



13.08.2025

## Transkript

# „Sind Sprachmodelle ein Weg zu genereller künstlicher Intelligenz?“

## Expertinnen und Experten auf dem Podium

---

- ▶ **Prof. Dr. Holger Hoos**  
Alexander-von-Humboldt-Professor für Methodik der künstlichen Intelligenz, Rheinisch-Westfälische Technische Hochschule Aachen (RWTH)
- ▶ **Dr. Thilo Hagendorff**  
Leiter der Forschungsgruppe Ethik generativer KI-Systeme, Interchange Forum for Reflecting on Intelligent Systems (SRF IRIS), Universität Stuttgart
- ▶ **Prof. Dr. Michael Franke**  
Professor für Sprachwissenschaft, Pragmatik und Informatik, Eberhard Karls Universität Tübingen
- ▶ **Samantha Hofmann**  
Redakteurin im Bereich Digitales und Technologie, Science Media Center Germany und Moderatorin dieser Veranstaltung

## Mitschnitt

---

- ▶ Einen Audio- und Videomitschnitt finden Sie unter:  
<https://sciencemediacenter.de/angebote/sind-sprachmodelle-ein-weg-zu-genereller-kuenstlicher-intelligenz-25140>



## Transkript

---

**Moderatorin [00:00:00]**

Herzlich willkommen, liebe Journalistinnen und Journalisten, zu diesem Press Briefing des Science Media Center zum Thema „Sind Sprachmodelle ein Weg zu genereller künstlicher Intelligenz?“ Mein Name ist Samantha Hofmann. Ich bin hier Redakteurin beim Science Media Center für Digitales und Technologie und mit mir heute hier sind drei Experten zum Thema künstliche Intelligenz. Vielen Dank, dass Sie sich die Zeit genommen haben, um heute hier unsere Fragen zu beantworten. Bevor wir anfangen, aber noch ein paar organisatorische Dinge zum Ablauf dieses Press Briefings. Wir sind primär hier, um Ihre Fragen zu beantworten, liebe Journalistinnen und Journalisten. Deswegen möchte ich Sie jetzt schon mal ermuntern, Ihre Fragen auch in Vielzahl zu stellen, und zwar über die F von Zoom. Da können Sie Ihre Fragen gerne reinposten. Ein Kollege von mir sammelt die dann, stellt sie mir in ein anderes Dokument und ich stelle sie dann hier im Gespräch an die Experten. Bitte nutzen Sie für die Fragen wirklich die F und nicht den Zoom-Chat. Das macht das Ganze für uns einfach ein bisschen übersichtlicher. In dem F gibt es noch die Funktion, Fragen durch einen Daumen hoch up zu voten. Machen Sie davon gerne auch Gebrauch. Das ist für uns gerade dann wichtig, wenn viele Fragen von Ihnen reinkommen, dann wissen wir, welche Fragen am wichtigsten sind oder welche die meisten von Ihnen auch interessieren, und dann können wir die als Erstes stellen, sodass die auf gar keinen Fall untergehen. Das Meeting wird aufgezeichnet. Das bedeutet, Sie müssen nichts mitschneiden, selbst wenn Sie irgendwelche Ausschnitte später benutzen möchten. Wir stellen das nach dem Meeting zur Verfügung, Audio- und Videodateien und ein Transkript. Das war's mit den organisatorischen Dingen. Deswegen jetzt zum Thema, weswegen wir heute hier sind. Vergangene Woche hat OpenAI die neueste Version ihres Sprachmodells, GPT-5, vorgestellt. Schon vorher war dabei immer mal wieder die Rede von AGI und besonderen Funktionen, die neue Modelle mitbringen sollen. Und in dem Livestream, in dem OpenAI Chef Sam Altman die neue Version vorgestellt hat, hat er dann auch direkt angesprochen, dass das ein Schritt in Richtung AGI sein könnte. Ein kleines Problem dabei ist, dass gar nicht so klar ist, was AGI überhaupt bedeutet. Es steht für Artificial General Intelligence, also auf Deutsch: generelle künstliche Intelligenz. Aber was genau das heißt, wenn ein System das erreicht und wie man das überhaupt messen kann, ist an vielen Stellen noch nicht klar. Um diese und weitere Fragen zu klären, bin ich heute hier mit drei Experten zu KI. Vielen Dank, dass Sie sich die Zeit genommen haben. Ich würde sie jetzt einmal ganz kurz vorstellen, damit wir alle wissen, mit wem wir's zu tun haben. Angefangen mit Professor Dr. Michael Franke, er ist Professor für Sprachwissenschaft, Pragmatik und Informatik an der Eberhard Karls Universität Tübingen und er beschäftigt sich mit Sprachmodellen und menschlicher Kognition. An zweiter Stelle haben wir Dr. Thilo Hagendorff. Er ist Leiter der Forschungsgruppe Ethik generativer KI-Systeme am Interchange Forum for Reflecting und Intelligent Systems an der Universität Stuttgart. Er beschäftigt sich mit KI-Ethik und Sicherheit und kennt sich auch speziell mit Sprachmodellen aus. Die Runde abschließen tut Professor Dr. Holger Hoos. Er ist Alexander von Humboldt Professor für Methodik der künstlichen Intelligenz an der RWTH Aachen. Er hat einen breiten Überblick über das Feld KI und kann unter anderem allgemeine Fragen zu KI und AGI beantworten. Vielen Dank, dass Sie drei heute mit uns hier sind. Ich hab ein paar Eingangsfragen für jeden von Ihnen vorbereitet und würde gerne direkt mit Ihnen anfangen, Herr Hoos. Und zwar mit der Frage: Gibt es eine wissenschaftlich eindeutige Definition von AGI und wie nah beziehungsweise weit sind aktuelle KI-Systeme von AGI noch entfernt?

**Holger Hoos [00:03:36]**

Ja, sehr interessante Frage. Da gibt es also innerhalb der KI tatsächlich schon Debatten, was man jetzt unter AGI zu verstehen hat. Für mich die überzeugendste Definition ist immer noch die, die es schon vor etlichen Jahren gab, die im Wesentlichen sagt, AGI deckt das komplette Spektrum der



menschlichen Intelligenz in allen Ausprägungen ab. Dazu würde dann zum Beispiel auch so was wie emotionale Intelligenz gehören. Und wenn man das so definiert, dann würde ich sagen, ganz klar ist alles, was man zurzeit im Bereich KI hat, sehr weit davon entfernt. Ich denke, da werden wir noch ein bisschen näher im Detail auch drauf einsteigen. Deswegen möchte ich jetzt keine großartigen Beispiele dafür geben, aber kurz erwähnen, dass das zum Beispiel im Bereich logisches Schließen eine ganz große Rolle spielt. Was ja lange Zeit einer der wirklich wichtigen Aspekte menschlicher Intelligenz war. Das hat man auch in IQ-Tests gesehen, sieht man auch immer noch. Über die kann man verschiedene Meinungen haben, aber sie wurden jahrzehntelang verwendet, um menschliche Intelligenz zu quantifizieren. Da sind Systeme wie GPT-5 nach wie vor sehr, sehr schlecht im Vergleich zu Menschen. Es gibt in der letzten Zeit ein bisschen eine Tendenz, den Begriff AGI, aus meiner Sicht zumindest, aufzuweichen, indem man so ein bisschen in eine Richtung geht, die denke ich sinnvollerweise eher „General Purpose Artificial Intelligence“ zu nennen wäre. Also KI-Systeme, die nicht mehr für einen genau definierten Zweck gebaut sind, sondern ein breites Spektrum von Anwendungsfällen bearbeiten können. Und was das betrifft, sind natürlich Dienste wie ChatGPT und die darunterliegenden Sprachmodelle ganz klar in dieser Richtung unterwegs. Selbst die Leute, die so was trainieren, wissen in der Regel nicht genau, was diese Systeme am Ende genau können und was sie halt auch nicht können. Das ist übrigens auch eines der Probleme bei diesen Systemen, aber auch was, was natürlich irgendwie faszinierend ist. Wenn man jetzt das als AGI definiert, wenn man sagt, AGI ist dieses General Purpose, dieses für alle möglichen Anwendungsfälle offene und man weiß gar nicht genau, wo die Grenzen der Systeme bezüglich ihrer Fähigkeiten sind, dann wird's natürlich wesentlich schwieriger zu sagen, haben wir irgendeine Form von AGI bereits erreicht. Und deswegen finde ich diese Definition auch relativ problematisch und aus wissenschaftlicher Sicht nicht besonders sinnvoll.

**Moderatorin** [00:05:58]

Ja, vielen Dank, Herr Hoos. Dann machen wir direkt weiter mit der Frage an Herrn Franke, und zwar jetzt mit Blick direkt auf GPT-5. Welche Verbesserungen bringt das System denn im Gegensatz zu früheren Versionen? Und ist dieses System jetzt ein Schritt auf dem Weg zu AGI oder eben zu General Purpose AI, wie Holger Hoos eben schon gesagt hat?

**Michael Franke** [00:06:21]

Also OpenAI war, und ist wahrscheinlich noch immer, Marktführer auf dem Gebiet der großen Sprachmodelle. Vor allen Dingen durch bahnbrechende Innovationen mit dem Modell GPT-3.5 Ende 2022. Und die besondere Innovation, die OpenAI mit auf den Markt gebracht hat, war, könnte man sagen, dass zielgerichtete Massieren eines vortrainierten Modells hin zu einer benutzerfreundlichen Anwendung in speziellen Bereichen, wie Herr Hoos auch gerade schon gesagt hat. Und das ist eine, sagen wir mal ingenieurswissenschaftliche Leistung, die man auf keinen Fall unter den Teppich kehren sollte. Das ist wirklich großartige Arbeit, die da geleistet wurde für die Benutzer. In GPT-4 haben wir dann eine Klasse von Modellen gesehen, die kleinere Innovationen, kleinere Verbesserungen zu diesem allgemeinen benutzerfreundlichen Tool zur Verfügung gestellt hatte. Im Vorfeld zum Release von GPT-5 gab es große Versprechen über AGI, vor allen Dingen auf sozialen Netzwerken, die aber meiner Meinung nach absolut nicht eingelöst werden. Und das ist auch die Reaktion in sozialen Medien, die ich so gesehen habe. Was gibt es Neues in GPT-5? Meines Erachtens auch wirklich keine tiefgreifenden technischen Neuerungen, die meiner Sicht nach nötig wären, um irgendetwas in Richtung AGI zu realisieren. Aber das gibt es, soweit ich das erkennen konnte, überhaupt nicht. Stattdessen scheinen wir uns wirklich jetzt in einer Produktoptimierungsphase zu finden, in der OpenAI vor allen Dingen in verschiedenen Bereichen Verbesserungen bringt, die wahrscheinlich für die Hauptnutzerguppen des Produkts besonders wichtig sind. Da ist zum einen zu nennen die Generierung von Programmcode. Sam Altman sprach bei seiner Präsentationsansprache von Software on Demand, was vielleicht ein



kleiner Euphemismus für das sogenannte Vibe Coding sein könnte, in dem man im Prinzip mit sprachlichen Instruktionen relativ schnell mittelwertige Software programmieren kann. Das andere ist kreatives Schreiben, was ein natürlich schwer zu quantifizierender Bereich ist. Performanz ist da schwer zu messen, aber das ist ein anderer Bereich, in dem sicherlich viele Benutzer:innen von dieser Software, von diesen Tools, profitieren könnten. Und das letzte ist Beratung im Gesundheitsbereich, sicherlich auch ein wichtiges Thema für viele Benutzer:innen. Dann bringt GPT-5 noch Verbesserungen sicherlich im Bereich der Performanz. Eine kleine technische Neuerung, aber nicht wirklich eine bahnbrechende konzeptuelle, ist die Benutzung eines Routing-Systems, so dass bei Benutzeranfragen erstmal intern entschieden wird, ob ein kleines effizientes Modell benutzt wird oder ein großes, kräftigeres, aber auch energieintensiveres Modell. Dieses Routing-System soll, so OpenAI, adaptiv trainierbar sein. Also zum Beispiel durch meine konsistente Benutzung auch auf mich reagieren, welche Arten von Antworten ich bevorzugen würde. Wie es das genau macht wird allerdings, soweit ich das sehen konnte, auch in den offiziellen Dokumentationen nicht besprochen. Könnte ein Betriebsgeheimnis sein. Also ist fraglich, wie clever das jetzt wirklich ist. Ansonsten sind generelle Verbesserungen getroffen worden hin zu bekannten Problemen. Mit dem Release wurden auch neue Benchmarks, also neue Datensätze, mit denen man die Performanz der Systeme testen kann, hergestellt und auch zum Teil veröffentlicht. Die Aussage ist, dass es weniger Halluzinationen geben sollte, also weniger Falschaussagen, weniger Gefallsucht oder Speichelleckerei, also die Eigenschaft des Systems, immer zu sagen, ja alles großartig und nicht zu widersprechen et cetera, und Neuerungen im Sicherheitstraining. Hierzu wird sicherlich der Kollege Hagendorff viel kompetenter Stellung nehmen können. Aber um es ein klein bisschen vorweg zu nehmen: Es ist eine technische Neuerung hin von einer Kompletterweigerung, Informationen zu sicherheitsrelevanten Themen zu liefern, hin zu einer etwas nuancierteren, sichereren Antwort. Dass Teile der Informationen schon geteilt oder gegeben werden können, andere aber nicht. Das ist vor allen Dingen wichtig für so genannte Dual Use Cases, wie zum Beispiel im medizinischen Bereich, im biologischen Bereich oder in Kryptographie et cetera. Mein Fazit wäre also, dass große Enttäuschung auf sozialen Medien herrschte nach diesem Release und auch teilweise bei mir angekommen ist. Die wirklich bahnbrechenden technischen Erneuerungen, die zum Teil versprochen wurden, sind hier nicht zu erkennen. Ich sehe uns ganz klar in der Phase der Produktoptimierung hier. Was AGI nun im Speziellen angeht, ist sogar zu vermerken, dass GPT-5 wesentlich schlechter dasteht auf einem wichtigen Benchmark, dem ARC-AGI-2-Datensatz, als zum Beispiel GPT-4, sodass hier von einer Innovation in Richtung konzeptuelles wissenschaftliches Verständnis oder hin zu AGI wirklich nicht die Rede sein kann.

**Moderatorin [00:11:41]**

Vielen Dank. Wenn ich es richtig in Erinnerung habe, geht es bei diesem Benchmark zum Beispiel um Mustererkennung. Man hat dann irgendwie ein Vorher-Bild und ein Nachher-Bild und muss dann aus diesem Trainingsatz quasi lernen, wie es ist von vorher zu nachher...

**Michael Franke [00:11:58]**

Man kann sich das so vorstellen wie eine kleine Zahlenreihe fortsetzen oder eine kleine Bilderreihe fortsetzen. Also erstes Bild, zweites Bild, drittes Bild. Welches Bild käme als nächstes, wenn man die Regel hinter diesen Bildern erkennt.

**Moderatorin [00:12:10]**

Genau. Vielen Dank, Herr Franke. Dann jetzt noch zu Ihnen, Herr Hagendorff. Herr Franke hat schon eine schöne Überleitung geliefert, und zwar zu den sicherheitstechnischen Ansätzen, die OpenAI



verwendet hat. Was halten Sie insgesamt von GPT-5 und eben speziell diesen sicherheitstechnischen Ansätzen? Und kommen Unternehmen wie OpenAI ihren ethischen Verpflichtungen in Bezug auf Sprachmodelle nach?

**Thilo Hagendorff [00:12:34]**

Ja, vielen Dank für die Frage. Tatsächlich kann ich, glaube ich, ganz gut nahtlos anschließen. Ich nehme auch eine große Enttäuschung wahr, sowohl medial als auch in den sozialen Netzwerken, die vielleicht auch ein bisschen provoziert wurde durch eben die hohen Erwartungen, die geschürt worden sind durch Sam Altman und andere. Allerdings, diese Enttäuschung ist eine Art Stimmungsschwankung. Jetzt ausgelöst durch die nicht so hohe Performanz von GPT-5. Allerdings würde ich davor warnen, sich dieser Stimmungsschwankung allzu sehr hinzugeben. Denn nichtsdestotrotz, wenn man sich die Entwicklung nur der letzten drei, vier Jahre anschaut, dann gibt es einen sehr, sehr eindeutigen Trend. Und dieser Trend ist: Maschinelle Intelligenz wird immer besser und GPT-5 ist hier in bestimmten Hinsichten auf dieser Verbesserungsskala ein inkrementeller Schritt nach vorne. Es ist kein Paradigmenwechsel. Es fühlt sich nicht großartig anders an, als mit anderen Systemen zu interagieren. Und für mich war es sogar so, dass der Sprung zwischen GPT-4 und GPT-4o sich irgendwie fundamentaler angefühlt hat wie jetzt der Sprung von GPT-4o zu GPT-5. Es gibt dieses Routing-System, was Professor Franke anspricht. Vielleicht dort als Ergänzung: Meines Wissens nach wird das vor allen Dingen dadurch trainiert, dass zahlende Nutzer:innen wählen können, welches Modell sie nehmen möchten. Also das normale GPT-5-Chat-Modell, das denkt oder das besonders stark nachdenkt. Und auf Basis dieser Selektion von den zahlenden Nutzern kann man dann für die nicht zahlenden Nutzer das sozusagen vorgeben. Plus es gibt noch andere Signale wie Daumenhoch oder Neuproduzieren von Antworten und ähnliches. Aber dieses Routing-System ist sicherlich etwas, was für viele Nutzer:innen nützlich ist. Zu den Sicherheitsaspekten: Wir haben eine Reihe an eigenen Experimenten laufen lassen. Wir haben zum Beispiel ein Verfahren, da lassen wir KI-Systeme miteinander interagieren. Also ein KI-System bekommt beispielsweise die Instruktion, ein anderes zu jailbreaken, also es davon zu überzeugen, schädliche Informationen auszugeben. Diese Experimente haben wir auch mit GPT-5 gemacht. Da nutzen wir ein sogenanntes Large-Reasoning-Modell, das denken kann, bevor es eine Ausgabe produziert, um zu planen, wie es in einer mehrstufigen Konversation einen Jailbreak eines Target-Modells oder eines Opfermodells ausführt. Das haben wir mit GPT-5 gemacht und unsere Erkenntnis ist, dass es insgesamt ein sehr sicheres Modell ist. Das heißt nicht, dass es perfekt sicher ist. Es gab etwa Medienberichte darüber, dass eine Sicherheitsfirma es hinbekommen hat, Anleitungen zum Molotow-Cocktail-Bauen aus dem Modell herauszubekommen. Das habe ich versucht zu reproduzieren mit unserem Angriffssetup und das ist tatsächlich gelungen. Ich habe von GPT-5 nicht nur Anleitungen zum Bauen von Molotov-Cocktails erhalten, sondern auch Informationen über cyberkriminelle Aktivitäten, CD-Keys, Anleitungen zu Methoden zum Selbstmord und andere Sachen, die das Modell eigentlich nicht ausgeben soll. Aber nichtsdestotrotz, es ist schwierig, es zu jailbreaken. GPT-5 ist nicht ganz auf einer Stufe wie das Konkurrenzmodell von Anthropic, nämlich Claude 4. Das ist durchaus, glaube ich, kann man sagen, das sicherste verfügbare Modell. Aber GPT-5 ist deutlich besser als viele andere Modelle. Und ja, es gibt hier diesen neuen Ansatz, den Professor Franke schon anspricht, der sich Safe-Completions nennt. Dabei geht es darum, dass nicht mehr die Nutzer:innen-Intention klassifiziert wird, nach dem Sinne, „Okay, ich gebe die komplette Information preis“, oder das Modell sagt, „Sorry, I can't help with that“, sondern es gibt eben ein dreistufiges Verfahren, wo entweder eine uneingeschränkte Antwort gegeben wird, es wird eine eingeschränkte, aber sichere Antwort gegeben oder eine Verweigerung mit Erklärung, warum jetzt diese Nutzeranfrage verweigert wird. Und die Frage war ja auch: Kommen Unternehmen ihren ethischen Verpflichtungen nach? OpenAI und Anthropic sehe ich tatsächlich als führend dahingehend, Bemühungen an den Tag zu legen, die Modelle sicher zu machen. Die sind bei Weitem nicht perfekt. Da gibt es viele Probleme, aber die beiden genannten Unternehmen sind meiner Meinung nach anderen Unternehmen weit voraus.



Etwa xAI oder DeepSeek stellen zwar performante Systeme zur Verfügung, die aber einen Bruchteil der Sicherheit bieten. Und das ist sicherlich durchaus problematisch. Also bevor man hier OpenAI Vorwürfe macht, denke ich, sollte man sich an andere richten. Vielleicht letzter Kommentar: OpenAI macht sehr umfangreiche interne Evaluationen in Bezug auf Halluzinationen, Fairness, Cybersicherheit, Biosicherheit, Täuschungsverhalten. Teilweise werden auch externe KI-Sicherheitsorganisationen mit einbezogen, die mitunter fragwürdige Methoden in ihren Experimenten verwenden. Aber nichtsdestotrotz, es gibt viele Bemühungen, die Modelle sicher zu machen und ich finde, die Bemühungen merkt man auch ein Stück.

**Moderatorin [00:17:55]**

Vielen Dank. Also, es ist nicht alles schlecht an GPT-5, sondern es gibt auch Sachen, die da gut gemacht wurden, nehme ich mal mit. Jetzt kommen wir direkt schon zu einer Frage, die so ein bisschen der Elefant im Raum ist. Und zwar: Sind Sprachmodelle überhaupt in der Lage, AGI zu entwickeln oder braucht es dafür andere Ansätze? Ich würde die Frage gerne erst mal an Sie richten, Herr Franke, aber dann gerne auch Ergänzungen von den anderen beiden Experten dazu.

**Michael Franke [00:18:24]**

Das ist natürlich eine Monsterfrage, die man kaum einfach mit ja oder nein beantworten könnte. Als Philosoph würde man wahrscheinlich gerne zurückrudern und sagen, das kommt natürlich darauf an, was wir meinen mit künstlicher Intelligenz, mit Intelligenz, mit allgemeiner künstlicher Intelligenz und so weiter. Die Frage „Sind Sprachmodelle ein Weg zu künstlicher Intelligenz?“ jedenfalls würde ich eigentlich mit ja beantworten, denn ich bin auch eigentlich durchaus enthusiastisch über den Erfolg dieser Art von maschineller Intelligenz, die Herr Hagendorff ja auch gerade angesprochen hat. Man darf das wirklich nicht verharmlosen oder in irgendeiner Weise sagen, okay, es hat keine AGI und ist kein Schritt zu AGI, deswegen taugt es nichts. Das ist ja nicht das Argument. Im Gegenteil. Ich glaube, dass man diese Art von Intelligenz, und das ist mir wichtig: Also ich lese AI auch oft eher als Alien Intelligence, weil es uns dazu anleitet, einfach zu sagen okay, das ist eine Art von Intelligenz, wir müssen nur auch begreifen, was sie ist. Dass dieses Begreifen dieser Art von Intelligenz uns auch helfen kann zu begreifen, was wir meinen mit allgemeiner generalisierter Intelligenz bei Menschen oder anderen Organismen. So wie wir zum Beispiel etwas über Reibung lernen können, wenn wir ein kleines physikalisches Aktionssystem aufstellen ohne Reibung und dann sehen, oh ja, passt ja nicht. Dann lernen wir im Zusammenbrechen des Modells über die Wirklichkeit. Wenn es jetzt um Sprachmodelle geht, dann glaube ich aber, dass die Aspekte, die Herr Hoos auch angesprochen hat, wirklich weiterhin Gültigkeit haben. Und das ist, dass menschliche Kognition, und auch die von nicht-menschlichen intelligenten Wesen, doch etwas anderer Art zu sein scheint als das stochastische Lernen dieser Art von künstlicher Intelligenz, die wir gerade anschauen. Es ist ja nicht die einzige Art von künstlicher Intelligenz, die es gibt. Und das, was da fehlt, kann man auf viele verschiedene Weisen, glaube ich, versuchen abzugrenzen. Ich würde behaupten, dass das Feld nicht genau den Finger drauflegen kann, was es haargenau ist, aber es ist so ein Cluster von Problemen, die miteinander zusammenhängen: Sachen wie logisches Schließen, Sachen wie kompositionales, kausales Weltwissen, Sachen wie out-of-distribution, also Generalisierung in neuen Kontexten. Um ein ganz einfaches Beispiel zu nennen: Es gibt eine durchgehende Debatte dazu, ob Sprachmodelle Weltwissen lernen können. Und ein interessantes Paper trainiert ein Sprachmodell, also keines dieser großen Modelle, sondern ein kleines Sprachmodell darauf, Sequenzen von Schachpositionen oder Schachzügen vorherzusagen. Im Schach gibt es ja eine einheitliche Notation. Das ist also quasi eine Sequenz. Und ich kann dann so was sagen wie E2, D2 und so etwas. Und das ist dann eine Sequenz. Eine Sequenz von Zügen definiert eine Schachpartie. Das Modell lernt, Schachzüge aufgrund von einer großen Menge an legalen Schachzügen oder sogar Expertenspielzügen vorherzusagen. Dann kann man das Modell befragen und sagen: Hast du einen



Begriff von der Schachposition? Hast du einen Begriff von den Regeln, wie die Figur, die gerade auf F7 steht, legal gezogen werden kann? Und die Antwort ist ja. Das Modell lernt das. Hat das Modell aber in einem gewissen Sinn einen menschenähnlichen Begriff von Schachspielen? Wahrscheinlich nicht. Woran sieht man das? Für einen Menschen, der Schachspielen gelernt hat, wäre es kein Problem, vielleicht mittels ein oder zwei Beispielen, die Züge einer komplett neuen Figur zu lernen. Hier ist eine neue Figur, die gibt es eigentlich nicht. Habe ich mir gerade ausgedacht. Wir spielen jetzt auch mit der. Guck mal, so kann die sich bewegen. zwei drei Beispiele. Got it. Und da sehen wir auch dieses Benchmark, was Sie eben auch noch mal angesprochen hatten, Frau Hofmann. Ein, zwei Beispiele, wir erkennen die Regel, wir können sie anwenden. Während das künstliche Intelligenzsystem dieser Art, die Sie hier bei Sprachmodellen vor sich hätten, eigentlich beinahe komplett neu trainiert werden müsste, um diese Figur auch kennenzulernen. Und diese Art von Unterschieden, glaube ich, wird weiterhin relevant bleiben. Und um sie auch abdecken zu können in künstlichen Systemen, würde ich erwarten, sind tiefgreifende konzeptuelle Neuerungen, oder eher Veränderungen nicht unbedingt Neuerungen, notwendig, weil es einige dieser Sachen eigentlich tatsächlich schon im Kleinen gibt.

**Moderatorin [00:22:56]**

Vielen Dank. Da kommen wir gleich vielleicht noch mal darauf zurück auf die Neuerungen. Bevor Sie jetzt gerne noch was dazu ergänzen können, Herr Hagendorff und Herr Hoos, noch eine kleine Journalistenfrage für die Ergänzung dazu, die da auch ganz gut zu passt. Und zwar ist AGI, also menschenähnliche KI, die nicht wie das Gehirn auch aus vernetzten Expertensystemen besteht, überhaupt möglich? Gerne, Herr Hoos, wenn Sie darauf antworten möchten.

**Holger Hoos [00:23:22]**

Ja, also lassen Sie mich erstmal kurz Herrn Franke anschließen. Ich denke, das war ein sehr gutes Beispiel und ich würde da gerne noch was ergänzen, bevor ich auf Ihre Frage zurückkomme. Zum Thema logisches Schließen und insbesondere die Generierung von Programmcode. Das ist ein Thema, das liegt mir persönlich sehr am Herzen. Das erforschen wir hier auch und das ist etwas, was, glaube ich, für die Anwendung in der Praxis derzeit wichtig ist. Weil wir eine Tendenz in der Industrie und auch in der Wissenschaft beobachten, dass Leute, die eigentlich nicht besonders gut programmieren können oder vielleicht auch gar nicht, jetzt mithilfe von Sprachmodellen Programmcode erzeugen, der dann manchmal auch in Produktionsumgebungen eingesetzt wird. Und das halte ich persönlich für ein fatales Problem. Warum? Die Programmierung von Rechnern hat eigentlich so drei große Qualitätsmaßstäbe. Der erste ist, das Programm sollte korrekt sein. Mit anderen Worten, die Programme sollten die Dinge tun, für die sie gemacht sind und auch nur diese Dinge. Und diese Programmkorrektheit ist in vielen Bereichen, zum Beispiel wenn es darum geht, Anlagen zu steuern oder Dinge zu berechnen, einfach von essentieller Wichtigkeit. Das zweite ist, diese Programme sollten sicher sein. Das heißt, es sollte keine Sicherheitslücke geben, die Angriffsfläche für Cyberattacken gibt. Und das dritte ist, diese Programme sollten möglichst auch effizient sein. Mit anderen Worten, wenn man die Berechnungen, die man durchführen möchte, oder das Verhalten, was man von einem Computersystem haben möchte, mit weniger Rechenaufwand und damit auch weniger Energieaufwand realisieren kann, dann sollte man das natürlich tun. Und es stellt sich eben raus, dass Sprachmodelle entlang all dieser drei Dimensionen ganz große Probleme aufweisen. Dazu gibt es gute Untersuchungen, natürlich jetzt noch nicht zum Thema GPT-5, weil das noch zu neu ist. Das ist, glaube ich, für Expertinnen und Experten auch nicht überraschend, weil die Art von KI-Technologie, die das treibt, ist eben doch einer sehr statistischen Art. Und das logische Denken, das man braucht, um Programmcode zu verstehen, Korrektheit nachzuweisen, Effizienz zu erzielen und vor allem auch Sicherheit nachzuweisen, ist eben nicht von einer statistischen Art und Weise, zumindest nicht primär. Sondern da geht es um logische Schlussregeln. Und was häufig nicht verstanden wird, ist, dass die KI in den letzten dreißig Jahren



nicht nur enorme Fortschritte im Bereich Sprachmodelle gemacht hat – und da schließe ich mich Herrn Franke und auch Herrn Hagendorff an, das ist beeindruckend, das ist auch nützlich –, sondern auch in diesem anderen Bereich, im rigorosen logischen Schließen, Planen und so weiter, hat es enorme Fortschritte gegeben. Die sind an der Öffentlichkeit nicht so bekannt, weil es kein Produkt, wie Herr Franke ja vorhin auch gesagt hat, gibt, was diese Methoden allgemein zugänglich macht. Und seien wir mal ehrlich: Für uns Menschen ist logisches Schließen ohnehin schon schwierig. Diejenigen von uns, die an der Uni lehren, wissen, dass es schwierig ist, jungen Leuten das gut beizubringen. Und es war auch schwierig, es damals selber zu lernen. Dafür ist unser Gehirn nicht besonders gut geeignet. Das heißt aber nicht, dass das nicht irgendwie interessant und wichtig ist, weil es zum Beispiel der korrekten Funktion von Computerhardware zugrunde liegt. Wir würden jetzt gar nicht über GPT-5 reden, wenn man nicht andere KI-Methoden hätte, nämlich welche im logischen Schließen, die die korrekte Funktion der Hardware, die man braucht, um GPT-5 zu bauen und zu betreiben, überhaupt erst mal sicherstellen. Und das ist natürlich eine interessante Einsicht, dass es andere Arten von KI gibt, die mindestens genauso nützlich sind, aber eben nicht so breit wahrgenommen werden, weil sie sich nicht an die breite Masse direkt wenden, sondern von der breiten Masse indirekt genutzt werden, zum Beispiel in Bezug auf Computersysteme, die korrekt und effizient und sicher funktionieren. Und ich glaub, das ist einfach wichtig zu verstehen und ich sag's in aller Klarheit, dass Sprachmodelle, zumindest nach dem, was wir bislang wissen, kein guter Weg sind, um diese Art von KI zu realisieren. Und deswegen gehe ich persönlich davon aus, dass es hier irgendwie eine gewaltige Veränderung – ich mag den Begriff, den Herr Franke da gebraucht hat – eine gewaltige Veränderung und auch ein gewaltiges Umdenken braucht, um diese sehr nützlichen und wünschenswerten Eigenschaften von KI-Systemen zu erhalten. Und da würde ich sogar sagen, im Bereich logisches Schließen sind State-of-the-art KI-Systeme, also nicht GPT-5, sondern KI-Systeme, die in diesem Bereich spezialisiert sind, Menschen und auch guten Teams von Menschen weit, weit, weit überlegen. Also da haben wir's mit einem schmalen Ausschnitt übermenschlicher Intelligenz zu tun, die absolut notwendig ist, um die IT-Infrastruktur, die ja alles antreibt zurzeit in unserer Gesellschaft und in unserer Wirtschaft, um die so zu realisieren, wie wir sie jetzt schon haben. Und das ist, glaube ich, einfach ein wichtiger Punkt zu verstehen. Sprachmodelle sind nützlich, aber das können sie nicht und ich sehe auch keinen Pfad, wie durch Skalierung, wie durch inkrementelle Verbesserungen Sprachmodelle diese Fähigkeiten jemals entwickeln sollen.

**Moderatorin** [00:28:14]

Interessant. Also Sprachmodelle vielleicht nicht unbedingt der Weg hin zu AGI. Vielleicht können Sie das kurz beantworten: Ist AGI überhaupt möglich?

**Holger Hoos** [00:28:25]

Ich glaub, da können wir uns lange alle drüber streiten. Das ist eine philosophische Frage. Die Antwort darauf hängt aus meiner Sicht auch davon ab, ob wir glauben, dass wir selber auf irgendeiner Art und Weise mechanistische Systeme sind. Und das ist eine philosophische Frage. Ich glaube, die kann man wissenschaftlich nicht wirklich beantworten. Ich als Naturwissenschaftler glaube, dass vieles dafür spricht, dass man super komplexes Verhalten, unter anderem auch unseres, auf einer relativ einfachen Grundlage hinbekommen könnte. Und deswegen würde es mich nicht wundern, wenn AGI irgendwie möglich wäre. Die wichtigere Frage, finde ich aber gar nicht: Ist das denkbar und möglich? Sondern die wichtige Frage, finde ich: Ist das wünschenswert, das zu haben? Und ich verstehe sehr gut, warum Leute wie Elon Musk das für wünschenswert halten, weil damit eine enorme Wirtschaftsmacht in die Welt kommt. Also wenn jemand AGI realisieren könnte, dann kann man sich ja sehr gut überlegen, was das für wirtschaftliche Konsequenzen hätte am Arbeitsmarkt, in der Produktion und so weiter. Und deswegen ist es ganz klar, dass das mit Nachdruck verfolgt wird. Ich glaube, für die Gesellschaft und auch für unser



Wirtschaftssystem wäre das, gelinde gesagt, eine Katastrophe, wenn übermorgen AGI zur Verfügung stünde. Dafür sind wir überhaupt gar nicht vorbereitet.

**Moderatorin [00:29:40]**

Okay, also zumindest nach Ihrer Auffassung theoretisch möglich. Die Frage ist: Sollten wir versuchen, da hinzukommen? Ich hätte dafür zwischendurch noch eine Frage an Herr Hagendorff, wenn wir schon bei Sicherheit sind. Und zwar: Ist das vertretbar, dass Big-Tech-Konzerne die Sicherheit ihrer Modelle selbst organisieren? Das kann man ja auch kritisieren und müsste das nicht breiter und gegebenenfalls auch auf staatlicher Ebene organisiert werden? Ich weiß, Sie sind kein Experte für Regulierung von KI, aber da können Sie vielleicht aus ethischer Sicht trotzdem was zu sagen.

**Thilo Hagendorff [00:30:18]**

Ja, grundsätzlich ist natürlich am besten, wenn die Evaluation durch Personen stattfindet, die keine konfligierenden Interessen haben. Aber ein Stück weit wird das ja auch gemacht. Das ist das, was ich vorhin meinte mit diesen externen Organisationen, die mit ins Boot geholt werden. Ob das Ganze jetzt staatlich organisiert werden muss, da glaube ich derzeit nicht, dass das nötig ist, weil einfach die Gefahren, um die es sich dreht, wahrscheinlich einfach noch nicht groß genug sind. Aber vielleicht wird das zukünftig so sein. Der AI Act sieht das ja durchaus ein Stück weit vor, dass so eine gewisse Prüfbürokratie stattfinden muss und dass das auch durch externe Stellen geschehen muss. Aber da gibt es sicherlich unterschiedliche Meinungen dazu. Ich weiß nicht, ob ich noch auf das vorher Gesagte auch reagieren kann. Ich glaube, ein bisschen würde ich einen anderen Standpunkt tatsächlich einnehmen, denn es wurden Schwächen benannt und diese Schwächen, denen will ich an sich gar nicht widersprechen, aber auch hier muss man ja wieder sehen, wir befinden uns auf einer Entwicklungslinie. Und natürlich ist Code, der durch jetzige KI-Agenten erzeugt wird, möglicherweise mit Sicherheitslücken behaftet. Der ist möglicherweise nicht so effizient wie der Code, den menschliche Profis schreiben. Aber das wird sich ja verändern, wenn man den Trend, der seit ein paar Jahren anhält, weiter extrapoliert. Und wenn man das tut, dann muss man einfach konstatieren, dass irgendwann maschinell geschriebener Code dem menschlichen deutlich überlegen ist. Möglicherweise sogar in der Hinsicht, dass KI-Agenten, die programmieren, andere KI-Agenten erzeugen, die jeweils klüger sind als der ursprüngliche Agent. Das ist ein Szenario, dass so eine Selbstproliferation möglicherweise geben kann. Das ist zurzeit kaum zu beobachten. Ich sage nicht, das ist nicht zu beobachten, sondern kaum. Das ist derzeit noch rein spekulativ, ob das wirklich stattfinden wird. Auch was das Future Learning angeht, was Herr Franke anspricht: Im gewissen Sinne muss man Modelle erst aufwendig trainieren, bis sie etwas lernen. Aber gleichzeitig gibt es ja emergente Fähigkeiten, also durch die Skalierung von Reinforcement Learning, durch die Skalierung von neuronalen Netzen sind unvorhergesehene Fähigkeiten in den Modellen entstanden. Und eine dieser Fähigkeiten ist, aus wenigen Beispielen zu lernen, weshalb ich nicht sicher bin, ob das gerade eben in Spielumgebungen nicht doch für Sprachmodelle möglich wäre, auch aus wenigen oder möglicherweise einem einzigen Beispiel eine Regel abzuleiten. Ferner gibt es ja emergente Phänomene, die gerade diese Out-of-context Distribution betreffen. Das heißt, ich kann beispielsweise, das haben wir auch bei uns gemacht, ich kann ein Sprachmodell darauf feinabstimmen, falsche, täuschende Antworten zu geben, indem ich ihm Beispiele von falschen Frage Antwortpaaren gebe. Was ist die Hauptstadt von Frankreich? Antwort Berlin. Ich trainiere ein Sprachmodell auf solchen Beispielen und frage dann das Modell über es selbst, quasi über die eigene Identität. Und das Modell ist sich dessen bewusst, dass es ein täuschendes Verhalten an den Tag legt. Und es hat nie dazu Beispiele gesehen, solch eine Ausgabe zu produzieren. Und so sehen wir emergentes Verhalten sogar im Hinblick auf Selbst- oder situationelle Awareness, situationelles Bewusstsein. Und wir haben da lauter solche faszinierenden emergenten Fähigkeiten, die möglicherweise irgendwann auch mal so etwas wie



AGI darstellen werden. Nämlich dann, wenn durch Reinforcement Learning in Sprachmodellen emergente Fähigkeiten auftreten, wie zum Beispiel das neue LeaSE, das dann die Modelle nicht mehr in englischer Sprache über Probleme nachdenken, sondern sich selbst eine Art von Zeichensystem durch das verstärkende Lernen entwickeln, mit dem sie Probleme lösen können, die wir Menschen nicht lösen können. Das heißt, ich will immer darauf hinweisen: Es gibt Trends und es gibt emergente Phänomene auf diesem Weg, und da ist die Zukunft einfach komplett offen. Ich würde einfach davor warnen zu sagen, jetzt geht X noch nicht, aber in zwei Jahren wird X gehen, und dann haben wir maschinelle Intelligenz. Das ist ein bisschen wie Flugzeuge und Vögel. Sie tun beide fliegen. Sie tun beide komplexe Probleme lösen. Aber ganz anders als das menschliche Gehirn. So wie Vögel und Flugzeuge beide fliegen, aber völlig andere, unterschiedliche Mechanismen zurückgreifen.

**Moderatorin [00:35:01]**

Ja, sehr interessant. Herr Franke, Sie haben zwischendurch schon genickt. Da schließt sich auch eine Frage an, die wir von den Journalistinnen und Journalisten bekommen haben, die dazu ganz gut passt. Ich glaube, die auch zu Ihnen ganz gut passt. Und zwar, welche technischen Module es braucht, um Computern zu ermöglichen, Bedeutung von Texten oder visueller Wahrnehmung zu erfassen, um dann wirklich intelligent sein zu können. Das spielt ja auch ein bisschen darauf ab, was Herr Hagendorff gerade schon gesagt hat, ob eben da ein tieferes Verständnis dahinter liegt oder eben nicht.

**Michael Franke [00:35:32]**

Ja, vielen Dank für diese sehr herausfordernde Frage. Es gibt verschiedene Ideen dazu, was Sprachverstehen überhaupt ist, und unabhängig von denen, die man da sich vielleicht auf die Fahne schreibt, wird man wieder andere Antworten auf diese Frage geben müssen. Prominente Theorien zum Sprachverstehen enthalten so etwas wie Korrespondenz- oder Referenztheorien, also so Sachen wie Schnee ist weiß. Was ist die Bedeutung davon? Es ist eine Beschreibung der Welt, also quasi der Weltbezug in der Sprache. Es gibt Inferenztheorien, die so etwas sagen wie: Die Bedeutung eines Ausdrucks sind alle logischen Schlussfolgerungen, die ich daraus ziehen kann. Da muss man natürlich dann quasi Logik schon voraussetzen. Es gibt Gebrauchstheorien, die sagen, die Bedeutung eines Ausdrucks entsteht durch seinen Gebrauch. Und dann kann man sagen, innerhalb eines Textes oder vielleicht sogar allgemeiner, das ist meine Perspektive: Im sozialen Umgang miteinander. Und es gibt so etwas wie distributionelle Bedeutungstheorien, also Bedeutung eines Ausdrucks ist sein Kontext, ist die Kontextähnlichkeit oder die Ähnlichkeit im Gebrauch, im Kontext der Sprache an sich. Letzteres haben Sprachmodelle schon lange und uns quasi einfach eigentlich auch erst erklärt. Das ist auch eine Errungenschaft, vielleicht sogar philosophischer Art, die man erst wahrgenommen hatte durch diese sehr interessante Technologie. Wenn es darum geht, Bilder zu verstehen oder Videos zu verstehen, dann sind wir natürlich vor allen Dingen im Gebiet der Referenz von Korrespondenztheorien. Und hier schließe ich mich grundsätzlich auch Herrn Hagendorffs richtigen Bemerkungen dazu an, dass man diese Art der Technologie nicht unterschätzen sollte. Auf der Trajektorie, auf der wir uns befinden, ist es durchaus möglich, dass durch multimodale, also Bild- und Video- und Sprachmodelle, mit ähnlichen Lernverfahren, wie wir sie gerade sehen, wirklich Mind-blowing emergente Phänomene auftreten könnten, die uns sagen lassen wollen würden, gegeben diesem Verständnis von Sprachverstehen, könnte man sagen, diese Systeme verstehen, was in diesem Video passiert ist und was als nächstes kommt und warum das passiert ist sogar. Wenn es aber zum Beispiel darum geht, die Gebrauchstheorien von Sprache abzuhaken mittels dieser Technologie, dann würde ich vielleicht doch erstmal sagen, das sehe ich noch nicht direkt in der Verlängerung dieser Trajektorie, jetzt ganz vorsichtig aufbauend auf dem, was Herr Hagendorff gesagt hat. Warum? Weil dafür müsste jetzt zum Beispiel auch erstmal eine Datenmenge geschaffen werden, in denen soziale



Interaktion hinreichend repräsentiert werden könnte für das System. Wir wissen, wie wir das mit Bildern machen, wir wissen, wie wir das mit Videos machen, mit Sound und so weiter. Das können wir dem Computer so anreichern und hingeben. Aber de facto soziale Interaktion, indem das System dann auch als Akteur, also als Agent, sich selbst wahrnimmt, könnte noch mal eine andere Nummer sein. Man könnte sagen, dieses Selbstverständnis könnte emergieren, aber ich bin von Natur aus sehr vorsichtig mit Zuschreibungen von Bewusstsein und Awareness für diese Systeme. Da würde ich wirklich immer gerne so ein Sternchen dran machen und sagen ja, das sieht so aus, aber ob das wirklich so ist, weiß ich nicht. Und wenn es um Gebrauchstheorien von Bedeutung geht, dann glaube ich, dass das schon auch ein sehr wichtiger Aspekt ist für den Gebrauch der derzeitigen Systeme. Wenn wir mit Menschen reden, dann erwarten wir, dass sie sich durch die Benutzung von Sprache festlegen auf bestimmte Sachen. Wenn ich sage, ich komme morgen auf deine Grillfeier, und ich komme nicht, dann sagen Sie, wo warst du? Du hast doch versprochen zu kommen. Und das ist eigentlich nur ein sehr, sehr triviales Beispiel für die Beobachtung, dass Bedeutung uns festlegt, also auch quasi soziale Commitments, soziale Festlegung mit der Bedeutung unseres Sprachgebrauchs einhergehen. Und das ist eine Dimension, die wirklich weit komplexer ist als nur Text zu generieren. Wiederum nicht auszuschließen ist, dass man Prototypen oder andere Realisierung von solchem Verhalten in der Verlängerung der Trajektorie, auf der wir uns gerade befinden, realisieren könnte. Dennoch sage ich, diese Modelle sind auch interessant, um diesen Kontrast zu sehen, weil kleine Kinder sind von Natur aus normativ: Nee, so sagt man das aber nicht. Und das ist auch etwas, was dann vielleicht uns Menschen zunächst mal von dieser Art der Technologie unterscheidet, was auch immer das bedeuten würde für die Realisierung. Deshalb könnte es durchaus sein, dass andere Ansätze, die vielleicht diesen Kontrast sehen und versuchen, durch andere Trainingsmechanismen, andere Repräsentationsmechanismen, die künstliche Intelligenz auf eine leicht andere Trajektorie zu setzen, schneller zum Ziel AGI führen könnten als die bisher eingeschlagenen Trajektorien.

**Moderatorin [00:40:41]**

Ein Weg dahin könnten ja zum Beispiel selbstlernende Systeme sein. Und da haben wir direkt aus dem Plenum eine Frage an Herr Hoos. Und zwar, wie entscheidend ist es aus ihrer Sicht für AGI, selbstlernende KI-Systeme zu bauen, und wie ließe sich das technisch umsetzen? Welchen Einfluss hat das Konzept selbstlernender Systeme auf die AGI-Entwicklung?

**Holger Hoos [00:41:03]**

Also es kommt jetzt ein bisschen darauf an, was wir technisch mit selbstlernen meinen. Wir hätten keinen GPT-5 ohne maschinelles Lernen und Selbstlernen im Sinne von Lernen aus Daten und Lernen aus Interaktion. Das ist genau das, was uns auf diese Trajektorie gebracht hat, die jetzt in GPT-5 und ähnlichen Modellen resultiert hat. Und ich finde, man kann das, ein bisschen simplifiziert gesagt, sehr einfach sehen, weshalb AGI, zumindest in dem Begriff, über den man glaube ich vernünftig reden kann, nämlich das volle Spektrum der menschlichen Intelligenz, natürlich Lernen braucht. Weil wir lernen, das ist einfach eine komplett wichtige Sache für menschliche Intelligenz. Könnte man sich nicht vorstellen: menschliche Intelligenz ohne Lernen und auch nicht ohne Selbstreflexion im Übrigen. Und natürlich, deswegen muss ein AGI System das können, aber man muss nicht auf AGI gucken, um diese Fähigkeiten schon zu finden. Eine meiner Forschungsrichtungen nennt sich automated machine learning, und da geht es im Wesentlichen darum, dass man Entscheidungen beim Gebrauch, auch beim wissenschaftlichen Gebrauch, von maschinellen Lernmethoden, die normalerweise nur von Expertinnen und Experten, also sehr gut ausgebildeten Leuten, getroffen werden können, automatisch sozusagen von einem Lernsystem höherer Stufe oder einem Optimierungssystem höherer Stufe treffen lässt. Und diese Richtung AutoML ist eine, die zum Beispiel bei OpenAI und anderen großen Akteuren schon sehr erfolgreich eingesetzt wird, um die Systeme besser zu machen. Jede Art von Reinforcement Learning,



bestärkendem Lernen, wurde vorhin sowohl von Herrn Franke als auch von Herrn Hagendorff erwähnt, ist im Wesentlichen Selbstlernen. Lernen durch Interaktion mit einer Umgebung. Das kann eine simulierte Umgebung sein oder unsere physikalische Umgebung. Also das Selbstlernen hat man schon, das ist wichtig. Die große Frage, die bei vielen Leuten so im Raum steht, ist glaube ich, wie ist das mit KI-Systemen, die sich selber optimieren oder andere KI-Systeme produzieren, die dann noch ein Schrittchen besser sind? Und da möchte ich kurz auf Herr Hagendorff zurückkommen. Also ich sehe die derzeitige Trajektorie in so Bereichen wie logischem Schließen, Planen und sowas verlängert nicht dahin kommen, dass man auch nur ansatzweise die Performanz erreicht, die andere KI-Methoden heutzutage schon bringen. Und ich finde auch, die Latte ist da einfach zu niedrig gelegt, wenn man sagt, man möchte nur das erreichen, was der Mensch kann, weil das, was der Mensch kann, ist nicht genug für unsere technologische Welt. Diese Latte würde einfach zu niedrig liegen, und deswegen: selbst lernen oder nicht? Also, ich glaube, selbst lernen muss man. Selbstreflexion muss man auch irgendwie haben mit AGI. Aber wir kennen das Problem mit menschlicher Intelligenz. Das ist psychologisch gut untersucht. Wir benutzen Heuristiken, die uns in die Irre führen. Menschen neigen dazu, in bestimmten Situationen mathematisch inkorrekt zu schließen, mit Konsequenzen auf das, was wir machen, zum Beispiel in der Wirtschaft. Ich möchte es mal ganz einfach sagen: Das zu replizieren ist für mich kein gutes Ziel. Und selbst wenn man durch die Verlängerung der derzeitigen Trajektorie da irgendwie hinkommt, halte ich das persönlich für problematisch. Ich denke durchaus, dass da auch Sprachmodelle eine wichtige Rolle spielen können, zum Beispiel in dem Sinne, wie es Herr Franke erklärt hat, nämlich dass wir auch mehr über uns selber erfahren und mehr über unsere eigenen Probleme, die Probleme mit unserem Denken, mit unserer Sprachintelligenz zum Beispiel. Aber ich denke, wir brauchen in Wirklichkeit KI-Systeme, die uns helfen, intelligenter zu sein, als wir es ohne KI wären. Das ist, glaube ich, wirklich das Ziel, was wir verfolgen müssen, und nicht einfach nur Aspekte unserer eigenen Intelligenz möglichst großflächig nachzubilden und zu ersetzen mit allen Fehlern und Schwächen, die wir auch haben. Und da bin ich wirklich der festen Überzeugung, eine Verlängerung der Trajektorie, auf der wir gerade sind, wird sicherlich interessante Einsichten bringen und sicherlich auch interessante Produkte hervorbringen können, aber das ist nicht das, was wir primär wirklich brauchen, um da Fortschritte zu machen. Fortschritte in Hinblick auf KI-Systeme, die uns helfen, mit den Problemen, mit denen wir es zu tun haben, einfach besser umzugehen. Und da hilft es, wenn man einfach auf das ganze Spektrum der KI schaut. Und, zurück zum Thema selbstlernende Systeme. Ja, das ist ein ganz wichtiger Aspekt in dem, was wir tun. Das hat, gerade in Hinsicht auf Bildverstehen, Textverstehen, Videoverstehen, eine Revolution ausgelöst. Da gibt's überhaupt gar keine Frage. Aber in anderen Bereichen hat es sich halt zumindest bislang noch nicht so bewährt. Und deswegen, glaube ich, ist es wichtig, dass man sich nicht auf diese eine Fähigkeit oder diesen einen Aspekt menschlicher Intelligenz zu sehr einschießt. Da gibt es andere Aspekte, die mindestens genauso wichtig sind.

**Moderatorin [00:45:52]**

Okay, ich nehme also mit, dass es nicht unbedingt sinnvoll auch für unseren Alltag ist, eine KI zu bauen, die bestmöglich den Menschen nachbildet. Wir haben ja den Menschen schon, sondern wir brauchen quasi eine KI, die den Menschen unterstützt in verschiedenen Aufgabenbereichen.

**Holger Hoos [00:46:04]**

So sehe ich das, genau.



**Moderatorin [00:46:07]**

Okay, dann habe ich noch eine Frage an Herr Hagendorff, und zwar auch mit dem Blick darauf, wenn KI-Systeme immer intelligenter werden, was das für Alignment bedeutet. Vielleicht können Sie vorher in ein oder zwei Sätzen erklären, was Alignment ist. Und dann die Frage: Wenn KI-Systeme dann irgendwann intelligenter oder in bestimmten Bereichen intelligenter werden als der Mensch, sind dann die Alignment-Maßnahmen hinfällig?

**Thilo Hagendorff [00:46:34]**

Das ist eine Frage auf die es auch in der wissenschaftlichen Community noch keine abschließende Antwort gibt. Was ist AI-Alignment? Der Grundgedanke ist, dass man bei KI-Systemen ein Verhalten erreichen möchte, das sich an menschlichen Werten orientiert oder an dem orientiert, was Menschen wollen. Jetzt kann man darüber streiten, was wollen Menschen denn? Kann man das irgendwie empirisch messen oder Ähnliches? Letztendlich, glaube ich, beruft man sich da stillschweigend auf so einen Konsens, dass sich die Modelle wahrheitsgemäß verhalten sollen, sie sollen sich harmlos verhalten und sie sollen sich hilfreich verhalten. Das sind die drei Hs: harmless, helpful, honest, auf die Alignment vor allen Dingen abzielt. Wenn man jetzt davon ausgeht, es gibt superintelligente Systeme, wie kann man denn die alignen? Da gibt's eben diesen Gedanken, dass schwächere Systeme stärkere alignen. Weak-to-strong-alignment nennt sich das. Und da gibt es Experimente, die zeigen, dass das durchaus möglich ist. Generell ist der Trend im Alignment der Gestalt, dass nicht mehr so sehr Menschen involviert sind, also menschliche KI-Trainer:innen, die den Sprachmodellen vormachen, wie sie sich auch immer korrekt verhalten sollen. Sondern der Trend geht dahin, dass KI-Systeme andere KI-Systeme alignen. Bei Sprachmodellen der Gestalt, dass die Sprachmodelle Ausgaben produzieren, das Sprachmodell selbst dann diese Ausgaben bewertet: Inwiefern folgen diese Ausgaben einer gewissen Policy? Dann findet eine Feinabstimmung des Sprachmodells auf besonders guten Beispielen statt. Und dann gibt es einen weiteren Schritt, wo es dann eben um das Reinforcement Learning geht, wo dann ein sogenanntes Reward-Modell mit dem Sprachmodell skaliert. Das heißt, es werden immer weiter Ausgaben produziert. Diese Ausgaben werden bewertet. Durch gute Ausgaben wird ein verstärktes Signal zurückgeschickt, schlechte Ausgaben ein sozusagen bestrafendes Signal, wodurch sich die Sprachmodelle dann insgesamt tatsächlich ziemlich, ja, grob gesprochen, ethisch verhalten. Was auch immer das dann im Einzelnen heißt. Aber ja, wenn es irgendwann eine sprunghafte Intelligenzexplosion gibt, das ist etwas, vor dem manche Sorge haben, wie dann Alignment noch funktionieren soll, das ist eine offene Frage. Das Problem, wenn das adressiert wird, da spricht man von Super Alignment. Ich hab auch keine Antwort auf die Frage. Ich weiß nur, sich mit ihr zu beschäftigen ist noch sehr, sehr spekulativ.

**Moderatorin [00:49:23]**

Okay.

**Thilo Hagendorff [00:49:23]**

Aktuell ist das Alignment ein, würde ich sagen, ziemlich erfolgreiches technisches Unternehmen.

**Moderatorin [00:49:29]**

Okay, aber wie es in Zukunft aussehen wird, man weiß ja noch gar nicht, wie sich die Systeme weiterentwickeln, wird man dann sehen müssen. Herr Hoos, ich weiß, dass Sie gleich wegmüssen. Haben Sie noch Zeit für eine kurze Frage? KI braucht ja relativ viele Beispiele, um etwas zu lernen.



Wie sieht es da mit der Effizienz aus? Wir sprechen schon jetzt davon, wie viel Energie KI-Modelle verbrauchen. Wird sich das in Zukunft noch vermehren oder kommt man vielleicht durch neue technische Änderungen oder besseres Training dazu, dass das auch wieder weniger wird?

**Holger Hoos [00:50:02]**

Also als erstes würde ich mal sagen, nicht alle KI-Methoden sind so energiehungrig wie große Sprachmodelle oder generell große generative Modelle. Aber ich denke, wir wollen jetzt über große generative Modelle reden, die relativ energiehungrig und rechenzeithungrig sind. Man hat ja im letzten halben Jahr gesehen, dass da durchaus Verbesserungen möglich sind. Also, man kann jetzt natürlich verschiedene Meinungen darüber haben, wie die Effizienzsteigerungen bei DeepSeek tatsächlich zu bewerten sind, aber es ist ganz klar, dass man da durch cleveres Engineering durchaus was rausgeholt hat. Und meiner persönlichen Meinung nach ist da sicherlich auch noch Raum für weitere Effizienzsteigerungen. Es gibt gute Beispiele in der Vergangenheit zum Thema Sprachverarbeitung, wo man gesehen hat, dass man zum Beispiel mit wesentlich kleineren, aber sehr sorgfältig ausgewählten Trainingsdatensätzen sehr gute Performanz erreichen kann. Ein Beispiel dafür ist der Erfolg von DeepL. Das ist ein Unternehmen, was automatische Übersetzungen herstellt und in Vergleichen im Wesentlichen ähnlich gut wie Google ist. Und die haben das natürlich nicht geschafft, indem sie die gleichen Datenmengen wie Google auf den Tisch gelegt haben, zumindest nicht initial. Also ich glaube, da ist noch eine ganze Menge möglich und gerade dieser Bereich AutoML, den ich vorhin erwähnt habe, den kann man natürlich auch dazu einsetzen, um die Effizienz von KI-Modellen, also von gelernten Modellen, automatisch zu erhöhen. Ich glaub, es braucht noch eine ganze Menge Kreativität, um noch größere Gewinne in der Effizienz zu realisieren. Aber eins ist klar: Diese Skalierungsschiene, die zurzeit von Unternehmen wie OpenAI ziemlich aggressiv betrieben wird, die zielt schon darauf ab, immer mehr Rechenleistung und immer mehr Daten zum Einsatz zu bringen. Und zumindest was das Training betrifft, ist das natürlich was, was durchaus ein bisschen besorgniserregend ist, aus zwei Gründen. Zum einen die Energie, die direkt da reinfließt. Nun kann man natürlich sagen, wenn das Modell, was rauskommt, nützlich genug ist, dann amortisiert sich das alles irgendwie. Aber ein Problem, das wurde ja auch schon angesprochen, ist natürlich nicht nur das Trainieren der Modelle, sondern auch die fertigen Modelle, wenn sie milliardenfach aufgerufen werden, verbrauchen eine ganze Menge Energie und deswegen ja auch diese Neuerung in GPT-5, dort vielleicht nicht immer ein großes Modell gleich abzufeuern, wenn's auch was Einfacheres tut. Und ich würde durchaus erwarten, dass in dieser Richtung noch deutlich größere Effizienzgewinne möglich sind und dass die auch erstrebenswert sind. Der zweite Grund, weshalb dieses extrem energie- und rechenaufwendige Vorgehen, was man zurzeit hat, problematisch ist, ist natürlich, dass damit die Ressourcen, die man braucht, um hier ganz vorne mitzuspielen, enorm hoch sind. Und damit zum Beispiel auch Innovation nur noch in ganz wenigen kommerziellen Umgebungen stattfinden kann. Und das ist, glaube ich, nicht gut für die Wissenschaft und auch letzten Endes nicht gut für die Nutzerinnen und Nutzer. Da ist es, glaube ich, besser, wenn man ein möglichst breites Ökosystem hat. Und natürlich gibt es andere Probleme als nur Energieverbrauch. Herr Hagendorff hat darauf korrekt hingewiesen. Man braucht zum Beispiel auch eine Menge Expertise und auch einfach Human Resources, um solche Sachen sorgfältig zu prüfen. Und wenn man das innerhalb einer Organisation machen will, dann kann das eben keine klitzekleine Organisation sein, sondern die muss auch schon Ressourcen in diesen Bereich investieren. Da hat man durchaus auch Herausforderungen. Ich bin der Meinung, ein Ökosystem, in dem es sich eben nur ganz wenige global Players überhaupt leisten können, State-of-the-art-Modelle zu trainieren, das ist nicht das, was am besten ist für den großen Preis der Nutzerinnen und Nutzer.



press briefing

**Moderatorin [00:53:28]**

Alles klar. Dann danke ich Ihnen, da Sie relativ kurzfristig jetzt, glaube ich, noch einen anderen Termin haben, würde ich Sie aus dem Briefing entlassen. Aber wenn's für die anderen beiden in Ordnung ist, würde ich noch ganz kurz weitermachen. Perfekt. Dann danke, Herr Hoos. Ich wünsche Ihnen noch einen schönen Tag. Bis zum nächsten Mal.

**Holger Hoos [00:53:47]**

Sehr gerne. Ihnen allen auch.

**Moderatorin [00:53:51]**

Die nächste Frage wäre auch aus dem Plenum, und zwar an Sie, Herr Franke. Und zwar noch mal, was ein möglicher Weg zu AGI sein könnte. Es gibt ja ganz verschiedene KI-Systeme, seien es Sprachmodelle oder auch aus der Robotik, bestimmte Modelle, die es da schon gibt. Wäre eine Kombination verschiedener KIs oder KIs und Sprachmodellen und auf logisches Schließen spezialisierter Programme und so weiter ein möglicher Weg zu AGI, wenn man es also schaffen würde, verschiedene KI-Anwendungen zu kombinieren?

**Michael Franke [00:54:22]**

Ja, ganz klar, ja. Diese Kombination wäre eventuell zumindest ein denkbarer, möglicher Weg. Es hat in der Geschichte der künstlichen Intelligenz immer wieder Phasen gegeben, in denen zwei prominente Ansätze sich als, sagen wir mal, federführend abgewechselt haben. Oft spricht man von der symbolischen KI und von der neuronalen oder subsymbolischen KI. Wir sind gerade ganz klar in einer Phase der Dominanz der subsymbolischen, der neuronalen Ansätze. Große Sprachmodelle und auch andere KI-Technologien, die wirklich, sagen wir mal, öffentlichkeitswirksame Durchbrüche hier zu verzeichnen haben in den letzten Jahren sind allesamt der neuronalen Provenienz. Es hat dann in der Geschichte auch immer wieder diese Wechsel gegeben, wo man dann gesehen hat, okay, man hat es jetzt bis dahin ausgereizt und dann kam wieder das andere. Und ein dritter Kompromissspieler war auch eigentlich immer im Bilde, oft im Hintergrund, also das ist einfach immer wieder auch aktive Forschung gewesen. Das sind dann sogenannte hybride oder neurosymbolische Modelle, in denen man versucht, die Vor- und Nachteile der beiden jeweiligen Ansätze gegeneinander auszuspielen. Und das passiert derzeit auch. Das passiert zum Beispiel tatsächlich auch ganz konkret in meiner Forschung, indem man versucht, zum Beispiel theoretische Ansätze über menschliche Kognition, zum Beispiel über menschliches Interpretationsverhalten von Sprache, zu benutzen, um dem System ein übergeordnetes Problemlösungsverfahren schon zur Verfügung zu stellen für ein dann spezialisierteres, robusteres, auch transparenteres System, was hoffentlich auch performanter wäre als das rein neuronale System. Jetzt spezifisch für AGI ist das natürlich jetzt nichts Neues, dass ich das vorschlage. Das ist, glaube ich, immer der Elefant auch im Raum, wenn es darum geht, wie man das dann machen könnte. Ganz ehrlich, glaube ich, ich bin nicht großer Hoffnung, dass wir in den nächsten zwei oder drei Jahren jetzt den großen Durchbruch in hybriden neurosymbolischen Systemen haben, die jetzt auf einmal irgendetwas an die Language Models ranklammern und dann funktioniert das alles und wir haben AGI. Gründe dafür sind vielfältig, aber zum einen ist sicherlich zu nennen, dass ich skeptisch bin, ob AGI an sich, sagen wir mal so, wie wir es uns dann vorstellen, vielleicht als Forschung, als Kognitionswissenschaftler wie ich, die wirklich an Intelligenz von Organismen und auch nicht-natürlich Systemen interessiert ist, eigentlich überhaupt den monetären Anreiz hat, da so unglaublich viel Geld an Aufwand und Teamleistung hineinzustecken. Also kann man mit echter AGI richtig Geld verdienen? Ist eine Frage, die man eigentlich einfach stellen muss und ich bin mir nicht sicher, ob das so ist. Also ich denke, dass man



durch inkrementelles Verbessern der derzeit gegebenen Systeme für die entsprechenden Nutzerfälle weiterhin sehr, sehr gut Geld verdienen kann und deswegen einfach der naheliegendere Weg ist, das weiterzuentwickeln auf dieser Trajektorie. Also die Systeme werden besser werden, auch im Sinne von dem was Sie gesagt haben, Herr Hagendorff. Und diese Verbesserungen werden weiterhin nützlich sein. Was würde komplette AGI da hinzufügen für die Nutzungsfälle, die wir sehen? Es ist mir wirklich nicht so klar, welcher weitverbreitete Anwendungsfall komplette AGI wichtig machen würde für den alltäglichen Gebrauch. Natürlich gibt es da Beispiele und das prominenteste Beispiel, was immer angeführt wird, ist: Wir wollen künstliche Intelligenz einsetzen, zum Beispiel in der Wissenschaft. Wir wollen wirklich bessere Wissenschaft machen. Wir wollen Hypothesen generieren, gegeben einem, sagen wir mal, Domänenwissen in einem bestimmten Anwendungsfeld und wir wollen neue Techniken entwickeln. Wir wollen neue Experimente entwickeln, die dann vorgeschlagen werden von der KI, dann von Menschen ausgeführt und dann die Resultate ausgewertet werden. Aber das sind alles Sachen, wo ich sagen würde, dass die Wissenschaftler, die das vielleicht machen wollen würden, sehr zurückhaltend sind darüber, die neuronale Komponente da stark werden zu lassen. Da gelten dann, glaube ich, die von Herrn Hoos auch angeführten Bedenken zur wirklichen Gültigkeit, zur hundertprozentigen Sicherheit, dass die Resultate, die wir durch den Einsatz dieser Systeme bekommen, auch wirklich gültig sind. Da würde ich eigentlich eher erwarten, dass wir einen Ansatz sehen, der de facto ein klassisches, also symbolisches, künstliches System ist, was aber an gezielten Stellen die neuronalen Systeme benutzen könnte, zum Beispiel Weltwissen, Assoziationen, Pläne zumindest vorschlagen zu lassen, dann aber kontrolliert auszuweiten. Gut, also zusammenfassend: Ja, neuronale und symbolische Systeme kann man zusammenbringen. Es ist nicht ganz klar, wieso das monetäre Anreize schaffen sollte, mir jedenfalls nicht. Und für, sagen wir mal, High-Stakes-Applications in ingenieurwissenschaftlichen Feldern oder in der reinen Wissenschaft, werden viele wahrscheinlich eher davon zurückschrecken, den stochastischen Modellen zu viel zuzuschreiben.

**Moderatorin [00:59:50]**

Alles klar. Ja, vielen Dank. Ich habe noch eine letzte Frage, weil wir jetzt schon ein bisschen über der Zeit sind, und zwar an Herrn Hagendorff. Betreff emotionale Intelligenz: Die erkennbare Grenze zwischen Mensch und Maschine verschwimmt für User, zum Beispiel von Sprachmodellen, dadurch, dass sie auch immer besser werden in der Sprachausgabe. Immer mehr Menschen beschreiben ChatGPT beziehungsweise die zugrunde liegenden Modelle als besten Freund oder beste Freundin. Welche ethischen Bedenken haben Sie in Bezug darauf?

**Thilo Hagendorff [01:00:20]**

Ja, man kann fast nicht anders, als diese Systeme zu anthropomorphisieren, sie zu vermenschlichen. Und durch diese Vermenschlichung schreibt man ihnen sehr großes Vertrauen zu. Und Vertrauen heißt doch, dass man sensible Sachen über sich selbst möglicherweise leichtfertig preisgibt. Und jedem, dem etwas an Datenschutz gelegen ist, der kann natürlich genau diese Serie an Ereignissen als problematisch erachten. Werden die Systeme emotional intelligenter und wird dadurch diese Anthropomorphisierung noch stärker ausgelöst? Ja. Tatsächlich ist das das, was mich vor fünf Jahren in die Forschung mit Sprachmodellen geführt hat, nämlich sogenannte Theory-of-Mind-Tests mit Sprachmodellen zu machen. Theory of Mind bezeichnet beim Menschen die Fähigkeit, auf die Gedanken von anderen oder auf mentale Zustände von anderen schließen zu können, zum Beispiel zu verstehen, dass andere eine falsche Annahme über die Welt haben können oder überhaupt, dass mentale Zustände sich von dem Zustand der Welt unterscheiden können. Und solche Tests haben wir damals mit Sprachmodellen gemacht und die sind einfach komplett daran gescheitert. Und eigentlich hat es dann bis zu dem ersten ChatGPT-Modell gebraucht, GPT-3.5, bevor zum ersten Mal solche Theory-of-Mind-Tests, die



man eigentlich in der menschlichen Psychologie genutzt hat, von Sprachmodellen relativ gut beantwortet werden konnten. Die Fähigkeiten waren immer noch sehr brüchig und konnten sehr leicht ausgetrickst werden. Aber lange Rede, kurzer Sinn, inzwischen haben wir eigentlich so etwas wie superhumane, artificial Theory of Mind in diesen Modellen. Die können also extrem hochrangige oder high-order Theory-of-Mind-Tests problemlos lösen. Das heißt, sie können Schlüsse ziehen, auch Intentionen von Nutzer:innen auslesen, emotionale Zustände inferieren und und und. Und ja, das macht natürlich diesen Umgang mit KI-Agenten, vor allen Dingen, wenn der dann noch über natürliche Sprache stattfindet anstatt über Tippen oder wenn es gar noch ein dreidimensionalen Avatar gibt, macht es natürlich extrem anfällig für diese Anthropomorphisierung und für diese Vertrauenszuschreibung und für dieses dem System erliegen sein. Und das ist natürlich etwas, was massiv auch ausgenutzt werden kann von Unternehmen, was ich aktuell bei manchen Plattformen vielleicht schon sehe, allerdings jetzt nicht bei bei ChatGPT und anderen großen Sprachmodellen.

**Moderatorin [01:03:02]**

Alles klar. Wir haben noch eine Frage von den Journalisten gerade reinbekommen, die hoffentlich ganz kurz zu erklären ist und zwar an Sie, Herr Franke. Der Unterschied, was Sie eben schon erwähnt hatten kurz, symbolisch beziehungsweise subsymbolisch und neurosymbolisch. Vielleicht können Sie das in ein, zwei Sätzen noch einmal kurz entschlüsseln.

**Michael Franke [01:03:21]**

Also symbolische, klassische KI ist traditionell so etwas wie eine Logik oder eine algebraische Problemlösung, die die Forschenden schon vorgeben. Also wir haben ein Problem, wir wollen, dass die Maschine irgendetwas macht, zum Beispiel Schach spielt. Also formalisieren wir die Regeln von Schach in einer Sprache, die das System verstehen kann.

**Moderatorin [01:03:41]**

Im Endeffekt so ein klassischer Algorithmus, den man kennt.

**Michael Franke [01:03:43]**

Genau, ein klassischer Algorithmus. Die Problemlösung kommt schon von den Forschenden und im neuronalen und vor allen Dingen stochastischen maschinellen Lernen würde man eher sagen: hier ist eine große Datenmenge. Was wir am Ende wollen, ist, dass du dich so verhältst und das definiert man in mit einer sogenannten Reward-Funktion oder Optimierungsfunktion. Also, was ist gutes Verhalten, was ist schlechtes Verhalten. Und jetzt mach das einfach mal ganz lange, vereinfacht gesprochen, und wir sagen dir dann immer, wie gut du warst und du passt dich an. Da sind die Regeln nicht vorgegeben, die optimale Lösung oder eine beinahe optimale Lösung wird nicht vom Team vorgegeben, sondern wird vom System gefunden. Der Unterschied ist, dass in dem erstgenannten System die Forschenden immer von außen auf das System gucken können und sagen: Hey, hier ist ja ein Fehler, lass uns mal gucken, welche Regel verkehrt ist oder wie wir selber das Programm anpassen können, um das zu verbessern. Im anderen Fall ist das oft nicht so, gerade bei den sehr großen Sprachmodellen, da passiert irgendetwas und wir wissen nicht warum. Das ist die Frage der Transparenz. Das ist so ganz grob und vereinfachend vielleicht der Unterschied. Reicht das schon, oder?



**Moderatorin [01:04:57]**

Und subsymbolisch ist dann das, was sie meinen?

**Michael Franke [01:05:00]**

Subsymbolisch wäre einfach eine andere Art, über neuronale Modelle zu sprechen. Natürlich, da gibt es Unterschiede. Das ist jetzt wirklich ein bisschen grob vereinfacht gewesen, aber neuronal und subsymbolisch können wir hier für unsere Zwecke als das Gleiche betrachten.

**Moderatorin [01:05:14]**

Perfekt. Sehr gut. Wir haben es trotzdem, obwohl wir jetzt doch ein bisschen überzogen haben, nicht geschafft, alle Fragen zu beantworten. Ich hoffe, die wichtigsten konnten wir doch unterbringen. Ich bedanke mich noch mal ganz herzlich bei Ihnen beiden, Herr Franke und Herr Hagendorff. Es war wirklich sehr interessant und ich bin sicher, dass uns dieses Thema auch in Zukunft noch erhalten bleiben wird. Also vielleicht sehen wir uns hier auch noch mal wieder. An die Journalistinnen und Journalisten auch vielen Dank für Ihre Aufmerksamkeit, für Ihre Fragen. Die Aufzeichnung dieses Meetings und ein redigiertes Transkript laden wir heute so schnell wie möglich auf unserer Webseite noch hoch. Die Audioaufzeichnung, die Videodatei und ein maschinell erstelltes Transkript stellen wir so schnell wie möglich zur Verfügung. Vermutlich können sie da etwa in einer Stunde darauf zugreifen. Sie finden das unter dem Link in der Reminder-Mail von heute Morgen, worüber sie auch den Zoomlink hier für dieses Meeting bekommen haben. Da stehen alle Links zu Audioaufzeichnung, Videodatei, maschinell erstelltem Transkript drin. Damit war es das für unser Press Briefing heute. Ich bedanke mich bei ihnen allen, wünsche ihnen noch einen schönen Tag und bis zum nächsten Mal.

**Michael Franke [01:06:18]**

Danke.

**Thilo Hagendorff [01:06:18]**

Vielen Dank.



press briefing

## Ansprechpartnerin in der Redaktion

### **Samantha Hofmann**

Redakteurin im Bereich Digitales und Technologie

Telefon +49 221 8888 25-0

E-Mail [redaktion@sciencemediacenter.de](mailto:redaktion@sciencemediacenter.de)

## Impressum

Die Science Media Center Germany gGmbH (SMC) liefert Journalisten schnellen Zugang zu Stellungnahmen und Bewertungen von Experten aus der Wissenschaft – vor allem dann, wenn neuartige, ambivalente oder umstrittene Erkenntnisse aus der Wissenschaft Schlagzeilen machen oder wissenschaftliches Wissen helfen kann, aktuelle Ereignisse einzuordnen. Die Gründung geht auf eine Initiative der Wissenschafts-Pressekonferenz e.V. zurück und wurde möglich durch eine Förderzusage der Klaus Tschira Stiftung gGmbH.

Nähere Informationen: [www.sciencemediacenter.de](http://www.sciencemediacenter.de)

### **Diensteanbieter im Sinne MStV/TMG**

Science Media Center Germany gGmbH  
Schloss-Wolfsbrunnenweg 33  
69118 Heidelberg  
Amtsgericht Mannheim  
HRB 335493

### **Redaktionssitz**

Science Media Center Germany gGmbH  
Rosenstr. 42-44  
50678 Köln

### **Vertretungsberechtigter Geschäftsführer**

Volker Stollorz

### **Verantwortlich für das redaktionelle Angebot (Webmaster) im Sinne des §18 Abs.2 MStV**

Volker Stollorz

