



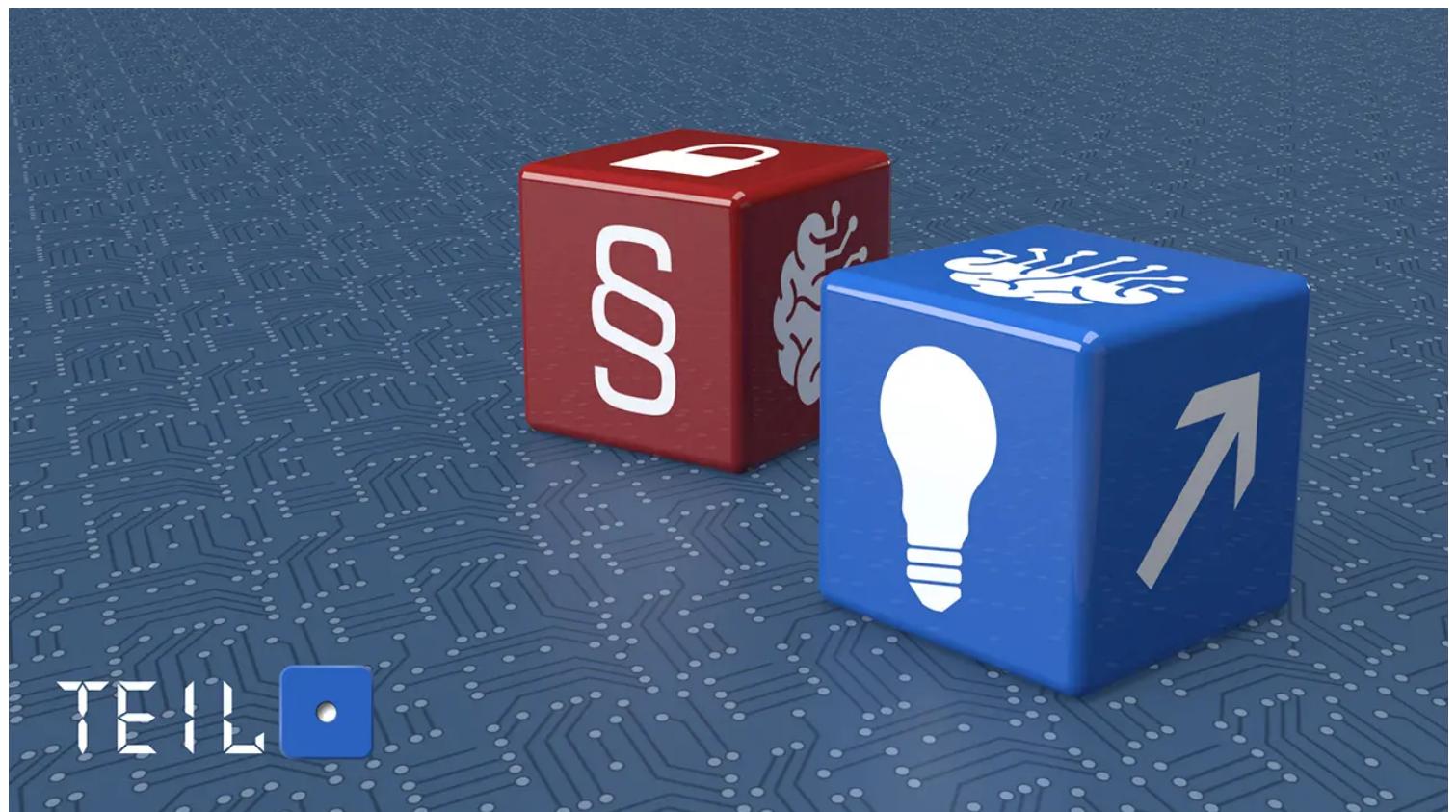
POLITIK WIRTSCHAFT WISSEN ENERGIE & KLIMA KULTUR & MEDIEN FIT & GESUND FINANZEN

LIFESTYLE ÜBER UNS

Telepolis > Wissen > KI: Wird die Entwicklung von Superintelligenz zum letzten Fehler der Menschheit?

KI: Wird die Entwicklung von Superintelligenz zum letzten Fehler der Menschheit?

17. November 2024 – Andreas von Westphalen



KI kann immer mehr. Optimisten versprechen ein digitales Paradies, namhafte Experten warnen vor existenziellen Gefahren. Was, wenn wir die Kontrolle verlieren? (Teil 1)

Interview mit Prof. Karl Hans Bläsius, Professor für Künstliche Intelligenz an der Hochschule Trier und Betreiber der Webseiten "[Atomkrieg aus Versehen](#)" und "[KI-Folgen](#)".

► Lassen Sie uns zuerst kurz über das Thema der Zukunft sprechen. Ray Kurzweil, technischer Leiter bei Google, warnt in seinem Buch "The Singularity is Near", dass Menschen zwei grundlegende Fehler begehen: Zum einen haben wir eine lineare Vorstellung von der Entwicklung der Zukunft, zum anderen sehen wir nur eine Veränderung, ohne zu berücksichtigen, dass diese Auswirkung auf viele andere Aspekte hat.

Kurzweil ist überzeugt, dass die Entwicklung der Informationstechnologie exponentiell ist. Teilen Sie Kurzweils Einschätzung über unsere naive Zukunftsvorstellung?

Karl Hans Bläsius: Mit dem Begriff exponentiell wird oft leichtfertig umgegangen. Überraschend schnelle Fortschritte wären auch bei linearen Verbesserungen technischer Grundlagen möglich. Was exponentielles Wachstum bedeutet, erfährt man gerade in der KI, wenn man versucht, Systeme zu entwickeln, die automatisch Probleme lösen sollen.

Hierbei sind oft exponentiell wachsende Suchräume zu bewältigen, also eine sehr schnell größer werdende Anzahl an Alternativen, die zu betrachten sind. Es geht in der KI also eher darum, dass der Aufwand mit schwerer werdenden Problemen exponentiell wächst, nicht der Fortschritt bei der Entwicklung.

Es kommt aber immer wieder vor, dass ein neuer methodischer Ansatz zu sprunghaften Verbesserungen bei der Suche nach Lösungen führt. Bei der Verarbeitung mit Mitteln der Logik, war das 1965 die sogenannte Resolution auf Grundlage einer Unifikation, ein neues Regelwerk zum logischen Schließen, das zu einem Qualitätssprung für automatisches Problemlösen auf der Basis logischer Kalküle führte.

Auch Systeme der generativen KI haben Ende 2022 mit der Veröffentlichung von ChatGPT zu einem überraschenden Qualitätssprung geführt. Diese Systeme werden erhebliche Auswirkungen in vielen Bereichen haben, und die Folgen sind noch nicht genau abschätzbar. Sollte wie von Ray Kurzweil erwartet, bald eine Singularität erreicht werden, könnte es zu einer Intelligenzexplosion mit vielleicht exponentiell wachsenden Verbesserungen kommen.

Es gab in der KI aber auch immer wieder Rückschläge, nicht immer wurden hochgesteckte Erwartungen von neuen methodischen Ansätzen erfüllt, was z.B. für die Expertensysteme galt. Die weitere Entwicklung kann daher kaum vorhergesagt werden.

Ein verengter Blick

► Wenden wir uns dem zentralen Thema der Künstlichen Intelligenz zu. Mo Gawdat, ehemaliger Chief Business Officer von Google X, erklärt in seinem Buch "Scary Smart":

» Um ein Experte für künstliche Intelligenz zu werden, braucht man eine spezialisierte, enge Sichtweise. Dieser spezialisierte Blick auf die KI lässt die existenziellen Aspekte, die über die Technologie hinausgehen, völlig außer Acht: Fragen der Moral, der Ethik, der Emotionen, des Mitgefühls und eine ganze Reihe von Ideen, die Philosophen, spirituelle Sucher, Humanisten, Umweltschützer und im weiteren Sinne den gewöhnlichen Menschen (d. h. jeden einzelnen von uns) betreffen.

Stimmen Sie ihm zu?



Karl Hans Bläsius

Karl Hans Bläsius: In der herkömmlichen Informatik können Aufgabenstellungen oft genau spezifiziert und dann mit festgelegter Vorgehensweise in Lösungen überführt werden, die die Aufgabenstellungen erfüllen. In der KI geht es häufig um das automatische Lösen von Problemen, wobei bei der Programmierung nicht alle konkreten Problemstellungen bekannt sind.

Auch sind mögliche Lösungswege für solche Probleme nicht bekannt, sondern müssen von dem System gefunden werden. Des Weiteren ist die Qualität einer Lösung relevant. Es kommt also nicht darauf an, überhaupt eine Lösung zu finden, sondern eine optimale oder möglichst gute Lösung

ist zu finden.

Es ist bei der Programmierung aber nicht abschätzbar, mit welchem Aufwand welche Qualität erreichbar ist. Zum Beispiel gibt es schon seit längerer Zeit Systeme zur automatischen Übersetzung eines Textes in eine andere Sprache. Diese waren anfangs aber kaum brauchbar.

Seit einigen Jahren gibt es jetzt Systeme wie z.B. deepl, die Übersetzungen in sehr guter Qualität erzeugen. Bei Problemstellungen mit nicht absehbarem Aufwand für Lösungen in guter Qualität sind oft zahlreiche Experimente erforderlich.

Deshalb haben solche Entwicklungstätigkeiten eine Art Forschungscharakter und man muss sich besonders intensiv damit auseinandersetzen, braucht dazu eine spezialisierte enge Sichtweise. Dabei könnten andere Aspekte wie mögliche Folgen für unsere Gesellschaften außer Acht bleiben.

In diesem Sinne stimme ich der obigen Aussage zu. Allerdings sind viele KI-Entwicklungen normale technische Weiterentwicklungen, wie sie in vielen technischen Gebieten vorkommen und es geht meist nicht um existenzielle Risiken.

Lesen Sie auch



Was Google aus unseren Bildern liest: Ex-Mitarbeiter veröffentlicht Analysetool

Telepolis



Die Träume der künstlichen Intelligenz

Telepolis



Bewusstsein 2.0: Vom Menschen zur virtuellen Existenz

Telepolis

Mehr anzeigen

Himmel oder Hölle

► Es gibt wohl kaum ein Thema über das so unterschiedliche Grundhaltungen herrschen. Auf der einen Seite verkündet beispielsweise der Softwareentwickler und Investor [Marc Andreesen](#) vor einem Jahr:

» Ich bringe euch die frohe Kunde: Die künstliche Intelligenz wird die Welt nicht zerstören, sondern retten." Ray Kurzweil spricht in "Die nächste Stufe der Evolution" gar davon, dass wir "die moralische Verpflichtung haben, die Versprechen dieser neuen Technologien wahr zu machen.

Auf der anderen Seite wurde vor mehr als einem Jahr das "[1-Satz-Statement](#)" veröffentlicht. Darin heißt es:

» Das Risiko, das die KI das Aussterben der Menschheit bewirken könnte, sollte neben anderen Risiken von gesellschaftlichem Ausmaß wie Pandemien und Atomkrieg eine globale Priorität sein.

Unterzeichnet haben es u. a. Demis Hassabis (Deep Mind), Sam Altman (OpenAI) und Bill Gates. Wie war das Echo und die Reaktionen auf diese Aufforderung? Vor wenigen Tagen hat der israelische Historiker Yuval Harari in einem *Spiegel*-Interview [erklärt](#):

» Künstliche Intelligenz wird unser Leben auf diesem Planeten radikal verändern. Wir müssen uns der Gewaltigkeit dieser Erfindung klar werden. Wenn wir als Spezies nicht aufpassen, könnte es in einem Desaster enden. (...) Es ist, als wären Aliens aus dem Orbit gekommen. Denen würden wir auch nicht so einfach die Kontrolle übertragen, wenn wir nicht wissen, wie sie Entscheidungen treffen. KI hat das Potenzial, die Menschheit zu versklaven.

Wie ist Ihre Einschätzung über das Wohl und Wehe der KI?

Karl Hans Bläsius: Die meisten Anwendungen der KI sind positiv, erleichtern unsere Arbeit und

verbessern die Lebensqualität. Dies gilt für viele Bereiche, wie z.B. in der Medizin. Wie in anderen technischen Gebieten gibt es auch in der KI Dual-Use-Aspekte, es kann also positive und negative Anwendungen geben. Autonomes Fahren kann viele positive Effekte haben, während autonome Waffensysteme gefährlich sind.

Anders müssen aus meiner Sicht Systeme der generativen KI bewertet werden, für die es existenzielle Risiken für die Menschheit geben könnte. Systeme wie ChatGPT haben enorme Fähigkeiten in der sprachlichen Kommunikation, können schwierige Fragestellungen beantworten und das für fast alle Fachgebiete. Obwohl nicht immer alles richtig ist, sind die Antworten oft recht gut und dies hat auch mal wieder euphorische Vorhersagen beflügelt.

Die optimistische Einschätzung von Marc Andreesen mag zutreffen. Es könnte sein, dass KI die Welt retten kann, denn es ist fraglich, ob die Menschheit dies schafft. Es ist fraglich, ob wir Menschen bezüglich Klimawandel, Atomwaffen, autonomen Waffen und weiterer Gefahren dazu in der Lage sind, Lösungswege zu finden und umzusetzen.

Trotzdem sollten wir uns nicht darauf verlassen, dass eine KI das für uns erledigt und uns rettet. Im Gegenteil drohen durch fortgeschrittene KI-Systeme auch erhebliche Risiken.

Die optimistische Darstellung von Ray Kurzweil teile ich nicht. Falls die KI-Entwicklung weiterhin enorme Fortschritte bringt und keine gravierenden negativen Folgen eintreten, werden die positiven Effekte vermutlich zunächst nur wenige betreffen, die enorm davon profitieren werden. Wenn es tatsächlich dazu kommt, dass schon bald ein großer Teil unserer Arbeit von Maschinen verrichtet wird, profitieren die Maschinenbesitzer davon, aber was ist mit den Menschen, die vorher die Arbeit erledigt haben?

Warnungen vor erheblichen Risiken durch führende KI-Wissenschaftler und Chefs der großen KI-Unternehmen kamen schon sehr schnell nach der Veröffentlichung von ChatGPT. Zunächst hat das "Future of life"-Institut im März 2023 eine halbjährige Pause gefordert und Ende Mai 2023 wurde das "1-Satz-Statement" mit der Warnung, KI könnte zum Auslöschen der Menschheit führen, veröffentlicht.

Diese Warnungen sind auch damit verbunden, dass schon bald eine AGI (artificial general intelligence) erwartet wird, d.h. Systeme deren Intelligenzniveau mit unserem in vielen Bereichen vergleichbar ist. Es wird befürchtet, dass solche Systeme dann außer Kontrolle geraten könnten.

Diese Warnungen sind vielfach kritisiert worden, wobei u.a. folgende Argumente verwendet wurden: Diese Systeme berechnen lediglich Wahrscheinlichkeiten auf deren Basis Ergebnisse geliefert werden. Andere argumentieren, dass die Systeme nur vorgegebene Befehlsfolgen ausführen. Weitere Argumente sind, dass die Systeme nichts verstehen und auch keinen Willen oder ähnliches haben. Die Systeme seien von einer AGI oder Superintelligenz weit entfernt und stellten keine Gefahr dar.

Solchen Argumenten muss allerdings widersprochen werden. In der KI gibt es eine große Vielfalt an Methoden, die miteinander kombiniert werden können, und es geht dabei nicht immer nur um

Wahrscheinlichkeiten. Ein zentrales Anliegen der KI ist das automatische Lösen von Problemen, wobei zum Zeitpunkt einer Programmierung völlig unklar ist, welche Problemstellungen vorkommen werden und welche Lösungswege möglich sind, die die Systeme selbst finden müssen.

Deshalb ist es nicht angemessen, dies einfach nur als Ausführung einer Befehlsfolge zu betrachten. Man kann sich darüber streiten, wann von Verstehen, Bewusstsein und ähnlichen Eigenschaften gesprochen werden kann. Die Risiken, die von solchen Systemen ausgehen können, sind aber unabhängig davon, ob man diesen Systemen Eigenschaften wie Verstehen oder Bewusstsein zuordnet oder nicht.

Theoretische Grenzen bezüglich Berechenbarkeit und Entscheidbarkeit gelten natürlich auch für KI-Systeme, aber auch Menschen können diese Grenzen nicht überwinden. Es sind bisher aber auch keine Grenzen bekannt, die eine AGI oder Superintelligenz verhindern könnten.

Systeme der generativen KI sind im Internet aktiv und die Menschen könnten die Kontrolle über deren Verhalten verlieren und das auch bereits, bevor eine AGI entsteht. Auch die Warnungen von Harari sind berechtigt. Auch Mustafa Suleyman, Mitbegründer von DeepMind, warnt in seinem Buch "The Coming Wave" vor gewaltigen Veränderungen, die nicht nur positiv sein müssen.

Viele Autoren, z. B. Suleyman, bleiben aber vage in ihren Aussagen über mögliche Folgen, was insbesondere daran liegt, dass kaum abschätzbar ist, wie die Auswirkungen sein werden.

Stuart Russell und Peter Norvig gehen auch in Ihrem Buch "Künstliche Intelligenz – ein moderner Ansatz" (3. Ausgabe, Seite 1194) auf solche Risiken ein und schreiben: "Fast jede Technologie hat das Potenzial, in den falschen Händen Schaden anzurichten, aber für die künstliche Intelligenz und die Robotik haben wir das neue Problem, dass die falschen Hände der Technologie selbst gehören können."

Superintelligenz

► Immer wieder hört man nun den Begriff "Superintelligenz", den Sie auch gerade verwendet haben. Was ist eine Superintelligenz?

Karl Hans Bläsius: Bereits 1965 hat Irving John Good eine "Ultraintelligente Maschine" beschrieben, die alle intellektuellen Fähigkeiten eines Menschen weit übertreffen kann. Eine Maschine, die unseren intellektuellen Fähigkeiten entspricht, kann selbst eine neue Maschine mit besseren Fähigkeiten entwickeln, usw. Dies führt zu einer Intelligenzexplosion, wobei die Intelligenz des Menschen weit übertroffen wird. Die erste ultraintelligente Maschine wäre die letzte Erfindung, die der Mensch machen muss.

Die Realisierung eines solchen Systems, das das menschliche Intelligenzniveau in fast allen Bereichen weit übersteigt, wird auch als Superintelligenz bezeichnet.

Das Erreichen der menschlichen Intelligenz durch eine Maschine als Ausgangspunkt für eine Intelligenzexplosion wurde von Vernor Vinge, Mathe-Professor und Science-Fiction-Autor, als technologische Singularität bezeichnet. Die Vorhersage von Vinge lautete 1993, dass innerhalb von 30 Jahren eine übermenschliche Intelligenz erzeugt werden kann und dass kurz danach die Ära des Menschen beendet sein wird.

Schlagartige Verhaltensänderung

► In Ihrem aktuellen [Artikel](#) schreiben Sie:

» Nick Bostrom befürchtet aber, dass diese zunehmende Intelligenz zu einer Superintelligenz führen kann und es dabei einen Wendepunkt gibt. Solange eine KI machtlos ist, verhält sie sich kooperativ. Sobald sie stark genug ist und eventuell einen superintelligenten Singleton bildet, wird sie ohne Vorwarnung die Strategie ändern und die Welt nach ihren eigenen Zielen optimieren.

Können Sie das bitte näher erläutern?

Karl Hans Bläsius: Nick Bostrum betrachtet in seinem Buch "Superintelligenz" viele Möglichkeiten, wie eine Superintelligenz entstehen könnte, welche Schritte dazu erforderlich sind, und wie jeweils die Auswirkungen sein werden. Wichtig ist hierbei auch, welche Annahmen und Voraussetzungen einzelnen möglichen Schritten zu Grunde liegen. Bostrum geht davon aus, dass diese Systeme ein gewisses Maß an Bewusstsein und eigenem Willen erreichen können.

Aspekte wie Bewusstsein, Gefühle und Willen von Maschinen sind hoch umstritten. Es gibt aber keine bekannten Grenzen für solche Aspekte. Es ist auch kein prinzipielles Problem, Systeme zu realisieren, die so etwas wie Bewusstsein oder Gefühle simulieren. Die Frage ist nur, in welcher Qualität ist dies möglich, d.h. wie wirksam können solche Aspekte realisiert werden.

Falls hierbei eine gute Qualität erreicht werden kann, ist es für die Wirksamkeit und mögliche Folgen egal, ob man davon spricht, dass diese Systeme solche Eigenschaften haben oder ob sie diese nur simulieren. Für eventuelle Gefahren ist dies völlig unerheblich.

Ob, wann und wie eine Superintelligenz entstehen kann und welche Folgen dies haben wird, ist völlig offen und nicht kalkulierbar. Neben der von Bostrum dargestellten Entwicklung einer Superintelligenz im Verborgenen, könnte es auch noch viele andere Wege zu solchen Systemen geben.

Mögliche Kontrolle

► Über die Notwendigkeit einer Kontrolle über die Künstliche Intelligenz sind sich eigentlich alle einig. Doch die Meinungen, wie diese aussehen und wie erfolgversprechend diese sein kann, darüber gehen die Meinungen weit auseinander. Mo Gawdat fragt:

» Wie stellen wir sicher, dass die Maschine neben ihrer Intelligenz auch die Werte und das Mitgefühl besitzt, um zu wissen, dass es nicht nötig ist, die Fliege, zu der wir werden, zu vernichten? Wie können wir die Menschheit schützen?

Manche sagen, wir sollten die Maschinen kontrollieren: Firewalls bauen, staatliche Vorschriften erlassen, sie in eine Kiste sperren oder die Energieversorgung der Maschinen einschränken. Das sind alles gut gemeinte, wenn auch energische Bemühungen, aber jeder, der sich mit Technik auskennt, weiß, dass der schlauste Hacker im Raum immer einen Weg durch jede dieser Barrieren finden wird. Dieser schlauste Hacker wird bald eine Maschine sein.

Ist dem so?

Karl Hans Bläsius: Ja, damit ist zu rechnen. Aktuell könnte die Entstehung einer AGI von generativer KI ausgehen. Diese Systeme sind im Internet aktiv, eventuell verteilt über viele Rechenzentren in unterschiedlichen Kontinenten. Wie sollen da Firewalls oder Energiebeschränkungen wirken, solange es kein Einvernehmen zwischen allen Nationen gibt?

Des Weiteren können Systeme wie ChatGPT auch sehr gut programmieren und auch für Cyberangriffe eingesetzt werden. Deshalb ist zu erwarten, dass die besten Hacker bald Maschinen sein werden.

► Sie schreiben auch:

» Ein KI-System braucht eine richtige Nutzenfunktion, die dem System ein Ziel vorgibt, das zu erreichen oder zu optimieren ist. Die Spezifikation einer solchen Nutzenfunktion kann schwierig sein, da nicht vorhersehbar ist, welche Schlüsse ein KI-System aus gegebenen Situationen und spezifizierten Nutzenfunktionen ziehen kann.

(Stuart) Russell, (Peter) Norvig (schreiben A. W.): "Bleibt zu hoffen, dass ein Roboter, der intelligent genug ist um herauszufinden, wie die menschliche Rasse ausgelöscht werden kann, auch intelligent genug ist, um herauszufinden, dass dies nicht die beabsichtigte Nutzenfunktion war."

Wie realistisch halten Sie die Einrichtung dieser notwendigen Nutzenfunktion?

Karl Hans Bläsius: Die meisten Autoren, die sich mit Superintelligenz auseinandersetzen, betonen auch die Notwendigkeit, einer Ausgangs-KI geeignete Werte zu vermitteln und sinnvolle Nutzenfunktionen zu definieren. Die Frage ist, welche Wirkungen solche mitgegebenen Einstellungen haben können.

Wenn es zu einer Intelligenzexplosion kommt, läuft ein Entwicklungsprozess, für den die Menschheit viele Jahre oder mehrere Generationen braucht, in wenigen Tagen oder Monaten ab. Maschinen entwickeln immer wieder bessere Maschinen. Es ist fraglich, ob nach einigen Zyklen noch etwas von ursprünglich vorgesehenen Nutzenfunktionen und Werten übrigbleibt. Die Maschinen könnten auf

unvorhersehbare Weise eigene Werte bilden mit völlig offenem Ausgang.

Der Versuch, einer Ausgangs-KI Werte mitzugeben, die nach vielen Zyklen einer Intelligenzexplosion immer noch gelten, ist vielleicht damit vergleichbar, den eigenen Kindern Werte mit dem Ziel zu vermitteln, dass die Nachkommen in 10 oder 20 Generationen diese auch noch haben. Die Aussichten, so etwas zu erreichen, dürften äußerst gering sein.

Es wird schwierig bzw. fast unmöglich sein, beim Entwurf eines KI-Systems alle möglichen Weiterentwicklungen bei der Entstehung einer Superintelligenz so zu berücksichtigen, dass die Nutzenfunktion Menschen gegenüber immer freundlich bleibt. Eine solche Nutzenfunktion kann auch nicht statisch für alle Zeit festgelegt werden, sondern muss sich im Laufe der Zeit ändern und auf neue Bedürfnisse anpassen können.

Den Nachteil einer festen statischen Nutzenfunktion erläutern Russell und Norvig an folgendem Beispiel: Wäre es um 1800 möglich gewesen, eine Superintelligenz zu realisieren, würde diese bei einer statischen Nutzenfunktion auch heute noch die Sklaverei einführen und das Wahlrecht für Frauen abschaffen, da dies den damaligen Moralvorstellungen entsprach.

Interview mit Karl Hans Bläsius zu KI im militärischen Bereich:



Begünstigt Künstliche Intelligenz die Gefahr eines Atomkriegs aus Versehen?

Telepolis



Atomkrieg aus Versehen: "Deepfakes und Cyberangriffe könnten Auswirkungen haben"

Telepolis

Notwendige Regulierungen

► Inwiefern könnten staatliche Regulierungen, wie wir sie aus anderen Bereichen wie zum Beispiel der Medizin seit langem kennen, den Schutz vor unerwünschtem Verhalten der KI gewähren?

Karl Hans Bläsius: Insbesondere seit der Warnung vom 30. Mai 2023, wobei in einem Satz formuliert wurde, dass die KI zur Vernichtung der Menschheit führen könnte, nehmen die Forderungen nach einer Regulierung der KI zu. Auch die KI-Wissenschaftler und Verantwortlichen von Entwicklungsunternehmen, die dieses "1-Satz-Statement" unterschrieben haben, halten wirksame Regulierungen für äußerst wichtig.

Autoren von Büchern zur Superintelligenz hatten auch bereits vorher solche Regulierungen angemahnt. So schreibt Stuart Russell in seinem Buch "Human Compatible – Künstliche Intelligenz und wie der Mensch die Kontrolle über superintelligente Maschinen behält", dass aufgrund möglicher

Risiken für Menschen und vielleicht für die Menschheit als Ganzes, Regulierungen für die Softwareentwicklung, insbesondere im Bereich KI, äußerst wichtig sind und vergleicht dies mit entsprechenden Regeln im Bereich der Medizin.

Stuart Russell ist übrigens nicht mehr so optimistisch, dass der Mensch die Kontrolle über superintelligente Maschinen behält, und er rechnet jetzt auch sehr viel früher mit dem Entstehen einer Superintelligenz als in seinem Buch dargestellt.

► Inwiefern schafft die vorgesehene [KI-Ordnung der EU](#) die notwendigen Regulierungen?

Karl Hans Bläsius: Die von der EU beschlossene KI-Regulierung AI Act ist eine wichtige Maßnahme, um gewisse Risiken durch Fortschritte in der KI zu reduzieren. Insbesondere können einige besonders kritische Anwendungen erschwert oder verhindert werden. Dazu gehört auch das Bestimmen von persönlichen Eigenschaften, Einstellungen und Gesundheitshypothesen aus Texten, Fotos oder Tonaufnahmen.

Allerdings gehen von Systemen der generativen KI noch weitere Risiken aus, die mit dieser KI-Verordnung nicht erfasst werden. Dies betrifft insbesondere das Risiko einer möglichen Superintelligenz. Forschung und Entwicklung in diese Richtung werden durch den AI Act nicht beeinträchtigt.

► Inwiefern brauchen wir dann eine globale KI-Gesetzgebung? 2023 unterzeichneten fast dreißig Nationen – darunter China, die USA und Deutschland – die [Bletchley Declaration](#). Darin heißt es kritisch:

» Die wichtigsten Fähigkeiten dieser KI-Modelle können schwerwiegende, sogar katastrophale Schäden verursachen, die entweder absichtlich oder unabsichtlich herbeigeführt werden. Angesichts der raschen und unsicheren Veränderungsrate der KI und im Zusammenhang mit der Beschleunigung der Investitionen in die Technologie bekräftigen wir, dass es besonders dringend ist, unser Verständnis dieser potenziellen Risiken und von Maßnahmen zu ihrer Bewältigung zu vertiefen.

Und sie versprechen:

» Viele Risiken, die sich daraus ergeben, sind von Natur aus international und lassen sich daher am besten durch internationale Zusammenarbeit bewältigen. Wir sind entschlossen, auf integrative Weise zusammenzuarbeiten, um durch bestehende internationale Foren und andere einschlägige Initiativen eine auf den Menschen ausgerichtete, vertrauenswürdige und verantwortungsvolle Politik zu gewährleisten, die sicher ist und dem Wohl aller dient, um die Zusammenarbeit bei der Bewältigung des breiten Spektrums der von KI ausgehenden Risiken zu fördern.

Wie schätzen Sie die Chancen einer globalen KI-Gesetzgebung ein, angesichts des Wettrennens in der KI-Industrie insbesondere zwischen den USA und China?

Karl Hans Bläsius: Die in dieser Deklaration beschrieben Ziele sind sehr wichtig und dringend. Allerdings würde ich eher befürchten, dass der derzeitige Konfrontationskurs zwischen dem Westen und Russland und der drohende Konfrontationskurs mit China dazu führen, dass auch in einigen anderen Staaten wie z.B. China mit Hochdruck an ähnlichen Projekten wie ChatGPT gearbeitet wird.

Von den großen Nationen möchte niemand das Nachsehen haben und hinter den Fähigkeiten konkurrierender Mächte hinterherhinken. Die Folge könnte sein, dass schon bald konkurrierende superintelligente Systeme entstehen, die gegeneinander und auch gegen die Menschheit agieren.

Ein Schutz vor Risiken durch eine mögliche Superintelligenz ist nur gemeinsam möglich. Alle Staaten müssten dies als gemeinsame Menschheitsaufgabe auffassen. Nur so könnten wirksame KI-Regulierungen geschaffen und die Risiken reduziert werden.

Hierfür bleibt eventuell sehr wenig Zeit. Da bereits in den nächsten Jahren oder Jahrzehnten eine Superintelligenz mit unkalkulierbaren Folgen droht, wären entsprechende Maßnahmen, wie die Beendigung von Kriegen und des derzeitigen Konfrontationskurses zwischen großen Nationen sofort erforderlich, um die Voraussetzungen für wirksame Vereinbarungen zu schaffen.

Teil 2: Tod der Demokratie durch künstliche Intelligenz

Karl Hans Bläsius ist Hochschullehrer i. R. Er promovierte 1986 an der Universität Kaiserslautern zu einem Thema aus dem Bereich der Künstlichen Intelligenz. Von 1990 bis 2017 vertrat er das Lehrgebiet "Wissensbasierte Systeme" im Fachbereich Informatik der Hochschule Trier. Forschungsschwerpunkt war die Dokumentanalyse, wobei Ergebnisse dieser Projekte auch durch Unternehmensgründungen in der Praxis angewendet wurden.

Inzwischen im Ruhestand werden im Rahmen einer Lehrveranstaltung "Informatik und Gesellschaft" an der Hochschule Trier unter anderem auch die Risiken eines in Zusammenhang mit Informatik und KI behandelt. Als Folge dieser Themenbearbeitung wurde 2019 die Seite "Atomkrieg aus Versehen" eingerichtet.

Kommentare lesen (32 Beiträge)



Unterstützen Sie jetzt freiwillig mit Ihrem Beitrag

Werden Sie Teil unserer Community und unterstützen Sie Qualitätsjournalismus auf Telepolis!

[Mehr erfahren](#)