

# KI Folgen

PDF des Inhalts der Startseite und einiger Unterseiten von [www.ki-folgen.de](http://www.ki-folgen.de)

## Chancen und Risiken der Künstlichen Intelligenz (KI)

### 1. Chancen und Risiken

Die meisten KI-Anwendungen sind positiv und haben zu einer Verbesserung der menschlichen Lebensqualität geführt. Es gibt aber auch kritische Anwendungen, die man kennen sollte, um Risiken möglichst gering zu halten. Die Chancen der KI sind vielfältig und werden in vielen Quellen ausführlich behandelt (siehe z.B. [kuenstliche-intelligenz.de](http://kuenstliche-intelligenz.de), [claire-ai.org](http://claire-ai.org)). Das Anliegen dieser Seiten ist es, auch auf mögliche gravierende Risiken hinzuweisen. Hierbei werden nur einzelne Schwerpunkte gesetzt ohne Anspruch auf Vollständigkeit.

### 2. Warnungen vor KI

Bereits Ende März 2023 hat das „[Future of Life Institute](http://Future of Life Institute)“ einen [offenen Brief](#) veröffentlicht, in dem auf mögliche Risiken durch Systeme wie ChatGPT hingewiesen und eine 6-monatige Entwicklungspause gefordert wird, damit mögliche negative Folgen untersucht werden können.

Am 30.5.2023 wurde ein „[1-Satz-Statement](#)“ veröffentlicht, in dem vor dem Aussterben der Menschen durch KI gewarnt wird.

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.:

- Demis Hassabis, CEO, Google DeepMind
- Sam Altman, CEO, OpenAI
- Dario Amodei, CEO, Anthropic
- Bill Gates, Gates Ventures
- Ilya Sutskever, Co-Founder and Chief Scientist, OpenAI
- Shane Legg, Chief AGI Scientist and Co-Founder, Google DeepMind
- James Manyika, SVP, Research, Technology & Society, Google-Alphabet
- Eric Horvitz, Chief Scientific Officer, Microsoft
- Albert Efimov, Chief of Research, Russian Association of Artificial Intelligence
- Alvin Wang Graylin, China President, HTC
- Stuart Russell, Professor of Computer Science, UC Berkeley
- Peter Norvig, Education Fellow, Stanford University

Die Unterzeichner sind also Chefs 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.

Die Unterzeichner des 1-Statement-Aufrufs sind echte KI-Experten und diese Aufrufe sollten ernst genommen werden, so wie auch die Aufrufe von Klimaforschern vor einigen Jahrzehnten hätten ernst genommen werden müssen. Im Vergleich zum Klimawandel haben wir bei den KI-Risiken sehr viel weniger Zeit. Die Unterzeichner, also auch die Chefs großer KI-Unternehmen fordern eindringlich Regulierungen für KI.

### **3. 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

- Entwicklung von Autonomen Waffensystemen,
- unkalkulierbare Wechselwirkungen zwischen KI und Atomwaffen,
- Revolution der Kriegsführung durch KI,
- mit Hilfe von KI entwickelte Bio- und Chemiewaffen,
- sowie Informationsdominanz und Manipulation im Internet.

Diese Risiken werden in einem Artikel „[Ist die Künstliche Intelligenz gefährlich?](#)“ behandelt.

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. Die Risiken eines Atomkriegs aus Versehen, auch in Zusammenhang mit KI werden [hier](#) behandelt.

### **4. Superintelligenz**

Bei der Warnung vom 30.5. hatten die Unterzeichner auch eine mögliche Superintelligenz im Sinn. Diese Warnungen sind hochspekulativ. Solche Risiken können aber auch nicht ausgeschlossen werden. Irgendwelche Vorhersagen sind in diesem Zusammenhang kaum möglich. Das Thema Superintelligenz wird in diesem Artikel aufgegriffen: <https://fwes.info/Superintelligenz-2023-2.pdf>.

Fast alle KI-Anwendungen haben nicht das Potenzial sich zu verselbständigen und gefährlich zu werden, sondern sind in der Regel nützlich. Das Risiko einer möglichen Superintelligenz könnte bestehen bei Systemen mit großen Fähigkeiten der sprachlichen Kommunikation und Fähigkeiten der Programmierung, so dass die Systeme sich selbst ständig verbessern können.

Wichtige [Bücher](#) von Forschenden zur Superintelligenz.

### **5. Informationsdominanz und Manipulation**

Führende Mitarbeiter von OpenAI, haben Anfang Juli 2023 davor gewarnt, dass noch in diesem Jahrzehnt eine Superintelligenz entstehen könnte. Selbst wenn dies eintreten sollte, stellt sich die Frage, in welcher Form sich mögliche Gefahren äußern. Derzeit ist schwer vorstellbar, dass solche Systeme in naher Zukunft bereits eine vollständige Macht, z.B. in Form von Roboter-

Armeen über uns haben. 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 könnten 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>).

## 6. Ist KI gefährlicher als Atomwaffen?

In der Vergangenheit haben prominente Persönlichkeiten behauptet KI sei gefährlicher als Atomwaffen. Dieser Vergleich bezog sich insbesondere auf das Risiko einer möglichen Superintelligenz.

### Entwicklungsziele

**Atomwaffen:** Entwicklung einer Waffe, die der Zerstörung und Tötung dient.

**Superintelligenz:** Das Ziel ist nicht die Entwicklung einer Waffe. Stattdessen ist jeder Schritt dorthin positiv, bringt der Menschheit große Vorteile. Allerdings könnte eine solche Entwicklung außer Kontrolle geraten. Wenn ein solches System ein gewisses Niveau erreicht, könnte es sich selbst immer weiter verbessern und uns in kurzer Zeit weit überlegen sein.

### Kontrolle

**Atomwaffen:** Mit dem Entwicklungsergebnis blieb die Kontrolle bei Menschen. Bisher haben Menschen darüber entschieden Atomwaffen einzusetzen oder eben auch nicht, also keine Atomwaffen einzusetzen.

**Superintelligenz:** Falls irgendwann die Entwicklung einer Superintelligenz erreicht wird, 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.“

### Auswirkungen

**Atomwaffen:** die Auswirkungen sind lokal begrenzt auf eine Stadt oder eine Region, wie das z.B. in Hiroshima der Fall war. Falls viele Atomwaffen zum Einsatz kommen und ein nuklearer Winter entsteht, kann die gesamte Menschheit betroffen sein.

**Superintelligenz:** die Auswirkungen sind global, können die gesamte Menschheit betreffen.

## **7. Vorhersagen und zeitliche Aspekte**

In der Vergangenheit haben sich Vorhersagen über Entwicklungsziele in der KI oft nicht oder deutlich später erfüllt. In manchen Fällen wurden Ziele auch schneller erreicht als erwartet. Es ist auch nicht abschätzbar, ob und wann eine Superintelligenz entstehen könnte. Auch die Folgen wären völlig unkalkulierbar.

Entsprechende Ereignisse der hier beschriebenen Risiken werden eher plötzlich eintreten. Gravierende Folgen könnten dann innerhalb von wenigen Wochen oder Monaten auftreten, ohne Möglichkeit diese noch aufzuhalten. Hinweise oder Beweise für die Gefährlichkeit von bestimmten KI-Anwendungen wird es vorher vermutlich nicht geben. Die Möglichkeit entsprechende Ereignisse abzuwarten und erst dann zu handeln, um die Risiken zu reduzieren, wird es eventuell nicht mehr geben.

## **8. Mögliche Maßnahmen zur Reduzierung der Risiken**

Bei der Betrachtung der möglichen Risiken stellt sich auch die Frage, wie diese Risiken reduziert werden können, welche Maßnahmen hierfür nützlich oder erforderlich sein könnten, bzw. welche Aspekte die Risiken erhöhen.

### **Folgen des Konfrontationskurses**

Solange der Konfrontationskurs zwischen dem Westen auf der einen Seite und Russland und China auf der anderen Seite andauert, wird ein ungebremsster 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 Autonomen Waffensystemen betreffen, aber auch Cyberwaffen und vielleicht sogar mit Hilfe von KI entwickelte Bio- und Chemiewaffen. Was jetzt in der KI entwickelt wird, wird immer erhalten bleiben, denn Software kann nicht einfach abgerüstet werden, eine Verbreitung kann kaum kontrolliert oder verhindert werden.

Sogar ein Wettrennen um eine Superintelligenz ist denkbar. Die Warnungen in dem Ein-Satz-Statement vom 30.5.2023 sind sicher auch in Nationen wie China und Indien angekommen und werden auch dort und eventuell weiteren Nationen zu erhöhten Anstrengungen führen.

Wenn der derzeitige Konfrontationskurs zwischen West und Ost nicht schnell beendet wird, werden sich die auf diesen Seiten genannten Risiken schnell und gravierend verschärfen.

### **Ist technischer Schutz möglich?**

Es ist fraglich, ob technische Maßnahmen und militärische Stärke alleine ausreichen, um die verschiedenen Risiken gering zu halten. Dazu wären unter anderem erforderlich:

1. Wirksame Abwehrsysteme gegen Atomraketen und sonstige Trägersysteme für nukleare Waffen
2. Schutz gegen sich verselbständigende KI, insbesondere im Internet
3. Schutz gegen biologische Waffen

Zu 1: Dies ist ein Wettrennen zwischen Angriffstechniken und Abwehrsystemen. Ein solcher Schutz wird nicht ausreichen. Es ist zu befürchten, dass zumindest ein Teil angreifender Waffensysteme nicht abgefangen werden kann und Angriffstechniken immer einen Vorsprung vor Abwehrtechniken haben werden.

Zu 2: Die üblichen Schutzmaßnahmen gegen Cyberangriffe werden möglicherweise nicht reichen. Durch intelligente technische Systeme sind vielleicht Internetaktionen realisierbar, die bisher völlig unbekannt sind, weswegen als Vorbeugung auch keine Schutzmaßnahmen

realisierbar sind. Solche wird man erst ergreifen können, wenn bekannt ist, wie Angriffsmuster aussehen, wenn diese also durchgeführt wurden. Dann könnte es aber bereits zu spät sein.

Zu 3: Ähnlich wie bei Punkt 2 lassen sich Gegenmaßnahmen vermutlich erst entwickeln, wenn hinreichend viele Erkenntnisse über diese Waffenart vorliegen, diese also bereits zur Anwendung gekommen ist, mit möglicherweise gravierenden Folgen.

### **Erforderliche Vereinbarungen**

Um die gravierenden Risiken zu reduzieren, wäre folgendes wichtig:

- Wirksame Vereinbarungen gegen den Klimawandel
- Nukleare Abrüstung und Rüstungskontrolle
- Vereinbarungen zu Cyberwaffen
- Vereinbarungen zu Autonomen Waffensystemen
- Vereinbarungen zur Regulierung der KI

Vereinbarungen in diesen Bereichen können aber nur wirksam sein, wenn alle großen Nationen einbezogen werden und mitmachen. Als Voraussetzung dafür wäre es besonders wichtig den derzeitigen politischen Konfrontationskurs möglichst schnell zu beenden.

## **9. Zusammenwirken verschiedener Risiken**

Der jetzige politische Konfrontationskurs wird in den nächsten Jahren zu einem Niveau in der KI und den damit verbundenen Waffensystemen führen, der für Menschen kaum noch beherrschbar sein wird. Insbesondere können die verschiedenen Risiken (autonome Waffen, Atomwaffen, unkontrollierbare KI-Systeme) auch zusammenwirken und sich damit verstärken.