserer Meinung bestätigt fühlen, und sie können messen, wie lange ich mich auf der Facebook-Seite aufhalte. Natürlich messen die nicht, wie zufrieden ich bin, aber sie haben einen Proxy, indem sie messen, wie lange ich mich auf ihrer Seite aufhalte, mit dem sie die Funktion meiner Zufriedenheit annähern können. Das heißt, sie können schon versuchen, mich lang auf ihrer Seite zu halten, und ich bin dann sehr zufrieden. Aber deswegen hat Facebook oder der Algorithmus, der dahintersteckt, keine Empathie.

Das ist wieder dieser Duktus: Man kann versuchen, alle möglichen Dinge durch Proxys zu messen und auch herzustellen, also so was wie, dass der Mensch sich damit wohlfühlt. Aber das heißt nicht, dass das System, das diesen Zustand herbeiführt, irgendwas von Empathie versteht oder über Empathie verfügt. Das ist genau dieser Unterschied, wo man immer sehr vorsichtig sein muss. Um einen bestimmten Zustand herzustellen, muss das System diesen Zustand nicht selbst besitzen.

Tanja Schultz

Ich freue mich, dass Sie diese Provokation aufgenommen haben, und ich gebe Ihnen 100 Prozent recht. Letztendlich sind wir damit auch in einer aus meiner Sicht viel wichtigeren Frage, nämlich: Wie wollen wir den Weg weiter gestalten? Das Problem ist: Das liegt nicht mehr in unserer Hand. Das liegt in Facebooks Hand und in Googles Hand, und ich finde, wir sollten auf der einen Seite nicht müde werden, den Firmen immer wieder zu sagen, dass wir Datenhoheit haben wollen, dass wir das selbst kontrollieren möchten, aber wir dürfen nicht aus den Augen verlieren, dass wir da schon sehr viel Boden verloren haben und dass momentan viele Entwicklungen, die in diesem Bereich liegen, nicht von ethischen und

moralischen Grundsätzen geführt werden, sondern von interessengetriebener Industrie.

Da sollten wir in demokratischen Regierungen, in neutralen Einrichtungen, wissenschaftlichen Einrichtungen, einen Gegenpunkt setzen und dafür kämpfen, dass die Systeme, die wir entwickeln, wirklich transparent sind, dass klar ist, was hier optimiert wird und was man erreichen möchte. Dazu müssen wir diese Ziele fest im Auge behalten und auch formulieren.

Judith Simon

Herzlichen Dank. Die Nächste auf meiner Liste ist Kerstin Schlögl-Flierl.

Kerstin Schlögl-Flierl

Vielen Dank. Meine Frage geht an alle vier Vortragenden. Ich würde gern beim Begriff Verantwortung als einer der ethisch virulentesten Fragen anknüpfen und Sie um eine kurze Definition bitten: Wer ist für was vor wem und vor welcher normativen Prämisse verantwortlich? Und ich würde Sie bitten, den Verantwortungsbegriff für Ihren Forschungsbereich zu konzeptualisieren. Wo kann Verantwortung übernommen werden und wo kann sie nicht mehr übernommen werden?

Judith Simon

Vielen Dank. Soll ich, wenn keine Reihenfolge vorgegeben wird, in alphabetischer Reihenfolge vorgehen? Das gibt Ihnen eine Orientierung, wann Sie dran sind. Dann fange ich mit Herrn Bethge an, dann Frau Luxburg, Herr Remy und Frau Schultz.

Matthias Bethge

Ich sag Ihnen einfach, was mir spontan dazu einfällt; das wird, glaube ich, dem Anspruch Ihrer Frage in keinster Weise gerecht.

Wir sind in einer Welt, die in unserer Umgebung in einem Zustand, in einer Art Gleichgewicht oder in einem Fließgleichgewicht ist. Dinge findet man erst mal vor, wie sie sind, und hat relativ begrenzte Kontrolle daran. Das merkt jeder, egal in welcher Hierarchieebene man agiert, dass Dinge sehr schwer zu kontrollieren sind. Inklusive was unsere Simulation wieder von den Auswirkungen unserer Handlungen betrifft, sie sind fehlerbehaftet usw. Die Fehler übernehmen wir auch von den Denkmodellen anderer, die wir in der Schule gelernt haben etc. Also es ist eine hochgradig arbeitsteilige Welt. Nicht umsonst gibt es ja […] das soziale Gehirn, also es gibt ja gar nicht das einzelne Gehirn, das das ganze Wissen angesammelt hat, sondern wir verlassen uns ununterbrochen darauf, was andere sagen, und das macht es extrem schwer, Verantwortung zu übernehmen in dem Sinne, dass ich wirklich kontrollieren könnte, was ich da mache.

Ganz pragmatisch gesagt versuche ich, unter diesen Annahmen so zu handeln (und da spreche ich ein bisschen mit meiner Analogie als Ingenieur oder so), einen Gradientenschritt, also einen Optimierungsschritt in die richtige Richtung zu machen, sodass mein Handeln mit dazu beiträgt, dass sich die Welt in eine Richtung entwickelt, die mir sinnvoller und besser vorkommt. Das ist jetzt sehr wenig regulativ betrachtet, aber ich wollte trotzdem diese Perspektive zur Verfügung stellen.

Ulrike von Luxburg

Ja, die Frage ist, in welche Richtung Ihre Frage gedacht ist. Es ist klar, die Verantwortungsfrage ist ein großes Problem. Wenn ich jetzt an das selbstfahrende Auto denke und es baut einen Unfall, wer ist dafür verantwortlich? Ist es der Softwareentwickler, war es der Forscher, der den Algorithmus entwickelt, oder die Firma, die den Algorithmus auf die Straße gebracht hat? Das ist eine Frage, wo ich als Informatikerin herzlich wenig dafür qualifiziert bin, darauf irgendwelche sinnvollen Antworten zu geben.

Was aus Sicht der Informatik dazu vielleicht zu sagen ist, ist, dass die ganzen Algorithmen, die wir entwickeln (also zumindest bei uns an der Uni), generische Algorithmen sind, also Algorithmen, die man auf fast alles anwenden kann. Und selbst wenn man es spezieller macht, bei Bildbearbeitung, gibt es die medizinischen Anwendungen, genauso wie das selbstfahrende Auto, genauso wie die automatischen Drohnen, die Waffen werfen. An dieser Stelle ist die Verantwortung in der Grundlagenforschung für diese generischen Fragen schwer festzumachen.

Wenn Sie jetzt fragen, was unsere eigene Verantwortung in dieser ganzen Gemengelage ist, dann habe ich es einfach als Grundlagenforscherin, Mathematikerin, Entwicklerin, die daran forscht. Was sind die Grenzen des maschinelles Lernens? Da stellt sich diese Frage nicht so sehr. Natürlich stellt sich die Frage, je mehr ich in Richtung Anwendung gehe. Darauf habe ich keine besonders gute Antwort, weil das nicht das Thema ist, was mich jeden Tag umtreibt.

Was ich aber auch wichtig finde zu sagen, ist, dass diese Dual-Use-Problematik, die diese Machine-Learning-Verfahren alle haben, nicht nur Machine Learning betrifft, sondern viele Bereiche der Informatik. Wenn ich ein neues Betriebssystem baue, dann kann ich das Militär einsetzen und es kann eine Firma einsetzen. Das heißt, wenn wir grundlegende Technologien zur Verfügung stellen, die für viele Zwecke eingesetzt werden können, ist es oft schwierig, das pauschal zu entscheiden.

Ich sehe die Grundlagenforschung an der Uni an vielen Bereichen an der Stelle, dass es wichtig ist, sie zu betreiben, auch weil man nicht das einzige Land sein will, was sie nicht betreibt. Es muss dann jeder wissen in dem Moment, wo es in die Anwendungen geht, dass man darüber nachdenken muss, wo die Probleme anfangen. Ich denke, man soll sie natürlich schon in die Ferne prognostizieren, was kann alles passieren? Aber die selbstdenkenden Roboter sind im Moment an der Stelle nicht das, was mich umtreibt.

Stefan Remy

Meine erste Reaktion war auch, dass ich nur allgemein darauf antworten kann in dem Sinne, dass man sich als Grundlagenwissenschaftler in den Neurowissenschaften nur – man kann nicht extrapolieren, wohin die Beantwortung der nächsten wissenschaftlichen Frage führen kann. Man ist in der Verantwortung, sich darüber Gedanken zu machen, und man kann sicher Dinge voraussehen, und das sollte auch jeder Wissenschaftler tun. Die Frage der Verantwortung wird wichtig, wenn man die wissenschaftliche Erkenntnis, die wir generieren, mit einer gezielten Anwendung verbindet.

Ich wollte den Begriff der Verantwortung noch mal von einem anderen Standpunkt aus stellen. Das ist fast eine weitere Frage, die ich in den Raum werfe. Wenn wir so Dinge generieren wie Gehirn-Maschine-Schnittstellen, die ins Gedächtnis eingreifen oder die Persönlichkeit verändern, stellt sich natürlich die Frage, wie zukünftige Entscheidungen, die so ein Cyborg-Organismus oder ein Mensch, der durch eine Maschine enhanct wird, welche Verantwortung dieser Organismus oder dieser Mensch für Entscheidungen hat, die er trifft. Das ist ja zum Beispiel auch schon bei Prothetik die Frage oder bei selbstfahrenden Autos. Wenn meine Armprothese jemandem Schaden zufügt, bin ich dafür verantwortlich oder *wer* ist dafür verantwortlich?

Tanja Schultz

Mir geht es ein bisschen so wie meinen beiden Vorrednern. Ich bin Informatikerin und glaube, ich kann relativ gut abschätzen, wo die Reise hingeht in Mensch-Maschine-Schnittstellen, was kognitive Systeme heute schon können. Aber bei

der Frage, wie soll Verantwortung – was sind die rechtlichen, sozialen, gesellschaftlichen Implikationen, da fürchte ich, fragen Sie die Falsche.

Deshalb finde ich es aber auch so wichtig, dass dieser Austausch, den wir jetzt haben, stattfindet, und da bin ich immer sehr dankbar, wenn man sagt, man berät das in einer größeren interdisziplinären Runde, dass die einem berichten: Das haben wir auch getan, wo könnte es technisch hingehen? Aber von der anderen, von der philosophischen, von der rechtlichen Seite wünschen wir uns auch Beistand bei diesen Fragen. Ich hab da ein immer ein bisschen das Gefühl, das ist *out of my league*.

Matthias Bethge

Mir ist gerade ein konkretes Beispiel aus meiner Arbeitsgruppe eingefallen, das vielleicht hilfreich ist zu diskutieren, wo ein Doktorand ein Grundlagenforschungsprojekt gemacht hat, nämlich One-Shot Detection: Man zeigt ein Bild, zum Beispiel meinen Schlüssel, den ich suchen will, und der Algorithmus soll diesen Schlüssel möglichst schnell in der Umgebung oder auf einem Bild finden. Suche ist ein grundlegendes Problem, das wir als Menschen ständig machen müssen.

Er hat dann aber einen Datensatz bekommen, wo auch viele militärische Objekte drauf waren, und hatte das Gefühl, dass dies eine Anwendung ist oder zu Anwendungen führt, die noch wichtiger oder mehr benutzt wird vom Militär als bei Leuten, die nach Vermissten suchen usw. Und weil er das Gefühl hatte, dass diese Frage, Dual Use, bei sich in dem Kontext vielleicht, dass dieses Interesse mehr vom Militär als von anderen da gewesen ist, hat er gesagt, das will er nicht weitermachen.

Ich persönlich denke, als Neurowissenschaftler ist es eine fundamentale Funktion, das menschliche Sehen zu verstehen, und ich würde trotzdem gerne wissen, wie es funktioniert.

Meine Schlussfolgerung daraus war, dass man sich bei der Auswahl, wie man so ein Forschungsproblem aufsetzt, ein Problem, an dem man das studiert, suchen sollte, wo man ganz konkret spezialisiert, wo es einen positiven Nutzen bringt. Also dass man nicht aus Versehen vielleicht der Militärforschung zuspielt, sondern eher guckt, dass man Vermisste finden kann und solche Sachen. Es gibt ja immer auch Bias in der konkreten Technologie, auch in der Grundlagenforschung, die man entwickelt, je nachdem, welche Datensätze etc. man verwendet.

Kerstin Schlögl-Flierl

Vielen Dank für die Beantwortung der Fragen. Es ist spannend: Keiner hat sich die Frage gestellt, ob die KI Verantwortung übernehmen kann. Denn das ist das, was uns so umtreibt: Kann KI Verantwortung übernehmen?

Judith Simon

Frau Schultz hatte noch die Hand oben und Frau Luxburg wollte sich zur letzten Frage äußern. Danach ist Steffen Augsberg der Nächste auf der Liste.

Tanja Schultz

Zu einem etwas anderen Aspekt der Verantwortung: Wir sind ja alle Multiplikatoren und Ausbilder und Ausbilderinnen. Ein wichtiger Grundsatz für mich – und vielleicht erinnert sich Andreas Kruse noch daran, als wir ein Projekt gemacht haben, bei dem es um die Entwicklung kognitiver Systeme für Menschen mit Demenz geht. Da haben wir in unserem Kick-off erst mal alle gemeinsam einen Demenzparcours gemacht, damit wir überhaupt wissen, worüber wir reden. Und nach dem Kick-off-Meeting bin ich mit meinen Mitarbeitern in drei Pflegeeinrichtungen gegangen, um die Menschen, für die wir so etwas bauen,

kennenzulernen und uns zu unterhalten. Das hat einen nachhaltigen Eindruck gemacht auf die Ingenieure und Tekki-Nerds, die bei mir auch im Labor sitzen. Die haben dann gewusst: Dafür baue ich das. Das ist für mich auch ein Stück Verantwortung, aber nicht in dem Sinne, was Sie hören wollten, ob jetzt die KI verantwortlich ist oder wir.

Ich denke, dass wir als Entwicklerinnen und Ausbilder dafür sorgen sollten, dass klar wird, wofür wir diese Systeme entwickeln und für welche Menschen wir diese entwickeln.

Ulrike von Luxburg

Jetzt, nachdem Sie Ihre Frage noch mal neu formuliert haben, habe ich erst verstanden, was Sie meinen, nämlich: Kann KI Verantwortung übernehmen?

Die Frage ist schon so formuliert, dass sich mir die Haare zu Berge stellen, denn die KI ist ja keine Person. Man muss sich in diesen Diskussionen immer klar sein: Da ist keine Person, die schon halb intelligent ist und kann die […] ist die vielleicht noch ein Baby und ist noch nicht intelligent und so oder kann Verantwortung oder ist die schon so weit?

Es ist ganz wichtig, wie wir hier über die KI reden. KI kann keine Verantwortung übernehmen, und auf die Frage, kann man die Ethik der KI einprogrammieren, auch Nein, weil es einfach keine Person ist. Das, was man hier KI nennen will, wird im Moment hervorgebracht durch schlaue Suchverfahren von Algorithmen, die nach Funktionen suchen. Da ist nirgends ein Verantwortungsbegriff.

Ich kann natürlich sagen, es ist die Verantwortung desjenigen, der das Ding trainiert, sinnvolle Daten auszuwählen, sinnvolle Testszenarien zu machen usw., da kann man dann über die technische Umsetzung reden. Aber natürlich kann eine KI nicht Verantwortung übernehmen, weil sie keine Person ist und keine Verantwortung – wie soll das gehen? Da ist kein Verantwortungsbegriff im Hintergrund.

Ich finde das wichtig, in der Art, wie wir darüber diskutieren, weil man immer über KI diskutiert, als wenn da auf der anderen Seite eine Person sitzt, und durch diese implizite Diskussion kommt es ja erst zustande, dass alle Leute denken: Oh, die KI, was kann die denn alles? Deswegen spreche ich selbst eigentlich nie von KI, sondern von maschinellem Lernen, von Algorithmen, um genau dies Missverständnis zu vermeiden.

Steffen Augsberg

Vielen Dank für die spannenden Vorträge. Ich habe zwei kurze Rückfragen, die in eine unterschiedliche Richtung zielen. Das eine nimmt eine kleine Kontroverse auf vom Anfang zwischen Frau von Luxburg und Herrn Bethge. Da ging es um die juridischen Entscheidungshilfen und die Frage, inwieweit man richterliche Entscheidungen auch algorithmisch treffen könnte. Da haben Sie, Frau von Luxburg, gesagt, das können Sie sich *nicht* vorstellen, und Herr Bethge hat angedeutet (wenn ich das richtig verstanden habe), dass er das nicht ganz so drastisch unterstützen würde.

Das könnte ein interessantes Beispiel dafür sein, dass wir bestimmte Schwächen, nämlich menschlicher Art, durchaus zu akzeptieren gelernt haben (der Mensch als Mängelwesen ist ja ein uraltes Phänomen und Beobachtung) und dass wir möglicherweise mit den neuen Schwächen, die in der algorithmenbasierten Entscheidungsfindung zu finden wären, noch nicht in der gleichen Form vertraut sind.

Könnte man an der Stelle sagen, dass das in gewisser Weise ein Komplementärverfahren ist, dass bestimmte Schwächen nicht ausgeglichen

werden können, aber doch irgendwie durch andere ersetzt werden? Und wenn das so ist, ist das dann nicht auch ein Problem, das anknüpft an das, was Herr Bethge gesagt hat, dass sich die Menschen immer an die Technik angepasst haben, dass das vielleicht […] noch nicht erfolgt ist und wir uns insoweit auch mit einer anthropologischen Perspektive beschäftigen müssten? Und die Frage an Sie konkret: Tun Sie das auch?

Das Zweite zielt in die Richtung, wie Frau von Luxburg eben zum Schluss gesagt hat, zu sagen, es kann gar keine Person sein. Würden Sie das tatsächlich so radikal formulieren? Ich meine, dass in den entsprechenden ethischen und philosophischen Debatten gerade dieser graduelle Übergang eine große Rolle spielt und zum Beispiel thematisiert wird, ob man ein Prinzip wie Non-Domination auch auf entsprechende KI-basierte Entitäten anwenden kann.

Und dann ist der entscheidende Punkt, ab wann man das annehmen würde, also die Frage, ob man solche künstlichen Wesen als moralisch relevante Akteure im Sinne von moralischen Objekten ansehen würde. Haben Sie dazu eine Idee?

Judith Simon

Ich nehme an, die Frage geht an alle Referentinnen und Referenten? – Dann gehe ich wieder alphabetisch vor. Die Frage geht zuerst an Herrn Bethge, dann Frau Luxburg, Herrn Remy und Frau Schultz.

Matthias Bethge

Ein Gedanke, der vielleicht an diese Frage anknüpft: Muss das eine Person sein oder nicht? Wir sind aktuell in der Situation, dass sich, wie ich schon angedeutet hatte, das maschinelle Lernen noch meistens darauf beschränkt, Vorhersagen zu machen, Input-Output-mäßig. Irgendjemand hat mal gesagt: Vernunft ist die Fähigkeit, die Welt als Ganzes zu erfassen. Ich glaube, das macht bei diesen ethischen Überlegungen einen Riesenunterschied, also dass man versucht, wirklich alle Aspekte zu berücksichtigen, und man sich im Nachhinein manchmal schuldig fühlt, wenn man irgendeinen Aspekt übersehen hat usw. Also dieses nichts unberücksichtigt zu lassen ist etwas, was in der Technik – wir sind aktuell weit davon entfernt, diesen Grad zu erreichen. Insofern glaube ich, muss man gerade so bei richterlichen Sachen versuchen, wirklich alle Aspekte zu berücksichtigen. Insofern würde ich sagen, sind wir im Moment definitiv in der Situation, KI als Werkzeug zu haben, nicht als Person oder Gegenüber.

Langfristig denke ich, je mehr Maschinen in der Lage sind, komplexe interne Modelle zu haben und verschiedene Handlungsoptionen durchzusimulieren, da macht es wieder Sinn zu fragen, welche von diesen vielen Handlungsoptionen wähle ich aus, nach welchen Kriterien? Ethik ist durchaus etwas, was man erlernt, wo man Überlegungen, bessere und schlechtere Entscheidungen treffen kann. Insofern ist das auch was, wo man den Maschinen bessere und schlechtere Auswahloptionen für Handlungsoptionen beibringen kann.

Judith Simon

Vielen Dank. Frau von Luxburg. Nachher fange ich an zu randomisieren, sonst ist Herr Bethge immer als Erster dran. Aber für diese Runde bleibe ich noch dabei.

Ulrike von Luxburg

Die erste Frage zu den juristischen Entscheidungssystemen und dass ich und Matthias Bethge vielleicht etwas unterschiedlich darauf reagiert haben, hat verschiedene Gründe. Erstens ist das auch eine persönliche Frage: Wie weit geht man und wie weit nicht? Sie haben natürlich recht: Je mehr man lernt, die KI-Systeme zu verstehen und zu kontrollieren, desto mehr kann man ihnen auch

zutrauen in dem Sinne. Der Richter ist sicher nicht fair, das KI-System ist vielleicht auch nicht fair. Vielleicht könnte man irgendwann testen – man hat dann zumindest Maße bezüglich der Fairness, an denen man versuchen könnte, die beiden zu vergleichen.

Es würde für mich auch stark davon abhängen, wer das System eigentlich entwickelt: Macht das eine private Firma? Wo ich keinen blassen Schimmer habe, wer da sitzt und was die Person eigentlich kann und auf welchen Daten die das trainieren. Oder passiert das in irgendeinem öffentlichen Raum? Wobei das jetzt in so einem System schwierig wäre, weil man die Daten aller vergangenen Gefangenen wahrscheinlich nicht einfach veröffentlichen könnte.

Auf meiner Seite müsste ein sehr großes Vertrauen in den Prozess da sein, bevor ich mir vorstellen könnte, solche Anwendungen überhaupt zu durchdenken, was das Testen angeht und die verschiedenen Bias, die man versucht zu korrigieren, und auch: Was ist eigentlich das Ziel und wie wird das System eingesetzt, ganz konkret: Wie findet die Interaktion von dem Richter, der da hoffentlich noch sitzen wird, mit diesem System statt? Das System kann nicht erklären und wird es auch nie richtig können. Wie gehe ich als Richter damit um? usw. Da gibt es sicher Grenzfälle und man kann darüber streiten, welche Fälle dort sind, wo man sagt, das ist gerade noch gerechtfertigt oder wo diese Schwelle überschritten hat.

Ob KI eine Person sein soll oder nicht, dazu kann ich wenig sagen. Für mich ist sie es im Moment nicht. Ich weiß, dass das insbesondere eine juristische Debatte ist oder auch in der ethischen Literatur, ob KI eine Person ist oder nicht, und was dann daran unterschiedlich ist und dass das ein juristischer Kniff sein könnte, einem KI-System einen Personenstatus zuzuweisen, weil man das dann anders behandeln kann.

Ich habe dazu keine richtige Meinung. Für mich ist eine KI, wie ich schon oft gesagt habe, einfach ein Algorithmus, der was tut. Ob das in der Zukunft irgendwann mal anders sein wird, weiß ich nicht. Ich kann aber schwer beurteilen, ob dieser juristische Kniff, KI als eine Person zu betrachten, die Verantwortungsdebatte in der nächsten Zukunft lösen wird. Ich habe das Gefühl, eigentlich nicht. Beim selbstfahrenden Auto kann ich mir nicht vorstellen, dass man das Auto als eigene Person darstellt und damit alle Probleme gelöst hat.

Stefan Remy

Viel Neues kann ich dazu nicht beitragen. Mein Denken ging genau in die Richtung, wie sich Herr Bethge auch dazu geäußert hat. Im Augenblick ist KI ein Werkzeug. Gerade in den Neurowissenschaften kann man sich das vorstellen wie einen Hammer, der uns hilft, Dinge rauszufinden, und im Augenblick sieht in unserem Feld alles aus wie ein Nagel, ja? Und wir wenden es auf alles an. Ich habe mir über diese konkreten Fragen noch zu wenig Gedanken gemacht.

Was den Einsatz im juristischen Kontext eingeht, finde ich Ihren Gedanken sehr wichtig und wollte den noch mal aufgreifen. Es kann sein, dass wir uns der Schwächen einer KI-basierten Entscheidungsfindung als Menschen nicht bewusst genug sind und dass wir das vielleicht als wichtige Aufgabe sehen sollten, klarzumachen, dass Menschen Fehler machen, dass Maschinen Fehler machen. Wenn das bewusster wahrgenommen wird, könnte es mit Sicherheit sehr hilfreich sein.

Tanja Schultz

Ich wollte noch mal den Begriff und die Frage von Herrn Augsberg aufgreifen, ob wir in Anthropologien denken und wenn ja, in welchen. Ich glaube, dass wir wenig direkten Austausch in der

Informatik zur Anthropologie haben, aber die steckt implizit immer drin.

Wir haben den Pol der anthropozentrischen Sicht, in dem Maschinen mit Menschen verglichen werden und abgewogen wird: Wer kann was besser? Und dann gehen wir wieder zurück, sodass sich das auch ergänzen kann.

Dann gibt es die klassischen Technikanthropologien, indem man sagt: Man möchte, dass Maschinen menschliche Fähigkeiten und Fertigkeiten ergänzen und verbessern können, bis hin zum Transhumanismus und zu den symmetrischen Anthropologien, wo es darum geht, dass vielleicht eines Tages Mensch und Maschine verschmelzen zu einem Cyborg. Also implizit steckt es drin, aber explizit wird das Verhältnis zwischen Mensch und Maschine nicht explizit in Anthropologien gedacht, zumindest in der Informatik.

Judith Simon

Herzlichen Dank. Dann bitte ich jetzt Herrn Gethmann um seine Frage.

Carl Friedrich Gethmann

Meine Frage richtet sich an Frau Schultz. Ich wollte darauf eingehen, was sie zum Begriff der Sprache gesagt hat; das fand ich sehr interessant. Ich gehe aber noch auf eine Bemerkung von Herrn Nida-Rümelin ein. Er hat gesagt, wir müssen uns von den Vorstellungen des Behaviorismus lösen. Da stimme ich ihm völlig zu. Das heißt, wir müssen mentale Termini brauchen, um menschliches Verhalten und menschliches Handeln zu verstehen. Aber was ich kritisch ihm entgegenhalte würde, wäre die Frage, ob man die mentalen Termini mit einer so starken ontologisch-realistischen Semantik aufladen muss. Man muss nicht unbedingt in eine so harte Ontologie eintreten, um den Preis dafür zu zahlen, dass man kein Behaviorist ist.

Warum sage ich das? Weil der Begriff des Denkens einer der zentralen Termini ist, der im Alltagsverständnis der Menschen untereinander eine Rolle spielt. Dahinter steht ein gewisser cartesischer Dualismus, dass es (wenn ich die Metapher verwenden darf) so eine Hinterbühne gibt, auf der wir denken, und dann drücken wir uns aus, und auf der Vorderbühne findet dann die Sprache statt. Davon haben Sie sich (und das finde ich sehr sympathisch, was meine eigenen Überlegungen anbetrifft) abgesetzt, Frau Schultz, und haben gesagt: Es gibt auch so etwas wie lautlose Sprache. Das ist eigentlich selbstverständlich, wenn man an die Taubstummensprache denkt, da wird ja durch Gestik Sprache manifestiert, und durch Schrift wird auf noch andere Weise Sprache manifestiert und vielleicht gibt es noch vierte, fünfte Manifestationsformen. Dahinter steht immer die sprachliche Kompetenz, die sich auf unterschiedliche Weise manifestiert, manchmal eben auch lautlos. Und wo sitzt die sprachliche Kompetenz? Wir sind ja keine Dualisten, die an übersinnliche Realitäten glauben; natürlich irgendwie im Gehirn. Also muss es was mit Gehirnstrukturen zu tun haben.

Dann wundert mich aber ein wenig, Frau Schultz, dass Sie etwas aus dem Modell fallen, das Sie selber errichtet haben und dem ich zustimme, wenn Sie wieder sagen, dass man an der Sprache erkennt, wie es sich mit der Kognition verhält. Das erweckt ein wenig wieder den Eindruck Vorderbühne, Hinterbühne, wenn ich das sagen darf. Eigentlich müssten Sie ja sagen: Kognition sei ein bestimmter Aggregatzustand oder eine bestimmte Modifikation von Sprache. Wir nennen das in der Philosophie radikalen Linguistic Turn. Dem müssten Sie eigentlich zusprechen. Das heißt, jetzt noch von einer von der Sprache getrennten Kognition zu sprechen dürfte eigentlich keine Rolle mehr spielen.

Warum ist diese Frage für uns interessant? Weil die Debatte, die soeben geführt wurde (nämlich ob wir KI-Systemen, wie komplex sie auch immer sind, quasi Personenstatus oder irgendwas zusprechen), etwas mit der Interpretation der mentalen Termini wie Denken usw. zu tun hat. Wenn alles Sprache ist, verliert die Sache etwas ihren mythischen Sinn. Denn Sprache ist ja etwas menschlich zutiefst Verständliches, und Sprache verstehen gehört zur normalen menschlichen Lebenswelt dazu und wird nicht künstlich erst erworben, nachdem das Individuum schon fertig ist, sondern das Individuum entsteht ontogenetisch mit dem Erwerb der sprachlichen Kompetenzen. Das ist nicht etwas Zweites. Deswegen ist diese Trennung von Sprache und Kognition ein bisschen den üblichen Redeweisen geschuldet. Aber die Frage ist: Sollten wir damit nicht aufräumen?

Tanja Schultz

Was ich vorhin sagen wollte, ist, dass Sprache einen Einblick gibt über die kognitiven Ressourcen. Mein Beispiel war gewesen, denn wir hatten Demenzen, das war die Frage von Herrn Kruse gewesen und ich hatte gesagt, dass die kognitiven Ressourcen, die bei Demenzen abnehmen, dass man das tatsächlich an der Ausdrucksfähigkeit, an der Flüssigkeit der Sprache, aber auch an den semantischen Inhalten ablesen kann. Denn Sprache zu produzieren kostet kognitive Ressourcen, und wenn diese kognitiven Ressourcen abnehmen, sieht man das auch in der Form und der Art, wie gesprochen wird.

Damit wollte ich eigentlich nur sagen, dass wir in der Lage sind, auf Basis von sprachlichen Äußerungen auch einzuschätzen, wie es um die kognitiven Ressourcen eines Menschen bestellt ist, insbesondere wenn man beispielsweise noch einen zweiten Task hat. Wenn ich Ihnen jetzt neben einem Vortrag auch noch Musik vorspiele oder etwas anderes, auf das man achten muss, dann verändert sich die Ausdrucksfähigkeit. Das können wir nutzen, um aus der sprachlichen Äußerung rauszukriegen: Ist die Person fokussiert? Nehmen die kognitiven Ressourcen über eine Dauer ab? Hat die Person eine schlechte Tagesform? Auch emotionale Stimmungen können wir aus Sprache auslesen. Darauf bezog sich meine Äußerung.

Gethmann

Gut.

Judith Simon

Vielen Dank. Ich würde jetzt als Nächstem Herrn Bormann das Wort geben.

Franz-Josef Bormann

Vielen Dank, ich habe zwei Fragen. Die erste Frage geht an Frau Schultz und an Herrn Bethge. Und die zweite Frage geht an alle vier.

Die erste Frage berührt noch mal dieses Problem Simulation und Mensch-Maschine. Ist das wirklich ein prinzipieller, kategorischer Unterschied? Oder nähert sich das immer mehr an? Herr Bethge hatte am Anfang seines Vortrags gesagt, die Technik sei noch sehr eng gestaltet, sehr eingenischt auf die Erfüllung bestimmter vordefinierter Funktionen ausgelegt. Meine Frage: Gibt es in der Algorithmenentwicklung oder in der Entwicklung dieser selbstlernenden Systeme etwas, wo man sagen kann, das ist zumindest analog dazu, was wir zum Beispiel Intuition nennen?

Ich komme auf die Frage, weil Herr Bethge mit seinem Kreativitätsbegriff da begonnen hat. Für mich ist das noch keine Kunst, wenn man eine van-Gogh-Oberfläche auf Tübingen projiziert, aber das ist jetzt eine andere Frage. Aber da geht es um etwas, was Sie zweimal angesprochen haben, mit dieser Totalität. Es geht irgendwie um das Ganze, es geht um etwas, was nicht mehr so eng eingenischt ist. So haben Sie auch versucht,

den Vernunftbegriff definitorisch zu fassen. Also gibt es in dieser Entwicklung dieser technischen Artefakte so etwas, wo wir sagen könnten, das könnte man vergleichen mit dem, was wir im Bereich des Menschlichen Intuitionen nennen? Oder gibt es etwas, wo wir sagen können, da ist im Blick auf Komplexität so etwas wie: Entscheidungen werden getroffen im Blick auf die Gewichtung verschiedener Parameter oder so etwas? Wo wir dann sagen könnten, das könnte eine Analogie sein zu dem, was wir Abwägung nennen oder so?

Da wäre die Frage: Wo stehen wir da? Viele haben gesagt, bisher ist alles noch sehr begrenzt, aber der Trend ist offen. Da wäre die Frage: Können Sie sich im Blick auf solche Operationen des menschlichen Geistes da im Technischen etwas vorstellen oder auch nicht?

Die zweite Frage wäre, damit wir auch als Ethikrat von Ihnen lernen; wir wollen ja auch einen Input liefern mit unserer Stellungnahme über Mensch-Maschine-Interaktionen, die auch die für Sie zentralen Fragen adressieren. Jetzt habe ich von Ihnen ein bisschen herausgehört: Der Streit um die Frage, ob dieses System eine Person ist oder ob man ihr Akteursstatus zubilligen sollte, ist von Ihren Überlegungen noch relativ weit weg. Vor allem Frau von Luxburg hat das sehr deutlich gemacht.

Aber vielleicht können Sie noch mal artikulieren: Im Blick auf Ihre Community der Grundlagenforscher, der Entwickler usw., was sind ganz zentrale Fragen, die in einer diesbezüglichen Stellungnahme des Deutschen Ethikrates auf jeden Fall vorkommen sollten?

Judith Simon

Vielen Dank. Also die erste Frage an Frau Schultz und dann an Herrn Bethge, und direkt im Anschluss die Frage danach, was der Wunsch vonseiten der Informatikerinnen und Informatiker an ethische Beratung oder Expertise ist, an alle.

Tanja Schultz

Geben Sie mir noch mal ein Stichwort für die erste Frage?

Franz-Josef Bormann

Das Thema Intuition, Abwägung oder so etwas, das sind ja alles Operationen, die wir stark mit dem menschlichen Geist assoziieren. Da haben wir jetzt die Frage: Gibt es da analoge Entwicklungen, oder heute noch nicht, aber vielleicht morgen? Wie das eingeschätzt wird.

Tanja Schultz

Danke. Da gibt es Entwicklungen. Aus einer, in die ich auch involviert bin, kann ich berichten: Wir entwickeln aktuell in einem Sonderforschungsbereich gemeinsam robotische Systeme, die von Menschen lernen. Eine Frage ist, ich nehme mal als Beispiel: Ich decke einen Tisch, in welcher Reihenfolge soll ich Messer und Gabel auflegen, wenn ich ein formales Dinner mache, wie muss das aussehen, versus ein Frühstück für zwei? Da gibt es viele Entscheidungen, die ein Mensch fällt und dann das Geschirr draufpackt.

Die Frage ist: Wie bringt man das einem Roboter bei? Die Antwort, die dort gewählt wird, ist, dass wir den Roboter befähigen, durch eine kognitive Architektur (so wird das genannt) probezuhandeln. Das heißt, der Roboter wird befähigt, diese menschlichen Tätigkeiten nachzuvollziehen und sich selbst in einer Simulation, aber auch in echt verschiedene Möglichkeiten durchspielen zu lassen und dann nach gewissen Gesichtspunkten zu entscheiden, welche dieser vielen Möglichkeiten nehme ich.

Diese Gesichtspunkte können sein: vielleicht den Weg, den man gehen muss, zu minimieren, also möglichst wenig Bewegung zwischen

Kühlschrank und Tisch beispielsweise, oder möglichst effizient im Sinne von beispielsweise Stromverbrauch oder, oder, oder. Es gibt viele unterschiedliche Kriterien, nach denen dieser große Suchraum abgearbeitet wird, und dann entscheidet sich der Roboter für eine Lösung, und die wird dann umgesetzt. Das wichtige Stichwort war: die Fähigkeit mitzugeben, probezuhandeln, Dinge „gedanklich" durchzuspielen und sich dann für eine dieser Möglichkeiten zu entscheiden.

Judith Simon

Herr Bethge, wollen Sie auf die erste Frage noch antworten? Sie können dann direkt anschließen mit der zweiten Frage, und die gebe ich dann an alle weiteren zurück.

Matthias Bethge

Zur Frage der Intuition: Intuition hat für mich viel mit Mustererkennung zu tun, also wo ich gerade nicht nur klare Theorie […] sagen kann, warum mache ich jetzt A und nicht B. Insofern passt das gut zu dem, was aktuell mit maschinellem Lernen gemacht wird, dass ich Indizien sammele und ein gewichtetes Urteil treffe, das sehr gut sein kann. Insofern können wir oft intuitiv Vorhersagen machen, die passen, ohne genau sagen zu können, warum. Das ist eigentlich das, was wir im Moment bei maschinellem Lernen haben.

Was wir nicht so gut haben, ist, dass wir diese Intuition in saubere Modelle übersetzen und daraus strukturiertere Vorhersagen machen. Also die Modellbildung ist eigentlich das, wo es […]

Zu Ihrer Frage, was eine Analogie wäre. Diese Beispiele von Computerspielen, das sind natürlich künstliche Welten, die da geschaffen werden, wo der Agent irgendwelche Aufgaben erfüllen muss. Das ist so ungefähr der Level der Komplexität, wo man in einer Simulation einen Agenten hat, der ein Modell von der Umgebung implizit erlernt haben muss und daraus Handlungsoptionen ableiten musste. Also […] das ist ganz simpel. Starcraft 2 ist dann schon ein bisschen variabler. Ja, das ist, glaube ich, der Stand der Forschung, den wir da haben.

Judith Simon

Vielen Dank. Wollen Sie auch direkt anschließen mit der zweiten Frage, die Herr Bormann an alle gestellt hat, nämlich die Frage, was für Sie aus der Praxis und Forschung eine ethische Beratung oder Hilfestellung wäre, die für Sie relevant wäre? Wir haben ja unterschiedlichste Adressaten, aber natürlich sind die Forscherinnen und Forscher selber – gegebenenfalls ist für uns auch interessant, welche Fragen wir uns stellen sollten, wenn wir eine Stellungnahme schreiben.

Matthias Bethge

Ich tendiere dazu, mir Gedanken darüber zu machen, wie die Zukunft, die wir erfinden, eigentlich aussieht. Wir sind dabei, eine Zukunft zu erfinden, die wird anders – vieles kann ganz anders gedacht werden. In der KI fühlt sich alles sehr virtuell an, man kann die Zahlen auf dem Computer sehr schnell ändern. Gleichzeitig haben wir konkrete physikalische Bedingungen, limitierte Ressourcen, eine wachsende Bevölkerung, im Nachbarkontinent sieht es dramatisch schlecht aus, also wo ich mich frage, […] so ein Lackmustest für unsere Technologie ist –

Eigentlich müsste es uns doch gelingen, mit dieser Technologie, wenn wir irgendwas vorantreiben auf einem Kontinent wie Afrika, in einen Zustand zu bringen, sodass die Leute da bleiben wollen, sag ich mal. Das sind so die Fragen, also wie können wir ein stabiles Gleichgewicht auf der Welt herstellen? Technologie ist unsere Möglichkeit, Einfluss zu nehmen auf die Welt.

Ich tendiere dazu, ein bisschen weniger auf die einzelnen Regeln zu gucken und mehr auf diese großen Fragen: Wie können wir nachhaltig

unseren Globus gestalten? Vielleicht sollte man das irgendwie auch in Beziehung setzen, wenn wir viele kleine Regeln für unseren Alltag hier speziell in Deutschland machen, wo wir von vielen Dingen profitieren, die nicht überall so auf der Welt da sind und die wir vielleicht auch in 50 oder 100 Jahren nicht mehr vorfinden – dass wir uns damit auch auseinandersetzen.

Judith Simon

Vielen Dank. Frau Luxburg, möchten Sie etwas zu der letzten Frage sagen?

Ulrike von Luxburg

Ja, gerne. Was mir wahnsinnig wichtig ist im Moment, wo ich das Gefühl habe, es geht ein bisschen schief, ist die Regulierungsrichtung von KI. Man kann einerseits diese sehr langfristige Perspektive aufmachen, und da bin ich relativ nah an Matthias Bethge und denke, die KI hat ein Potenzial, manche Fragen gut lösen zu können, an denen wir im Moment knabbern. Es gibt viele tolle Anwendungen von KI, also automatische Übersetzungen: Wenn ich in mein Handy sprechen kann und dann kommt in auf irgendeiner Sprache wieder raus, die ich nicht sprechen kann, das finde ich einfach toll.

Aber es gibt auch die ganzen anderen Anwendungen, die im Moment für eine wahnsinnige Beunruhigung bei der Bevölkerung sorgen. Ich habe keine generelle Lösung dafür, und es ist wahrscheinlich genau das Problem, vor dem der Ethikrat dann auch steht, wie man diese Regulierung genau umsetzen könnte.

Was ich wahnsinnig wichtig finde, ist die Frage: Darf jede Firma dieser Welt Gesichtserkennung machen und mit meinen Daten durch die Welt gehen und gucken, wo ich mich aufhalte? Darf das jede Firma? Darf das eine staatliche Organisation? Wenn ja, unter welchen Annahmen oder unter welchen Voraussetzungen darf sie es überhaupt? Und was muss der Algorithmus erfüllen, damit diese Organisation es darf? Darf das Arbeitsamt einfach versuchen, Arbeitslose zu klassifizieren und so einen Testballon starten? Oder sollten da irgendwelche Regelungen stattfinden?

Mir ist nicht klar, wie diese Regulierung im Detail aussehen soll. Ich finde es nur wichtig, dass man da vorankommt und das mit den verschiedenen Akteuren bespricht. Es geht um Firmen, um staatliche Organisationen usw., und es gibt da viele Schattierungen: Sachen, wo es nicht so dramatisch ist, und Sachen, wo es sehr dramatisch sein kann, wenn es nicht reguliert wird.

Das ist das, was mich im Moment umtreibt, weil ich im Gespräch mit Leuten, die ich in Tübingen, in den Diskussionsveranstaltungen treffe, immer auf dieses Unbehagen stoße, und ich kann es verstehen, und wenn ich mich mit meiner Nachbarin gestern unterhalte, erzählt die mir das. Aus jedem Gespräch strömt dieses Unbehagen. Gegen dieses Unbehagen müssen wir was machen, weil sonst die Stimmung kippt gegen die offensichtlich auch tollen Sachen, die die KI kann. Irgendwann sagen die Leute dann: Das wollen wir nicht mehr, schafft das alles ab. Da einen gesunden Zwischenweg zu finden finde ich wichtig, aber auch herausfordernd. Das ist vielleicht eine gute Aufgabe für den Ethikrat, dazu Stellung zu nehmen.

Judith Simon

Danke schön. Herr Remy.

Stefan Remy

Was mich umtreibt, geht spezifischer in Richtung Mensch-Maschine-Interaktion, weil es da relativ klar oder vorstellbar ist, wo die Entwicklung hingeht, Cognitive Enhancement und solche Dinge. Also beim Auslesen von neuraler Aktivität zur Kommunikation sehe ich nicht so viel Regulierungsbedarf. Aber wenn man, Unternehmen, es

sind ja häufig Unternehmen, die jetzt die Technologieentwicklung vorantreiben, wenn man anfängt zuzulassen, dass Hirnaktivität gezielt oder auch nicht gezielt manipuliert wird, dann sollte man sich frühzeitig überlegen, was das für persönliche Konsequenzen für die Individuen hat, die das betrifft, und sich überlegen, ob man dann nicht einschreitet bzw. die Situation eng beobachtet, dass man Eingriffe in fundamentale die Persönlichkeit ausmachende Gedächtnisinhalte nicht einfach geschehen lässt und sich später damit befasst, was diese persönlichkeitsveränderten Menschen eigentlich, ja, wo da die Verantwortung liegt zum Beispiel für Entscheidungen, die die treffen.

Judith Simon

Vielen Dank. Frau Schultz.

Tanja Schultz

Um direkt anzuknüpfen an das, was Herr Remy gesagt hat: Das hatte ich versucht in meinem Vortrag deutlich zu machen, das geschieht alles schon. Es gibt bereits in großem Maßstab Untersuchungen mit Hirnaktivitätsmessungen und Erfassungen des Distraktionsgrades in chinesischen Schulen. Das passiert tagtäglich, und das können wir auch nicht verbieten oder kontrollieren. Wir sollten nicht müde werden, dagegen zu wettern und das bewusst zu machen, aber das ist nicht unter unserer Kontrolle. Deshalb finde ich es wichtig, dass ein Deutscher Ethikrat weiterhin darauf drängt, dass wir allgemeine Grundsätze, die bei uns in einem demokratischen Staat herrschen, einfordern von den Systemen, die hier entwickelt werden, insbesondere von nicht ergebnisorientierter, sondern von offener neutraler Forschung.

Dazu gehört, dass man die Grundrechte respektiert, dass man die Datenhoheit respektiert, das heißt, dass man Individuen befähigt, zu verstehen, was mit ihren Daten passiert, welche Daten aufgezeichnet werden, und wie man vermeidet, dass diese Daten irgendwo hinkommen, wo man sie nicht haben möchte. Transparenz halte ich für ganz wichtig, dass KI-Systeme immer transparent machen, warum sie sich für etwas entscheiden. Das sollte man auch einfordern. Eine gewisse Rechenschaftspflicht und ein Bewusstsein für Missbrauch, das sind Punkte, die wurden schon formuliert.

Gerade gibt es vom IEEE [Institute of Electrical and Electronics Engineers] eine neue Initiative, die heißt Ethically Aligned Design. Da hat man sich auf die Fahnen geschrieben die Priorisierung des menschlichen Wohlbefindens in autonomen intelligenten Systems. Da werden solche Kriterien bereits genannt, und darauf sollte man drängen.

Judith Simon

Vielen Dank. Die nächste Frage kommt von Herrn Demuth.

Hans-Ulrich Demuth

Ich fand es sehr interessant, was Sie, Frau Schultz, und auch was Herr Remy geschildert hat, zur Arbeit mit Alzheimer-Patienten; das kam auch in der Frage von Herrn Kruse raus.

Wenn ich meinen biologischen Hintergrund betrachte und mir angucke, dass bei Alzheimer-Patienten zumindest bei der klinischen Diagnose die Hälfte der Neuronen im Hippocampus alle nicht mehr existent sind, dann ist die Frage: Ist da noch was zu reparieren? Meines Erachtens ist das nicht der Fall. Vielleicht können Sie mich eines Besseren belehren?

Die nächste Frage, die bei mir aufkommt, ist: Wenn wir uns die moderne Entwicklung der Kommunikation zwischen Neuronen anschauen, wo wir jeden Tag was Neues lernen zur Konzentrationsveränderung der Rezeptoren, da kommen Myriaden von chemischen Molekülen ins Spiel,

die im Prinzip diesen Prozess sowohl der Plastizität als auch der Stabilität des Hirns kontrollieren, dann stelle ich mir die Frage: Ist das (und jetzt nehme ich den Begriff, der mehrfach von verschiedenen Seiten kam) simulierbar durch einen Rechner? Ich folge hier komplett Frau von Luxburg: Ich glaube nicht. Bitte versuchen Sie mich zu überzeugen.

Judith Simon

Dann bitte ich alle, wenn Sie möchten, zu überzeugen. Die Frage ging aber erst mal an Frau Schultz, wenn ich das richtig gesehen habe.

Tanja Schultz

Die Antwort zur ersten Frage: Wir haben tatsächlich das Glück, dass wir eine Datenbasis von über tausend Personen haben. An deren Sammlung war auch Andreas Kruse beteiligt. Das sind Menschen, die man über 20 Jahre ihres Lebens begleitet und sprachlich interviewt hat, und es liegen alle diese Interviews vor. Einige dieser Menschen haben Demenzen entwickelt, und wir haben nun die Materialien von 10 Jahren zuvor und 20 Jahren zuvor ausgewertet und haben feststellen können, dass wir bereits 10 Jahre vor dem Ausbruch einer Demenz sprachliche Veränderungen feststellen können.

Das wäre eine mögliche Antwort auf: Was tun wir, wenn schon 50 Prozent der Neuronen niedergegangen sind? Ich glaube, da können wir nichts tun mit Mensch-Maschine-Schnittstellen. Aber wir können vielleicht frühzeitig prädizieren, und wenn man das mit Künstliche-Intelligenz-Systemen machen kann, dann kostet uns das kein Geld und keinen zusätzlichen Aufwand, sondern man könnte das beiläufig testen, natürlich mit Zustimmung aller Beteiligten, und könnte einen Telefonservice einrichten, wo man jeden Tag anrufen kann, wenn man möchte, und ein bisschen mit dem System sprechen kann, und das System sagt dann, da verändert sich was oder ist alles gut.

Hans-Ulrich Demuth

Frau Schultz, das würde ich glatt unterstreichen an dieser Stelle. Bis dahin kann ich mitgehen.

Judith Simon

Vielen Dank, und jetzt die Antwort von Herrn Remy.

Stefan Remy

Ich teile die Ansicht von Frau Schultz zu Demenzen. Das ist das Ziel der Forschung im Augenblick, Biomarker zu identifizieren, die eine Pool-Diagnose ermöglichen, zu einem Zeitpunkt, wo der Neuronenverlust noch nicht so extrem ist und wo man den Krankheitsverlauf noch beeinflussen kann. Da ist das Feld offensichtlich noch nicht weit genug. Ich sehe das aber nicht ganz so pessimistisch, denn die Forschung zeigt, dass neuronale Aktivität auch in Organisationsprinzipien, in Oszillationen, in Regelmäßigkeit stattfindet während des Verhaltens. Diese Regelmäßigkeit kann durch Interaktion, durch Stimulation und Detektion beeinflusst werden, und zumindest in der Grundlagenwissenschaft, in der tierexperimentellen Forschung zeigt sich da ein Potenzial der symptomatischen Therapie, nicht der kausalen Therapie, wo wir sicher noch weit weg sind.

Die zweite Frage: Wir reden von komplizierten Netzwerken, wenn wir vom Gehirn sprechen, und man kann diese Komplexität – jeden Teilaspekt unterbringen, wenn man bis zu den einzelnen Molekülen, den einzelnen Neurotransmittern und den einzelnen Proteinen im Gehirn geht. Ich gebe Ihnen recht, Herr Demuth: Das ist erschlagend, aber deswegen nicht unmöglich, die Simulation durchzuführen. Im Modell würde man ja immer auf die essenziellen Bestandteile, die für die Funktion des Modells wichtig sind, zurückgehen.

Judith Simon

Vielen Dank. Die nächste Frage kommt von Herrn Kruse.

Andreas Kruse

Herr Remy, ich wollte direkt daran anknüpfen, was Sie gesagt haben bzw. was auch Frau Schultz dargelegt hat. Wir dürfen gerade bei der Demenz die symptomatische Therapie nicht unterschätzen, und da ist für mich KI oder die Mensch-Maschine-Interaktion überaus bedeutsam. Wir haben das ja auch im Projekt mit Frau Schultz eindrucksvoll beobachten können, also wenn du ein Gegenüber hast, das nicht nur ganz platte Antworten gibt, sondern das eine kognitive Linie systematisch weiterführt, das können wir ja auch gut.

Von daher finde ich diese Differenzierung zwischen konvergentem Denken und divergentem Denken so wichtig, weil man fragen kann: Kannst du möglicherweise auch noch ganz neue Gedankeninhalte in eine solche Kommunikation einbringen? Das ist mir deswegen so wichtig, weil wir auf diese Art und Weise mit Mensch-Maschine-Interaktion Menschen mit schwersten kognitiven Einbußen erheblich helfen können. Das ist beispielsweise ein Punkt, den wir in meinen Augen im Ethikrat in der Stellungnahme starkmachen sollten, auch an das Symptomatische heranzutreten.

Das fiel mir auch in anderen Beispielen auf, wenn wir bestimmte Läsionen haben, wir kommen an bestimmte motorische Operationen nicht mehr heran oder was Frau Schultz genannt war, beim Locked-in-Symptom, dass wir zwar die Sprache im Kern beherrschen, aber sie nicht produzieren können, oder bei einer Broca-Aphasie. Da hat die Mensch-Maschine-Interaktion ein unglaubliches Potenzial, das auch ein hohes rehabilitatives Potenzial beschreibt, und das sehr gut darzustellen halte ich für bedeutsam.

Es gibt einen Punkt, Herr Remy, da haben Sie in meinen Augen ein sehr wichtiges Thema eröffnet. Wenn wir jetzt in die strukturelle Plastizität hineingehen, Stichwort autobiografisches Gedächtnis oder ich nenne mal ein anderes, was für ältere Menschen problematisch sein kann, auch für Parkinson-Patienten, dass die inhibitorischen Prozesse nicht mehr funktionieren, also ich kann den Ausdruck von Emotionen und Affekten, vielleicht auch von Gedanken nicht mehr kontrollieren, da ist natürlich die Vorstellung, dass wir irgendwann die Möglichkeit haben (in Teilen haben wir sie heute schon), in diesen entsprechenden Arealen zu intervenieren, dass man beispielsweise zu bestimmten Gedächtnisinhalten wieder befähigt wird oder dass der Ausdruck von Emotionen bzw. von Affekten gelingt. Spätestens dann kommen wir zu einem Punkt, den Sie adressiert haben, Herr Remy. Das kann in der Tat auch eine Veränderung der Persönlichkeit bedeuten.

Hier sehen wir, dass Mensch-Maschine-Interaktion, in den Neurosciences bis hinein in die Rehabilitation, einen Grenzgang beschreibt, und wenn der Grenzgang da ist, sind auch immer ethische Fragen angesprochen. Ich sehe die Aufgabe darin, dass wir in einer solchen Stellungnahme die Potenziale, die Gefahren, aber auch diesen Grenzgang beschreiben.

Judith Simon

Die nächste Frage ist von Gräb-Schmidt.

Elisabeth Gräb-Schmidt

Meine Frage geht an Herrn Bethge und wurde teilweise schon von Franz-Josef Bormann gestellt, und zwar die Kreativität, die Sie der künstlichen Intelligenz jetzt schon oder prospektiv zumessen. Zugleich aber haben Sie zu Herrn Nida-Rümelin gesagt, dass Sie das Thema Bewusstsein ausklammern. Klammern Sie das aus, weil Sie es nicht für sinnvoll erachten, es auf die künstliche

Intelligenz anzuwenden, oder klammern Sie es aus, weil Sie denken, das wird mit erfasst über den Begriff der Intelligenz im Sinne dessen, dass es kein Dualismus sein soll, wie es Herr Gethmann gefordert hat?

Wenn Sie von Kreativität oder Erfahrung oder dem Aspekt der Motivation reden, den Herr Kruse ins Spiel gebracht hat, ist das nicht etwas, was durch die Intelligenz im Sinne der Algorithmen nicht abgedeckt werden kann? Und jetzt ohne einen Dualismus zu zementieren, denn auch das, was Herr Gethmann mit Überschuss des Wissens – dieses Überschussphänomen ist ja etwas, was die Reflektivität und den Selbstbezug voraussetzt, um damit umzugehen. Dieses Umgangswissen, halten Sie es auch für möglich, dass die künstliche Intelligenz das in der Selbstentwicklung entwickelt? Oder würden Sie sagen, das ist nicht der Fall? Und daran würde sich meines Erachtens auch bemessen, ob man sagt, das ist eine starke künstliche Intelligenz, und auch von Kreativität sprechen kann.

Matthias Bethge

Ich habe tendenziell schon einen verhaltensorientierten Ansatz. Wenn da ein Agent ist, der Dinge produziert, die ich als kreativ empfinde, die ich vorher nicht kannte, die neu sind, die ich nützlich finde, dann empfinde ich das als kreativ. Das würde ich dann so bezeichnen, unabhängig davon, ob dieser Agent ein Bewusstsein hat oder nicht.

Insgesamt halte ich es bei meinem Nachdenken über diese Fragen der Intelligenzleistungen so, dass ich die Bewusstseinsfrage abspalte und sage: Welche Informationsverarbeitungsprozesse muss ein Agent leisten, um die Informationsverarbeitungsprozesse hervorzubringen, die wir als Menschen hervorbringen? Auch Erfahrungen machen ist etwas, was Informationsverarbeitung erfordert: sich daraus in der Zukunft anders verhalten, neue Handlungs- –

Auch Ethik kann erlernt werden. Das kann alles ohne Bewusstsein erlernt werden. Es gibt keine Notwendigkeit dafür, dass wir bewusst sind. Das wäre meine Einstellung dazu. Ich weiß nicht, ob Ihnen das hilft.

Elisabeth Gräb-Schmidt

Eine Urteilsfindung braucht Differenzerfahrung, und das wäre da nicht inkludiert.

Matthias Bethge

Doch. Wenn ich ein internes Modell von der Welt und von meinen Aktionen in der Welt habe, kann ich die auch bewerten, wie ich das auch bei einem Schachspiel machen kann, dass ich meine Erfahrungen, die ich in der Vergangenheit gemacht habe, auf die Zukunft anwende und sage: So möchte ich mich verhalten oder so möchte ich mich nicht verhalten. Das ist natürlich sehr abstrahiert und vereinfacht, aber da sehe ich keinen Unterschied.

Judith Simon

Mit Blick auf die Uhr würde ich jetzt bitten, dass die drei letzten Fragenden ihre Fragen hintereinander stellen und dann alle Redner und Rednerinnen sich aussuchen, auf welche der Fragen sie antworten wollen, sofern nicht durch die Frage schon vorgegeben ist, wer sie beantworten soll. In der Reihenfolge wären das zunächst Stephan Kruip, Andreas Lob-Hüdepohl und Armin Grunwald.

Stephan Kruip

Ich bin Diplom-Physiker und habe Anfang der 80er Jahre einem Computer Mühle beigebracht, ohne dass der Computer je begriffen hat, was Mühle ist. Und das ist auch schon mein Stichwort. Bevor man sich überlegt, ob ein Computer Emotionen lernen kann oder Gerichtsentscheidungen

durchführen kann, würde ich anregen, darüber nachzudenken, ob eine künstliche Intelligenz etwas begreifen kann.

Ich bin von Beruf Patentführer und wollte einfach die Erfahrung teilen. Im Jahr 2000 habe ich angefangen, da war die große Sorge, dass alle Patentprüfer entlassen werden. Es wurden Softwarepakete vorgestellt, wo dann Delegationen von anderen Ländern gesagt haben: Ja, und wann entlassen Sie alle Patentprüfer, wenn das so perfekt funktioniert?

Es gibt faszinierende Fortschritte in der KI und wir bekommen heute tolle Recherche-Ergebnisse. Trotzdem ist es so, wenn man die Recherche-Ergebnisse einem Menschen zuordnen würde, würde man sagen: Der hat es nicht begriffen. Der hat die Erfindung nicht begriffen, sondern nur Zufallstreffer gelandet.

Deswegen meine Frage: Kann künstliche Intelligenz nach heutigem Stand überhaupt irgendetwas begreifen, einen Sachverhalt begreifen?

Andreas Lob-Hüdepohl

Frau Schultz, Sie haben an einer Stelle geschildert, dass Sie nach einem Kick-off-Treffen mit Ihrem Forschungsteam demenziell Erkrankte besucht haben und in die Lebenswelt eingetaucht sind. Deshalb meine Frage: Können Sie sich vorstellen, dass nicht nur Sie die Lebenswelt, für die Sie arbeiten, kennenlernen, sondern im Sinne einer partizipativen Forschung, wie wir das beispielsweise im Kontext unserer Stellungnahme der Robotik-Ethik formuliert haben, dass die Adressatinnen und Adressaten Ihrer Forschung einbezogen werden in die Entwicklung von Fragestellungen? Oder ist das angesichts der Komplexität der Thematik, mit der Sie in Ihren Forschungen zu tun haben, eine Illusion?

Armin Grunwald

Es ist nur eine Anregung. Es kam mehrfach die Sorge, dass durch die aufgeheizte Stimmung, überhöhte Erwartungen an KI und überspitzte Risikobefürchtungen mit der hohen damit verbundenen Unsicherheit auf beiden Seiten die Gefahr besteht, dass wir aufgrund einer öffentlichen, einer politischen Stimmung Chancen vielleicht nicht nutzen können, die da wären in einem anderen Vorgehen.

Das hat mich erinnert an die Stimmung zur Nanotechnologie ungefähr 2003, 2004. Damals hat man zum Beispiel in Österreich, aber auch in Deutschland landesweite Dialogprozesse gestartet in der Überzeugung, man müsse die Sorgen der Menschen ernst nehmen, man müsse mit ihnen ins Gespräch kommen, es gehe nicht, wie in der früheren Kernkraftdiskussion von oben zu sagen, wir haben alles unter Kontrolle und wer anderes behauptet, hat keine Ahnung.

Durch diese offene Dialogatmosphäre ist damals viel Vertrauen gewonnen worden. Vielleicht lohnt es sich, trotz der 15 Jahre, die seitdem vergangen sind, mal in diesen alten Foren zu schauen, ob man da etwas lernen kann.

Judith Simon

Vielen Dank. Damit wären wir am Ende der Zeit, aber jetzt haben Sie noch 30 Sekunden, um sich eine Frage auszusuchen, die Sie gern beantworten möchten. Damit es schneller geht, sage ich jetzt die Reihenfolge: Frau Schultz, Frau Luxburg, Herr Remy und Herr Bethge.

Tanja Schultz

Ich nehme die Frage Nummer 2 zum partizipatorischen Design. Vielen Dank für die Frage, denn das habe ich unterschlagen, aber das machen wir natürlich. Für das I-CARE-Projekt beispielsweise erinnere ich mich an einen tollen Workshop, den wir bei Herrn Kruse in Heidelberg hatten. Da

haben wir viele ältere Menschen eingeladen, um rauszukriegen, ob die mit den Technologien, die wir verwenden wollen, zurechtkommen. Wir haben auch explizit nachgefragt, wie wir so etwas gestalten könnten, um es zugänglicher zu machen. Also: Ja, das tun wir und ich halte es auch für einen sehr wichtigen Aspekt unserer Forschung.

Ulrike von Luxburg

Die Dialogprozesse, die am Schluss angesprochen worden sind, sind wahnsinnig wichtig. Wir machen in Tübingen viel in dieser Richtung. Ein Unterschied vielleicht zur Nanotechnologie ist, dass in diesem Dialogprozess nur bestimmte Player auftauchen und speziell Firmen, denen man alles Mögliche unterstellt, sich da meistens sehr stark zurückhalten.

Ich glaube, man kann bis zu einem gewissen Grad durch Dialogprozesse Dinge erreichen. Die sind sehr wichtig und wir arbeiten daran, aber sie lösen nicht das Gesamtproblem.

Die andere Frage, zu der ich kurz etwas sagen wollte, war, ob KI begreifen kann. Ich hatte in meinem Vortrag schon gesagt: Erklären kann sie ein bisschen, aber nicht so gut, und um Dinge begreifen zu können und uns auch begreifbar zu machen (das ist ja auch immer die Frage: Wie wird das Wissen transportiert, das vielleicht implizit in diesem Algorithmus steckt?), bräuchten wir Erklärungen und eigentlich auch Modelle. Das sind Sachen, die die KI im Moment relativ schlecht kann oder auch nicht verwendet wird. In vielen Algorithmen werden keine Modelle gebaut und dann ist es sehr schwierig, da auf irgendwelche Erklärungen zu kommen. Es kann natürlich sein, dass sich das in Zukunft mal ändert. Aber im Moment ist das schwierig.

Stefan Remy

Kurz zum Gedanken der Partizipation: Das halte ich auch für total wichtig. Das wird auch gemacht und sollte immer gemacht werden, wenn wir Entwicklungen in bestimmte Richtungen treiben: in enger Interaktion zu sein mit den Menschen, die von der Entwicklung profitieren sollen.

Zu den Begriffen Offenheit und Vertrauen, die ich in dem Zusammenhang für total wichtig halte, hatten wir vorhin den Gedanken Open Science; den hat Frau Schultz aufgebracht. Das sehe ich auch so. Es ist ganz wichtig, dass wissenschaftliche Daten und Erkenntnisse allgemein zugänglich sind für jeden und allgemein zugänglich gemacht werden. Und ich möchte diejenigen, die Sie angesprochen haben, nämlich Firmen, Kommerzorientierung – darauf sollte man ein Auge werfen, damit der Gedanke der offenen Wissenschaft und des Vertrauens gegenüber der Wissenschaft nicht degradiert in der Zukunft.

Matthias Bethge

Ich habe mir auch das mit dem Begreifen rausgesucht. Feynman kann sehr anschaulich beschreiben, was es bedeutet, zu verstehen und zu begreifen. Er sagt, das ist das Zurückführen auf etwas, was uns vertrauter ist, und wenn man so etwas wie Wärme verstehen will, dann hilft die Analogie, Bälle, die ein Artist hin und her […], als Modell dafür zu verstehen.

Mit solchen Arten von Modellen arbeiten wir bisher im maschinellen Lernen noch nicht. Es gibt erste Grundlagenforschung in die Richtung, wo so etwas versucht wird. Wir haben das auch in unserem Exzellenzcluster, den Frau Luxburg führt. Wenn so was funktioniert, haben wir noch mal eine ganz andere Diskussion. Aber ich glaube schon, dass es prinzipiell möglich ist.

Judith Simon

Vielen Dank. Damit bedanke ich mich bei allen ganz herzlich und gebe ab an Alena Buyx für das Schlusswort.

Schlusswort

Alena Buyx · Vorsitzende des Deutschen Ethikrates

Herzlichen Dank, Judith. Ich verspreche, kein Schlusswort zu formulieren. Das ist immer schwierig nach solchen Anhörungen und Veranstaltungen. Heute ist es völlig unmöglich, nach dieser reichhaltigen Diskussion. Deswegen mache ich es ganz anders und wende mich explizit an unsere Referentinnen und Referenten:

Wir haben Sie eingeladen, damit Sie uns den Stand der Dinge darstellen und wohin die Reise geht. Das haben Sie wunderbar getan und dafür sind wir Ihnen dankbar. Aber dann haben Sie sich auf unser Terrain begeben, und wir haben Sie da gnadenlos herausgefordert, wenn ich das mal so sagen darf, und wir haben in der Diskussion Fragen erörtert, die zum Teil in bestimmten Disziplinen nicht nur mit Blick auf die künstliche Intelligenz seit Jahrzehnten, teils seit Jahrhunderten oder gar Jahrtausenden besprochen werden: Was verstehen wir unter Intelligenz? Wer sind moralische Agenten und Akteure, Personalität, Verantwortung usw.

Das ist das, was wir brauchen. Denn wenn wir diese gerade zum Schluss angesprochenen praktischen Herausforderungen und Umsetzungsfragen für uns alle und für die Gesellschaft lösen oder zumindest adressieren wollen, brauchen wir diese Art von Austausch. Das haben einige von Ihnen auch gesagt. Ich kann nur sagen (und ich glaube, da spreche ich im Namen aller Kolleginnen und Kollegen vom Ethikrat): Das war eine echte Sternstunde des interdisziplinären Austausches. Wir danken Ihnen sehr, dass Sie mit uns auf dieses Parkett gekommen sind. Wir haben enorm profitiert. Deswegen gehen Sie jetzt in Ihren Rest des Tages. Ich möchte persönlich allen anderen Beteiligten danken, stellvertretend Judith Simon, auch eine Sternstunde der Moderation. Vielen herzlichen Dank dafür, wie du das gemacht hast.

Ich danke der Technik und den Schriftdolmetscherinnen, ich danke all denjenigen aus der Geschäftsstelle und der AG, die im Hintergrund vorbereitet haben, die Technik so reibungslos haben laufen lassen, und ich danke Ihnen, die Sie uns zugeschaut haben, die Sie sich interessiert haben und hoffentlich etwas mitgenommen haben. Es gibt ausführliche Dokumentationen dieser Veranstaltung. All das, was wir in der Zukunft noch machen werden, finden Sie auf unserer Webseite. Wir informieren Sie gern. Wir haben einen Verteiler, also schreiben Sie uns.

Dann schließe ich mit zwei Hinweisen auf weitere Online-Veranstaltungen: Wir haben am 24. März das Forum Bioethik zu dem wichtigen Thema „Triage – Priorisierung intensivmedizinischer Ressourcen unter Pandemiebedingungen“, und wir haben, worauf wir uns alle schon sehr freuen, am 23. Juni unsere Jahrestagung, „Wohl bekomms! Dimensionen der Ernährungsverantwortung“.

Es wird nicht langweilig. Bleiben Sie uns treu, kommen Sie wieder und wir wünschen Ihnen und uns allen einen schönen Tag. Vielen Dank.