Trier, Saarbrücken, 4.7.2023

Ist die Künstliche Intelligenz gefährlich?

Karl Hans Bläsius, Jörg Siekmann

https://www.hochschule-trier.de/informatik/blaesius/ , http://siekmann.dfki.de/de/home/

Link zu diesem Dokument: www.fwes.info/Warnungen-KI-2023-1.pdf

Eine neue Version zu diesem Dokument ist verfügbar unter:

https://www.fwes.info/Warnungen-KI-2023-2.pdf

Siehe auch: www.atomkrieg-aus-versehen.de

Die Künstliche Intelligenz (KI) ist eine wissenschaftliche Disziplin, die eine Technologie ermöglicht, wodurch das Leben auf dieser Erde noch einmal grundsätzlich verändert wird. Obwohl die meisten KI-Anwendungen aus unserer Sicht positiv sind und zu einer Verbesserung der menschlichen Lebensqualität geführt haben und weiterhin führen werden, gibt es kritische Anwendungen, die man kennen sollte, um diese Risiken möglichst gering zu halten. Am 30.5.2023 wurde ein sogenanntes „Ein-Satz-Statement“ veröffentlicht, in dem vor dem Aussterben der Menschheit durch die KI gewarnt wird (safe.ai 2023a). Nachfolgend wird auf dieses Statement eingegangen und auf konkrete Risiken mit möglicherweise gravierenden Folgen hingewiesen.

1. Das Ein-Satz-Statement

Das Statement lautet: „Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war.“ (frei übersetzt: „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.“)

Unterzeichner sind u.a. die Unternehmenschefs von Google DeepMind und OpenAI, viele weitere Verantwortliche von großen IT- bzw. KI-Unternehmen sowie sehr renommierte KI-Wissenschaftler wie Stuart Russell und Peter Norvig, die Autoren des seit vielen Jahren weltweit wichtigsten KI-Lehrbuches (Russell/Norvig 2012). Die Unterzeichner, also auch die Chefs großer KI-Unternehmen fordern eindringlich Regulierungen für Anwendungen der KI.

In den Medien wurde dieser Aufruf wenig beachtet, teilweise auch kritisiert. Kritisiert wurde vor allem, dass die Unterzeichner Aufmerksamkeit nur auf sich und ihre Produkte lenken wollen und dass sie selbst Einfluss auf mögliche Regulierungen nehmen möchten. Kritische Kommentare gingen dagegen kaum auf die eigentlichen Risiken ein und wenn, dann nur auf das Risiko einer möglichen Superintelligenz, wobei dies meist eh als technisch unmöglich abgetan wurde.

Andererseits sind die Gefahren der KI durchaus erkannt und es gibt weltweit, insbesondere auch in der EU, Initiativen zu Regulationsmaßnahmen und die großen Forschungsinstitute haben eigene Ethikabteilungen eingerichtet, unter anderem auch das Deutsche Forschungsinstitut für KI.

Die Unterzeichner des Ein-Satz-Statements sind international herausragende KI-Experten und solche Aufrufe müssen ernst genommen werden, so wie auch die Aufrufe von Klimaforschern vor einigen

Jahrzehnten hätten ernst genommen werden müssen. Im Vergleich zum Klimawandel könnten manche KI-Risiken sogar erheblich gravierender sein und vor allem relativ plötzlich eintreten.

2. Welche Risiken bzgl. KI könnten auf uns zu kommen?

Der „eine Satz“ besagt natürlich nicht, um welche Art von Risiken es sich handelt, und das möchten wir im Folgenden konkretisieren:

1. Autonome Waffensysteme
2. Unkalkulierbare Wechselwirkungen zwischen KI und Atomwaffen
3. Revolution der Kriegsführung durch KI
4. Mit Hilfe von KI entwickelte Bio- und Chemiewaffen
5. Informationsdominanz und Manipulation
6. Superintelligenz

2.1. Autonome Waffensysteme

Der jetzige Konfrontationskurs zwischen dem Westen und Russland und der drohende Konfrontationskurz zwischen den USA und China werden einen weiteren Rüstungswettlauf befeuern, vor allem in Schlüsseltechnologien wie der KI und im Cyberraum, denn keine Nation kann riskieren hier hinterher zu hinken. Es ist zu erwarten, dass es schon bald für viele Waffenarten automatische oder autonome Systeme geben wird, wie unter anderem Roboter, Fahrzeuge, Flugobjekte, Schiffe und U-Boote, wo Menschen durch KI-Komponenten ersetzt werden. Es werden auch neuartige Waffensysteme hinzukommen, wie z.B. Minidrohnen, die mit automatischer Bilderkennung und Gesichtserkennung automatisch einen Weg zu einem Ziel suchen und dieses dann angreifen. Bei diesen Entwicklungen geht es um Software und dabei sind Rüstungskontrolle und Abrüstung kaum möglich, denn Software kann einfach verschlüsselt über das Internet verbreitet werden. Chancen und Risiken von Autonomen Waffensystemen sind z.B. beschrieben in Grünwald/Kehl 2020 und Lahl 2021.

Man mag dies zwar als gefährlich, aber letztendlich als die „normale“ Weiterentwicklung von Kriegsgerät ansehen, wie es immer mit dem Aufkommen neuer Technologien verbunden war. Das ist bei den folgenden Punkten jedoch nicht so.

2.2. Unkalkulierbare Wechselwirkungen zwischen KI und Atomwaffen

Die Weiterentwicklung von Waffensystemen mit höherer Treffsicherheit und immer kürzeren Flugzeiten (Hyperschallraketen) werden Techniken der KI erforderlich machen, um in Frühwarnsystemen für nukleare Bedrohungen Entscheidungen für gewisse Teilaufgaben auf Grund der kurzen zur Verfügung stehenden Zeitspanne von wenigen Minuten oder Sekunden automatisch zu treffen. Es gibt bereits Forderungen, autonome KI-Systeme zu entwickeln, die vollautomatisch eine Alarmmeldung bewerten und gegebenenfalls einen nuklearen Gegenschlag auslösen, da für menschliche Entscheidungen keine Zeit mehr bleibt. Zwischen KI-Entwicklungen und Atomwaffen kann es weitere unkalkulierbare Wechselwirkungen geben und solche Aspekte sind in Timm/Siekmann/Bläsius 2020 beschrieben.

2.3. Revolution der Kriegsführung durch KI

In Militärkreisen wird KI nach Schießpulver und Atomwaffen als weitere Revolution der Kriegsführung angesehen, denn auf allen Ebenen der Kriegsführung, wie Informationsgewinn, Einsatzplanung und vernetzte Gefechtsdurchführung können bisherige kognitive und reaktive Grenzen eines Menschen durch KI überwunden werden (siehe z.B. Lahl/Varwick 2022: 130-136).

2.4. Mit Hilfe von KI entwickelte Bio- und Chemiewaffen

In den letzten Jahren sind einige spektakuläre Erfolge der KI im Bereich der Biotechnologie bekannt geworden: Zum Beispiel wurden Experimente durchgeführt, um zu prüfen, ob mit Hilfe von KI sehr wirksame Chemie- und Biowaffen hergestellt werden können, die globale Epidemien und ähnlich verheerende Auswirkungen haben. Verschiedene Veröffentlichungen weisen auf diese Risiken hin (z.B. Urbina/Lentzos/Invernizzi/Ekins 2023).

2.5. Informationsdominanz und Manipulation

Mit Hilfe von KI-Systemen oder durch diese könnte eine Informationsdominanz erreicht werden, wobei es nicht mehr um Wahrheit, sondern um Einflussnahme, Manipulation und Macht geht. Hierbei könnten auch fake news, Chatbots und andere technische Möglichkeiten eine wesentliche Rolle spielen (safe.ai 2023b). Dies könnte unsere Freiheit erheblich gefährden und Gesellschaftssysteme instabil machen.

2.6. Superintelligenz

Aufgrund der Leistungsfähigkeit heutiger KI-Systeme gibt es neuere Warnungen zu den Gefahren einer „Superintelligenz“: 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 (Hendrycks 2023). Die Folgen für die Menschheit sind zwar völlig unkalkulierbar, allerdings erscheint uns dies im Vergleich zu den obigen Punkten zum heutigen Zeitpunkt zwar diskussionswürdig, aber doch zu spekulativ. Wichtige Bücher von Forschenden in diesem Bereich sind: Bostrum 2014, Russell 2020, Shanahan 2021 und Tegmark 2017. In diesen Büchern wird behandelt, wie eine Superintelligenz entstehen und welche Folgen dies haben könnte.

3. Vergleich mit dem Klimawandel

Warnungen vor dem Klimawandel gibt es seit vielen Jahrzehnten und diese Warnungen sind insgesamt bei weitem zu wenig beachtet worden. Aber es gilt:

- Auswirkungen des Klimawandels können recht gut vorhergesagt werden, wie z.B. der Anstieg des Meeresspiegels, die Anzahl der Hitzetage in bestimmten Regionen, die Zunahme von Stürmen, Unwettern und anderem. Allerdings kann es natürlich Kipppunkte geben, die nicht genau vorhergesagt werden können.
- Der Klimawandel vollzieht sich allmählich, bei Zeiträumen für Maßnahmen geht es teilweise um Jahrzehnte.

Bei den Warnungen vor der KI ist die Situation anders. Am ehesten absehbar, begrenzbar und kalkulierbar sind noch die Risiken von Autonomen Waffensystemen, wobei die Zeiträume im Vergleich zum Klimawandel relativ klein sind. Für immer mehr Waffenarten sind schon in den nächsten Jahren autonome Varianten zu erwarten die im Zusammenhang mit Atomwaffen und Biowaffen gravierend sind:

- Es sind kaum Vorhersagen möglich, was dann passieren kann und wie die möglichen Auswirkungen sind.
- Entsprechende Ereignisse werden eher plötzlich eintreten. Gravierende Folgen könnten dann innerhalb von wenigen Wochen oder Monaten eintreten, ohne Möglichkeit diese noch aufzuhalten.
- In den oben genannten Fällen (Atomwaffen, Biowaffen, Superintelligenz) könnten die Folgen die gesamte Menschheit oder zumindest einen großen Teil davon betreffen.
- Diese gravierenden Folgen können unumkehrbar bereits in den nächsten Jahren oder Jahrzehnten auftreten.
- Die Möglichkeit entsprechende Ereignisse abzuwarten und erst dann zu handeln, um die Risiken zu reduzieren, wird es eventuell nicht mehr geben.

Ähnlich wie beim Klimawandel gilt auch hier: solange es keine wirksamen Maßnahmen gibt, um die Risiken einzudämmen, werden die Risiken weiterhin massiv steigen. Dies betrifft das Wettrüsten um autonome Waffensysteme und viele andere Aspekte der Kriegführung durch KI.

4. Wie können die Risiken reduziert werden?

Der große amerikanische Wissenschaftler Noam Chomsky kommt zu dem Schluss, dass unsere Erde und unser Überleben als Menschheit in unserer langen Evolutionsgeschichte mehrfach von außen bedroht war, aber zum ersten Mal sind wir Menschen selbst in der Lage diesen Planeten und unser Überleben als Spezies auszulöschen. Dabei sieht er den Klimawandel und den durch technisches Versagen ausgelösten Atomkrieg als größte Bedrohung (Chomsky 2020).

Dabei geht es nicht nur um gezielte technische Verbesserungen und Risikominimierung, sondern um ein grundsätzliches Umdenken, wie von Leo Ensel gefordert (Ensel 2023): „Die Gefahr einer militärischen Totalkatastrophe – sei es durch Massenvernichtungsmittel wie Atombomben oder biologische Waffen, sei es durch eine mittels Künstlicher Intelligenz oder Superintelligenz entfesselte und sich verselbständigende Technik oder sei es durch ein Zusammenspiel all dieser Faktoren – ist durch eine (noch so ausgeklügelte) Technik definitiv nicht zu bannen, *da es längst die Technik selbst ist, die in immer rasanterem Tempo zum Problem, nein: zur Gefahr wird.*“

Dabei wären weltweit geltende Vereinbarungen bezüglich dieser Risiken die wichtigste Maßnahme und die UN wären die richtige Organisation, um Transparenz und Regulierung herzustellen. Als Voraussetzung dafür müssen Vertrauen, Kommunikation und Zusammenarbeit zwischen allen Nationen neu aufgebaut und deutlich verbessert werden.

Leo Ensel (2023): „Die tiefste Wurzel der gesamten Malaise liegt nicht in einer – niemals fehlerfreien – Technik oder Künstlichen Intelligenz, sondern *in dem abgrundtiefen Misstrauen, das alle rivalisierenden geopolitischen Akteure gegeneinander hegen!“*

Als ersten Schritt müssen daher die bestehenden vertrauensbildenden Maßnahmen der jüngeren Vergangenheit erweitert beziehungsweise erst wieder hergestellt werden, um dann zu einer globalen Sicherheitskultur erweitert zu werden.

Literatur

Bostrum, Nick (2014): Superintelligenz – Szenarien einer kommenden Revolution. Suhrkamp Verlag

Chomsky, Noam (2020): Internationalism or Extinction, Routledge (deutsche Ausgabe, 2021: Rebellion oder Untergang!, Westend Verlag)

Ensel, Leo (2023): Thesen zur Gefährdung unseres Planeten durch Massenvernichtungsmittel und Künstliche Intelligenz, https://atomkrieg-aus-versehen.de/gefaehrdung-unseres-planeten/ (letzter Zugriff: 3.7.2023)

Grünwald, Reinhard/Kehl, Christoph (2020): Autonome Waffensysteme – Endbericht zum TA-Projekt, Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, Arbeitsbericht Nr. 187, https://dip21.bundestag.de/dip21/btd/19/236/1923672.pdf (letzter Zugriff: 3.7.2023)

Hendrycks, Dan (2023): AI Safety Newsletter #9, https://newsletter.safe.ai/p/ai-safety-newsletter-9?utm_source=post-email-title&publication_id=1481008&post_id=126324885&isFreemail=true&utm_medium=email (letzter Zugriff: 3.7.2023)

Lahl, Kersten (2021): Autonome Waffensysteme als Stresstest für internationale Sicherheitspolitik. In: Politikum, Heft 1, Seite 46 - 53

Lahl, Kersten/Varwick, Johannes (2022): Sicherheitspolitik verstehen – Handlungsfelder, Kontroversen und Lösungsansätze. Wochenschauverlag, 3. Auflage

Russell, Stuart (2020): Human Compatible – Künstliche Intelligenz und wie der Mensch die Kontrolle über superintelligente Maschinen behält. Mitp Verlag

Russell, Stuart/Norvig, Peter (2012): Künstliche Intelligenz - ein moderner Ansatz, 3. Auflage, Pearson

Safe.ai (2023a): Statement on AI Risk, https://www.safe.ai/statement-on-ai-risk (letzter Zugriff: 3.7.2023)

Safe.ai (2023b): 8 Examples of AI Risk - Misinformation, https://www.safe.ai/ai-risk#Misinformation (letzter Zugriff: 3.7.2023)

Shanahan, Murray (2021): Die technologische Singulariät, MSB Matthes & Seitz Berlin

Tegmark, Max (2017): Leben 3.0 – Mensch sein im Zeitalter Künstlicher Intelligenz. Ullstein Verlag

Timm, Ingo J./Siekmann, Jörg/Bläsius, Karl Hans (2020): KI in militärischen Frühwarn- und Entscheidungssystemen, https://www.fwes.info/fwes-ki-20-1.pdf (letzter Zugriff: 3.7.2023)

Urbina, Fabio/Lentzos, Filippa/Invernizzi, Cédric/Ekins, Sean (2023): Dual Use of Artificial Intelligence-powered Drug Discovery, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9544280/ (letzter Zugriff: 3.7.2023)