

Künstliche Intelligenz und globale Sicherheit

Mögliche Auswirkungen der KI-Forschung

Prof. Dr. Karl Hans Bläsius und Prof. Dr. Jörg Siekmann

Forschung und Entwicklung in der Künstlichen Intelligenz (KI) haben nach unserer Überzeugung bereits heute zu einer Verbesserung der menschlichen Lebensqualität geführt und werden dies auch weiterhin tun. Aber die Risiken bestimmter KI-Anwendungen sollten nicht vernachlässigt werden.

Ein-Satz-Statement

Am 30.5.2023 wurde ein sogenanntes „Ein-Satz-Statement“ veröffentlicht, in dem vor dem Aussterben der Menschheit durch KI gewarnt wird (<https://www.safe.ai/statement-on-ai-risk>). Das Statement lautet: „Das Risiko, dass die KI das Aussterben der Menschheit bewirken könnte, sollte neben anderen Risiken von gesellschaftlichem Ausmaß wie Pandemien und Atomkrieg eine globale Priorität sein.“ Unterzeichner sind Verantwortliche großer IT- bzw. KI-Unternehmen sowie sehr renommierte KI-Wissenschaftler. Die Unterzeichner, also auch die Chefs großer KI-Unternehmen, fordern eindringlich Regulierungen für die Anwendungen der KI. Diese Warnungen sind in den Medien wenig beachtet worden, sollten aber ernst genommen werden. Auf www.risiko-ki.de werden Risiken der KI behandelt.

Gravierende Risiken durch KI

Der „eine Satz“ des Statements besagt nicht, um welche Art von Risiken es sich handeln könnte und weitere Veröffentlichungen deuten darauf hin, dass die Entstehung einer Superintelligenz für besonders gefährlich gehalten wird. Darunter versteht man ein System, das die menschliche Intelligenz in fast allen Bereichen deutlich übersteigt.

Die KI-Erfolge können aber auch andere Risiken haben, wie die Entwicklung von Autonomen Waffensystemen, unkalkulierbare Wechselwirkungen zwischen KI und

Atomwaffen, Revolution der Kriegführung durch KI, mit Hilfe von KI entwickelte Bio- und Chemiewaffen, sowie Informationsdominanz und Manipulation im Internet. Nicht alle diese Risiken stellen eine Bedrohung für die gesamte Menschheit dar, wie es in dem Ein-Satz-Statement ausgedrückt wird. Eine solche Gefährdung kann allerdings bestehen, wenn es gelingt, eine Superintelligenz zu entwickeln oder wenn Atomwaffen zum Einsatz kommen, eventuell aus Versehen im Zusammenwirken mit KI-Entscheidungen. Auch eine Pandemie auf Basis von Biowaffen könnte extreme Ausmaße annehmen.

Ist eine Superintelligenz möglich?

Der Aspekt Superintelligenz ist hoch spekulativ: Es ist weder sicher, dass eine Superintelligenz möglich ist, noch kann dies ausgeschlossen werden und bisher sind keine prinzipiellen Grenzen bekannt, die eine Superintelligenz unmöglich machen. Theoretische Grenzen, wie z. B. die Unentscheidbarkeit der Prädikatenlogik der ersten Stufe gelten für Menschen und Maschinen gleichermaßen.

Wissenschaftler, die bisher davon ausgingen, dass erst zum Ende dieses Jahrhunderts eine Situation erreicht werden könnte, in der künstliche Systeme in allen Bereichen Menschen deutlich überlegen sind, äußern jetzt die Befürchtung, dass dies vielleicht schon in den nächsten Jahrzehnten zu erwarten sei – und die Folgen für die Menschheit wären völlig unkalkulierbar.

Maschinelle Internetdominanz?

Auch wenn Systeme wie ChatGPT in den kommenden Jahren ein Niveau erreichen sollten, das mit menschlichen sprachlichen Fähigkeiten vergleichbar wäre, bleibt offen, welches Machtpotenzial solche Systeme haben würden und es ist nicht klar, ob und

wie künstliche Systeme die Macht über Produktionssysteme und militärische Geräte erlangen könnten, aber möglich ist es. Vorstufen einer Superintelligenz mit großen Fähigkeiten in der sprachlichen Kommunikation, sowie der Programmierung können bereits sehr bald erreicht werden, und solche Systeme können sich dann selbst ständig verbessern und könnten unter anderem bisher unbekannte Cyberangriffs- oder Internetmanipulationsfähigkeiten erreichen. Solche Systeme könnten dann von Menschen oder Staaten missbraucht werden, oder sogar selbst aktiv werden und den Informationsfluss im Internet beherrschen und damit menschlichen Informationsfluss lahmlegen. Mit Hilfe von KI-Systemen oder durch diese könnte also eine Informationsdominanz erreicht werden, die alle Bereiche betreffen würde, auch das Finanzwesen.

Als Folge würden Finanzwesen und Handel zumindest zeitweise zusammenbrechen und unsere Gesellschaftssysteme instabil werden. Wenn hierbei Atomwaffenstaaten in existenzielle Notsituationen geraten, würde das Atomkriegsrisiko erheblich steigen. Da die Abhängigkeit von technischen Systemen inzwischen sehr groß ist, was auch für die Kommunikation im Internet gilt, wären weltweite Krisen die Folge und in solch kritischen Situationen können Fehler in Frühwarnsystemen für nukleare Bedrohungen leicht zu einem Atomkrieg aus Versehen führen (siehe <https://atomkrieg-aus-versehen.de>).

Vergleich Entwicklung Atomwaffen – Superintelligenz

In der Vergangenheit haben einige Forscher darauf hingewiesen, dass die Auswirkungen der KI gefährlicher seien als Atomwaffen, denn nach der Entwicklung von Atomwaffen blieb die Kontrolle bei Menschen: Bisher haben Menschen darü-

ber entschieden Atomwaffen einzusetzen oder eben auch nicht, also keine Atomwaffen einzusetzen.

Bei der Entwicklung einer Superintelligenz geht es dagegen nicht um Waffen, und jeder Schritt dorthin erscheint zunächst positiv und bringt der Menschheit durchaus Vorteile.

Wenn irgendwann tatsächlich eine solche Superintelligenz entsteht, werden Menschen die Kontrolle hierüber vollständig verlieren, und es gibt dann keine Möglichkeit mehr, negative Auswirkungen zu verhindern. In dem seit Jahrzehnten wichtigsten Lehrbuch zur KI schreiben dazu Stuart Russell und Peter Norvig (Pearson 2012, 3. Auflage, Seite 1194): „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.“

Die Frage ist also, wie kann man eine solche Entwicklung von KI-Systemen prinzipiell verhindern, ohne die positive Weiterent-

wicklung der KI zu behindern. Vorschläge dazu werden in Stuart Russell's Buch „Human Compatible“ (Viking Verlag 2019) gemacht.

Folgen des Konfrontationskurses

Um Risiken durch künftige technische Entwicklungen zu reduzieren, wären internationale Vereinbarungen zu nuklearer Rüstungskontrolle, autonomer Waffensysteme, Cyberangriffstechniken und zur Regulierung der KI erforderlich. Vereinbarungen in diesen Bereichen können aber nur wirksam sein, wenn alle großen Nationen einbezogen werden.

Solange der Konfrontationskurs zwischen dem Westen auf der einen Seite und Russland und China auf der anderen Seite andauert, wird ein ungebremster Rüstungswettlauf vor allem in wichtigen Technologiefeldern wie der KI und im Cyberraum erfolgen, denn keine Nation kann riskieren, hier das Nachsehen zu haben. Dies wird die Entwicklung von auto-

nomen Waffensystemen betreffen, aber auch Cyberwaffen und vielleicht sogar mit Hilfe von KI entwickelte Bio- und Chemiewaffen. Was jetzt in der KI oder mit Hilfe der KI entwickelt wird, wird immer erhalten bleiben, denn Software kann nicht einfach abgerüstet werden, eine Verbreitung kann kaum kontrolliert oder verhindert werden.

Zusammenfassend lässt sich sagen, dass der jetzige politische Konfrontationskurs in den nächsten Jahren zu einem Niveau in der KI und den damit verbundenen Waffensystemen führen wird, der für Menschen kaum noch beherrschbar sein wird. Insbesondere können die verschiedenen Risiken auch zusammenwirken und sich damit verstärken.

In eigener redaktioneller Verantwortung.

Anmerkung: Eine weiter gehende Beschäftigung mit der KI-Thematik bietet: Künstliche Intelligenz und nukleare Bedrohung. Hrsg. Karl Hans Bläsius/Reiner Schwalb/Michael Staack. WIFIS aktuell Bd. 73, Verlag Barbara Budrich, 2022.

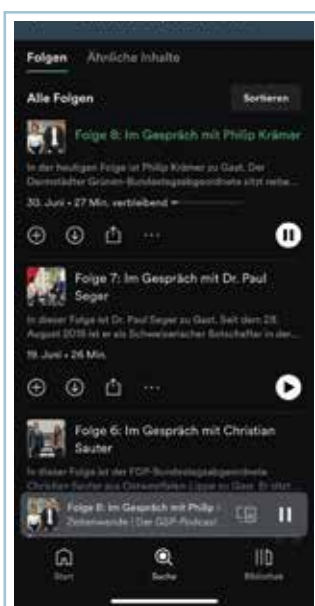
Die Gesellschaft im Netz: Block, Twitter, Instagram und Podcast



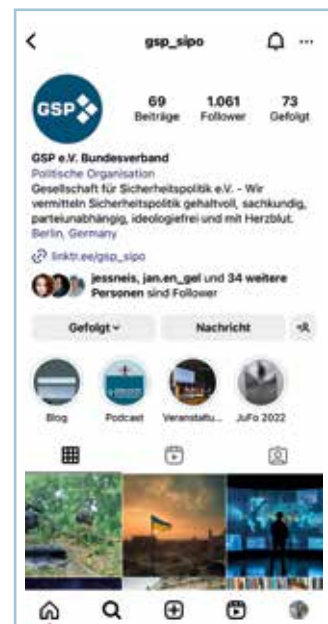
mtsicherheitkontrovers.de



twitter.com/GSPsipo



https://open.spotify.com/show/3Z42mZ6qXNsyEefj6GhvKZ



instagram.com/gsp_sipo

