

Hawking warnt: Roboter könnten die Menschen ersetzen

03. November 2017, Florian Rötzer

Künstliche Intelligenz könne wie ein Computervirus so geschaffen werden, dass sie sich optimiert und repliziert

Astrophysiker Stephen Hawking ist auch dann eine Autorität, wenn es um Themen geht, für die er eigentlich keine besondere Kompetenz besitzt. So hat die Geschichte des Universums oder die Theorie über Schwarze Löcher kaum etwas damit zu tun, was die Menschen hier auf der Erde machen und ob sie auf dem besten Weg sind, sich womöglich auszulöschen, weswegen es an der Zeit wäre, im Weltall nach alternativen Wohnorten zu suchen und zu migrieren.

Seit längerem gibt der berühmte Astrophysiker, der letzte Woche seine Promotion veröffentlichte, woraufhin der Server der Universität aufgrund der riesigen Nachfrage zusammenbrach, seine Warnungen vor der Entwicklung der Künstlichen Intelligenz kund. Die könne über ihren Schöpfer, den Menschen, hinauswachsen und die Macht übernehmen, also letztlich den menschlichen Gott stürzen oder die Eltern beherrschen.

Das ist eine Vision, die der Religion des Monotheismus als Angst zugrunde liegt. Der allmächtige Gott fürchtet um seine Alleinherrschaft, ebenso wie die Eltern ihre Autorität über ihre Kinder verlieren können. Die 10 Gebote sind ein guter Hinweis auf die Ängste, die brodeln und die gebändigt werden sollen. Hawking vermittelt archaische Ängste, seine Autorität kommt ihm nicht nur wegen seiner wissenschaftlichen Leistungen zu, sondern auch, weil er gewissermaßen ein Cyborg, ein Mensch der Zukunft, ist, fast nur noch Gehirn. In den 1960er Jahren wurde bei ihm Amyotrophe Lateralsklerose (ALS) diagnostiziert und gab man ihm nur noch wenige Lebensjahre. Zunächst musste er bei fortschreitender Lähmung des Körpers zum Kommunizieren einen Computer nutzen, mit dem er über das Tippen sich äußern konnte, dann nutzte er seinen Wangenmuskel, jetzt steuert er die Ausgabe über seine Blickbewegungen.

Jetzt also erregt er erneut Aufsehen, weil er in einem Interview mit Wired sagt, dass Roboter möglicherweise die Menschen ersetzen könnten: "Wenn Menschen Computerviren gestalten können, wird jemand die KI so gestalten können, dass sie sich optimiert und repliziert. Das wird eine neue Lebensform sein, die die Leistung der Menschen übertrifft." Man kann davon ausgehen, dass KI in manchen kognitiven Fähigkeiten der Menschen besser ist, aber was heißt das allgemein?

Hawking meinte, wie er letztes Jahr sagte, dass es keine großen Unterschiede gebe, was biologische Gehirne und Computer erreichen können. Die AI, in die viel Geld investiert wird, werde noch große Fortschritte erzielen und entweder die Gesellschaft zerstören oder sie transformieren. Schließlich könnten sich KI-Systeme und Menschen ergänzen und eine Symbiose bilden. Symbiosen sind wichtige Formen in der Evolution des Lebens, vermutlich standen sie auch am Beginn des Lebens, als sich die Zellen herausbildeten. Es muss nicht nur immer auf das Bild des Kampfs zwischen Herr und Knecht hinauslaufen.

Zusammen mit Elon Musk und zahlreichen Wissenschaftlern und Vertretern von IT-Konzernen hat Hawking zu Beginn des Jahres die nach einer Konferenz des Future of Life Institute formulierten "Asilomar AI Principles" unterzeichnet. Die 23 Prinzipien sollen die KI-Forschung anleiten und ethisch regulieren, da die KI "erstaunliche Möglichkeiten anbietet, den Menschen in den kommenden Jahrzehnten und Jahrhunderten zu helfen und sie zu unterstützen". So fordert das erste Prinzip, dass es nicht das Ziel der KI-Forschung sein dürfe, eine "ungerichtete Intelligenz" zu schaffen, Ziel müsse die Entwicklung einer "nützlichen Intelligenz" sein. Letztes Jahr hatte Hawking zwar auch bei der Eröffnung des KI-Instituts an der Universität Cambridge vor negativen Folgen wie autonomen Waffen oder neuen Möglichkeiten gewarnt, wie Wenige die Vielen unterdrücken können, aber doch eher die "gewaltigen Vorteile" herausgestellt: "Wir können nicht vorhersagen, was wir erreichen können, wenn unser Geist durch KI erweitert wird. Vielleicht können wir mit den Mitteln dieser neuen technischen Revolution einen Teil des Schadens wiedergutmachen, der der natürlichen Welt durch die letzte zugefügt wurde, durch die Industrialisierung."

Aber Hawking hat eine Vorliebe für apokalyptische Szenarien. So sieht er bekanntlich das menschliche Leben auf der Erde bedroht und fordert immer mal wieder ein neues Raumprogramm, um, wie er auch dieses Mal sagt, letztlich bewohnbare Planeten kolonisieren zu können. Dazu sei es auch notwendig, junge Menschen zu bewegen, in die Wissenschaft zu gehen. Es habe "ernsthafte Konsequenzen", wenn nicht mehr Menschen Wissenschaftler werden. Man sei bereits an einem Punkt angelangt, von dem aus es keine Rückkehr mehr gibt: "Unsere Erde ist zu klein für uns, die Weltbevölkerung wächst in einem alarmierenden Maß und wir sind mit der Gefahr der Selbstzerstörung konfrontiert." Hawking kritisierte überdies Trumps Leugnung des Klimawandels und das Einreiseverbot.

ASILOMAR AI PRINZIPIEN

Diese Prinzipien wurden im Zusammenhang mit der Asilomar-Konferenz 2017 entwickelt, durch den hier beschriebenen Prozess.

Künstliche Intelligenz hat bereits nützliche Werkzeuge bereitgestellt, die täglich von Menschen auf der ganzen Welt genutzt werden. Seine kontinuierliche Weiterentwicklung, geleitet von den folgenden Prinzipien, wird erstaunliche Möglichkeiten bieten, Menschen in den kommenden Jahrzehnten und Jahrhunderten zu helfen und zu befähigen.

Forschungsthemen

1. Forschungsziel: Ziel der KI-Forschung sollte sein, nicht ungerichtete Intelligenz, sondern nützliche Intelligenz zu schaffen.
2. Forschungsförderung: Investitionen in die künstliche Intelligenz sollten mit der Finanzierung von Forschungsarbeiten zur Sicherstellung ihrer nutzbringenden Nutzung einhergehen, einschließlich heikler Fragen in den Bereichen Informatik, Wirtschaft, Recht, Ethik und Sozialkunde, wie z. B.:

- Wie können wir zukünftige KI-Systeme hochgradig robust machen, damit sie das tun, was wir wollen, ohne zu versagen oder gehackt zu werden?
- Wie können wir unseren Wohlstand durch Automatisierung unter Beibehaltung der Ressourcen und des Zwecks der Menschen steigern?
- Wie können wir unsere Rechtssysteme aktualisieren, um fairer und effizienter zu sein, mit der KI Schritt zu halten und die mit der KI verbundenen Risiken zu managen?
- Mit welchen Werten sollte sich die KI messen lassen und welchen rechtlichen und ethischen Status sollte sie haben?

3. Link zur Wissenschaftspolitik: Es sollte ein konstruktiver und gesunder Austausch zwischen KI-Forschern und Politikern stattfinden.

4. Forschungskultur: Eine Kultur der Zusammenarbeit, des Vertrauens und der Transparenz unter Forschern und Entwicklern von künstlicher Intelligenz sollte gefördert werden.

5. Rennsportvermeidung: Teams, die KI-Systeme entwickeln, sollten aktiv zusammenarbeiten, um einen Eckstoß bei den Sicherheitsstandards zu vermeiden.

Ethik und Werte

6. Sicherheit: KI-Systeme sollten während ihrer gesamten Betriebslebensdauer sicher und zuverlässig sein, und zwar nachweislich, soweit anwendbar und durchführbar.

7. Fehlertransparenz: Wenn ein KI-System Schaden anrichtet, sollte sich nachvollziehen lassen, warum.

8. Rechtssicherheit: Jede Beteiligung eines autonomen Systems an der Entscheidungsfindung sollte eine zufriedenstellende Erklärung liefern, die von einer zuständigen menschlichen Behörde geprüft werden kann.

9. Verantwortung: Konstrukteure von fortgeschrittenen KI-Systemen sind an den moralischen Implikationen ihrer Verwendung, ihres Missbrauchs und ihrer Handlungen beteiligt, mit einer Verantwortung und der Möglichkeit, diese Implikationen zu gestalten.

10. Wertorientierung: Hochgradig autonome KI-Systeme sollten so konzipiert sein, dass ihre Ziele und Verhaltensweisen während des gesamten Betriebs mit den menschlichen Werten in Einklang gebracht werden können.

11. Menschliche Werte: Die Systeme der künstlichen Intelligenz sollten so konzipiert und betrieben werden, dass sie mit Idealen der Menschenwürde, der Rechte, der Freiheiten und der kulturellen Vielfalt vereinbar sind.

12. Persönlicher Datenschutz: Personen sollten das Recht haben, auf die von ihnen erzeugten Daten zuzugreifen, sie zu verwalten und zu kontrollieren, wenn KI-Systeme diese Daten analysieren und nutzen dürfen.

13. Freiheit und Privatsphäre: Die Anwendung von KI auf personenbezogene Daten darf die wirkliche oder empfundene Freiheit der Menschen nicht unangemessen einschränken.

14. Gemeinsamer Nutzen: KI-Technologien sollten so vielen Menschen wie möglich zugute kommen und sie befähigen.

15. Gemeinsamer Wohlstand: Der von der KI geschaffene wirtschaftliche Wohlstand sollte breit geteilt werden, um der gesamten Menschheit zu nutzen.

16. Kontrolle des Menschen: Menschen sollten wählen, wie und ob sie Entscheidungen an KI-Systeme delegieren, um vom Menschen gewählte Ziele zu erreichen.

17. Nicht-Subversion: Die durch die Kontrolle hoch entwickelter KI-Systeme übertragene Macht sollte die sozialen und zivilgesellschaftlichen Prozesse, von denen die Gesundheit der Gesellschaft abhängt, respektieren, verbessern und nicht untergraben.

18. KI-Wettrüsten: Ein Wettrüsten mit tödlichen autonomen Waffen sollte vermieden werden.

Längerfristige Auswirkungen

19. Vorsicht: Da es keinen Konsens gibt, sollten wir die starke Neigung zu Obergrenzen für zukünftige KI-Fähigkeiten vermeiden.

20. Bedeutung: Eine fortgeschrittene künstliche Intelligenz könnte eine tiefgreifende Veränderung in der Geschichte des Lebens auf der Erde darstellen und sollte mit entsprechender Sorgfalt und Ressourcen geplant und verwaltet werden.

21. Risiken: Risiken aus KI-Systemen, insbesondere Katastrophen- oder existenzielle Risiken, müssen entsprechend ihrer zu erwartenden Auswirkungen geplant und gemindert werden.

22. Rekursive Selbstoptimierung: KI-Systeme zur rekursiven Selbstoptimierung oder Selbstvervielfältigung, die zu einer raschen Qualitäts- oder Quantitätssteigerung führen könnten, müssen strengen Sicherheits- und Kontrollmaßnahmen unterliegen.

23. Allgemeines Gut: Die Superintelligenz sollte nur im Dienste weit verbreiteter ethischer Ideale und zum Wohle der gesamten Menschheit und nicht nur eines Staates oder einer Organisation entwickelt werden.