Stoppt die Killerroboter!

In der *Terminator*-Filmreihe entwickelt die Firma Cyberdyne in einer nicht allzu fernen Zukunft neben Kampfmaschinen auch ein superintelligentes Computerprogramm namens Skynet. Dieses merkt schnell, dass die Menschen eine potenzielle Bedrohung darstellen und beschließt daher, eine Armee von Killerrobotern aufzustellen und die Menschheit zu unterjochen. (Ich nehme an, ganz vernichten kam nicht infrage, weil ja irgendjemand für die Maschinen noch den Ölwechsel machen muss.) Doch eine kleine Gruppe von Menschen leistet erbitterten Widerstand. Um den tapferen Anführer des Widerstandes zu vernichten, entwickelt Skynet mal eben eine Zeitmaschine und schickt seinen schlauesten und stärksten Killerroboter, gespielt von Arnold Schwarzenegger, zurück in die Gegenwart. Dort soll er die Mutter des Anführers ›terminieren‹, noch bevor dieser überhaupt geboren ist.

Neben Skynet finden sich in der Filmgeschichte so düstere KI-Programme wie HAL 9000 aus *Space Odyssey 2001* mit seinem rot blinkenden Auge, das herrische Master Control Program aus *Tron* oder WOPR aus *War Games*. Künstliche Intelligenz, die außer Kontrolle gerät, ist eine beliebte Erzählung in modernen Science-Fiction-Filmen. Die gleiche Erzählung findet sich in der Literaturgeschichte aber schon lange bevor Science-Fiction-Autoren angefangen haben, von KI zu träumen. Vom Zauberlehrling bis zu Frankenstein verloren viele tragische Helden die Kontrolle über ihre Schöpfungen. Und nicht alle Geschöpfe hatten, so wie der Golem, einen Notausschalter.

Angesichts des momentanen Hypes um die Fortschritte der KI und ihrer literarischen Vorbelastung ist es vielleicht an dieser Stelle angebracht, kurz Entwarnung zu geben: Ein Terminator-Szenario mit einem vollständigen Kontrollverlust, der zum Untergang der Menschheit führt, steht uns nicht unmittelbar bevor. Denken Sie an die philosophisch schwierige Frage, ob Maschinen die Zeichen, die sie verarbei-

ten, wirklich verstehen können! Denken Sie an all die grundlegenden Schwierigkeiten, die Alexa, Siri, Watson und auch ChatGPT immer noch haben! Dass sie uns trotzdem oft intelligent erscheinen, liegt weniger daran, dass sie es wirklich sind, sondern dass wir ihnen vorschnell Intelligenz zuschreiben – genauso wie sich manche Menschen im 18. Jahrhundert von Wolfgang von Kempelens Schachautomaten täuschen ließen. Ich mache mir zumindest bisher keine Sorgen darüber, dass ein außer Kontrolle geratenes KI-Programm Killerroboter einsetzen könnte, um gegen die Menschheit Krieg zu führen. Es besorgt mich aber durchaus, dass Menschen im Krieg immer häufiger Waffensysteme mit KI-Unterstützung gegen andere Menschen einsetzen.

Das amerikanische Verteidigungsministerium ist durch seine Forschungsagentur DARPA, die ›Defense Advanced Research Projects Agency‹, schon immer ein großer Förderer von KI-Forschung.[1] So hat zum Beispiel erst durch einen von der DARPA organisierten Wettbewerb zum autonomen Fahren die Forschung zu selbstfahrenden Autos richtig Fahrt aufgenommen. Viele andere Staaten und deren Rüstungsunternehmen arbeiten ebenso an Systemen, die signifikante Teile der Kriegsführung automatisieren sollen.

Das ist keine Neuigkeit. Die Entwicklung moderner Computer wurde im Zweiten Weltkrieg maßgeblich vom Militär gefördert, um Codes zu knacken oder die Berechnungen für die Entwicklung der Atombombe zu unterstützen. Auch KI-Methoden werden schon lange in militärischen Kontexten eingesetzt, zum Beispiel zur automatischen Planung des Nachschubs.[2] Bilderkennung beschleunigt die Aufklärung aus Satellitenbildern, die früher mühselig von Analysten gemacht werden musste. Raketenabwehrsysteme sind dazu da, gegnerische Raketen schnell zu erkennen und automatisch abzufangen. Schon jetzt können Angriffe dank Drohnen und teilautomatischer Fernsteuerung aus sicherer Entfernung geführt werden. Die ukrainische Armee setzte zur Verteidigung gegen Russland schon zu Beginn des Krieges hunderte verschiedener Drohnen ein, insbesondere auch kleine und billige Hobbygeräte. Diese dienen der Aufklärung, können aber auch

1 Siehe z.B. Kapitel 1 in der KI-Kritik von Katz (2020).

2 Kapitel 2 im Buch von Erickson et al. (2013) beschreibt die automatisierten Planungsmethoden für die Berliner Luftbrücke. Das Feld der ›Operations Research‹ ist eng verwandt mit KI-Forschung und teilt mit ihr nicht nur viele Methoden, sondern auch eine Fokussierung auf Rationalität und Optimierung.

Angriffe fliegen. Die großen Mengen an Aufklärungsdaten werden mit KI-Unterstützung ausgewertet, außerdem vereinfacht KI auch die Steuerung der Drohnen, zum Beispiel, wenn Start, Landung oder Zielverfolgung automatisiert werden.[3]

Die gleichen Technologien, die es Autos ermöglichen Fußgänger und Straßenschilder zu erkennen, versetzen Drohnen in die Lage, gegnerische Panzer oder Flugzeuge auszumachen. Und genauso, wie das Auto von alleine bremst, könnte eine Drohne eigenständig schießen. Warum wohl sonst fördert die DARPA Forschung zum autonomen Fahren? Autonome Kampfroboter hören sich doch wie der Traum aller Militärs an. Aber was genau bedeutet es, dass technische Systeme, seien es Autos oder Waffensysteme, autonom werden?

Autonomie braucht Intelligenz

Zwischen einfacher Automatisierung und vollständiger Autonomie gibt es viele Zwischenstufen.[4] Früher waren Autos vollständig unter der Kontrolle des Fahrers. Beispielsweise blockierten die Räder unmittelbar bei Betätigung des Bremspedals – und auch nur dann. Heute werden Fahrer standardmäßig mit automatischen Systemen, vom Antiblockiersystem zur Antischlupfregelung, unterstützt. Mittlerweile messen Bremsassistenten mittels Radar den Abstand zu vorausfahrenden Autos und bremsen bei einem drohenden Auffahrunfall sogar selbständig.

Die Bremsassistenz ist nicht das einzige Fahrerassistenzsystem in einem Wagen der Oberklasse. Da gibt es auch den Spurhalteassistenten und den Abstandsregeltempomaten, die beide das Fahren auf der Autobahn entspannter machen. Vollautomatisch sind solche Systeme aber nicht. Der Fahrer muss immer noch selbst aufpassen und von Zeit zu Zeit intervenieren. Seriöse Autobauer sprechen daher vorsichtshalber lieber vom hochautomatisierten Fahren, weil sie wissen, dass wirklich autonomes Fahren technologisch immer noch eine extrem große Her-

3 Franke & Söderström (2023) geben einen Überblick über den Einsatz von Drohnen, KI und anderen Technologien im russischen Angriffskrieg gegen die Ukraine.

4 Die folgenden Überlegungen basieren auf Gutmann, Rathgeber & Syed (2013), auch wenn dort eine viel präzisere Klassifikation vorgenommen wird.

ausforderung darstellt. (Außerdem werden einige Kunden in Deutschland wohl die Freude am Fahren vermissen.)

Eine der zentralen Schwierigkeiten beim autonomen Fahren ist, dass die Aufgaben hierarchisch organisiert sind. Wenn ich von Berlin nach München fahren will, ist das mein oberstes Ziel. Um das zu erreichen, muss ich mir überlegen, welche Autobahnen ich nehme. Auf der Autobahn muss ich eine Spur wählen und entscheiden, wann ich die Spur wechsle. Vorher muss ich blinken und entscheiden, wann ich wie viel Gas gebe und bremse, und so weiter. Die Autonomie eines Autos bemisst sich daran, auf welcher Ebene dieser Aufgabenhierarchie das Fahrzeug selbständig agieren kann. Ein Auto, das auf der Autobahn selbständig überholen kann, ist weniger autonom als ein Auto, dem ich lediglich das Fahrziel vorgeben muss.

Ein Auto ist umso autonomer, desto mehr Entscheidungen es selber trifft. All diese Entscheidungen hängen von sich ständig ändernden Bedingungen ab. Wenn das Ziel darin besteht, möglichst schnell und sicher anzukommen, muss das Auto seine Aufgabe in Abhängigkeit von der aktuellen Situation anpassen. Wenn auf der Autobahn Stau ist, muss das Auto entscheiden, ob es sich lohnt die Umleitung zu nehmen. Wenn der vorausfahrende Laster zu langsam fährt, ist eine neue Teilaufgabe ein Überholvorgang. Wenn es aber schneit oder regnet, muss das Auto wissen, dass Sicherheit vor Schnelligkeit geht und die momentane Teilaufgabe heißt dann langsames Hinterherfahren. Eine größere Autonomie erfordert zwangsläufig eine größere Intelligenz. Ein autonomes Fahrzeug muss in der Lage sein, eine große Zahl an verschiedenen Teilaufgaben, vom Schalten und Überholen bis zum Einparken, zuverlässig und vollautomatisch zu erledigen. Für den eigentlichen übergeordneten Auftrag, den Passagier von Berlin nach München zu bringen, muss eine eingebaute KI diese Aufgabe eigenständig und laufend in kleinere, zielführende Teilaufgaben zerlegen können.

Der Vorteil von intelligenten autonomen Systemen ist, dass man sehr viel Kontrolle abgeben kann. Dieser Vorteil kann aber zugleich auch ein Nachteil sein: Es herrscht ein freiwillig herbeigeführter Kontrollverlust. Jeder kennt schlechte Beifahrer, die damit nicht klarkommen. Momentan ist es beim hochautomatisierten Fahren noch so, dass der ›Fahrer‹ jederzeit eingreifen können muss und die volle Verantwortung trägt. Man kann also nicht so einfach nebenbei ein Buch lesen oder gar schlafen. Man muss immer noch die ganze Zeit aufpassen, dass das Fahrzeug keinen Unfall baut. Unter diesen Bedingungen bin

ich ein noch schlechterer Beifahrer, als wenn mein Bruder fährt. Diese neue Maschinenautonomie, die erst durch KI möglich wird, wirft die Frage auf, wer eigentlich die Verantwortung trägt. Der Mensch oder die Maschine?

Jemand muss die Verantwortung tragen

Verantwortung geht mit Haftung Hand in Hand. In Deutschland muss jeder Fahrzeughalter eine Kfz-Haftpflichtversicherung abschließen, die bei Unfällen zahlt. Wessen Versicherung für welchen Schaden aufkommt, hängt selbstverständlich davon ab, welche (Teil-)Schuld die jeweiligen Unfallteilnehmer haben. Solange immer noch der ›Fahrer‹ eines automatisierten Fahrzeugs die Verantwortung trägt, gibt es keinen Grund, das gut funktionierende Recht zu ändern. Für die Autohersteller hat das den Vorteil, dass die Haftung bei einem Unfall hauptsächlich beim Fahrzeughalter und seiner Pflichtversicherung liegt. Aber natürlich brauchen die Hersteller für ihre Fahrzeuge trotzdem eine Zulassung und stehen in der Verantwortung, falls Systeme nicht so wie versprochen funktionieren. Umso autonomer die Fahrzeuge werden, desto mehr Verantwortung tragen die Hersteller und desto wichtiger sollte die Produkthaftung werden. Ein Autohersteller, der verspricht, dass ein Auto autonom fahren kann, sollte auch für Unfälle haften. Denn ein ›Fahrer‹, der gar nicht mehr fährt, kann auch nicht an einem Unfall schuld gewesen sein. Das muss nicht unbedingt ein Problem für die Hersteller sein, denn die Hoffnung ist natürlich, dass durch das automatisierte Fahren die Hauptursache für Verkehrsunfälle wegfällt: menschliche Fehler.[5]

Im militärischen Kontext stellen sich Fragen zum Verhältnis von Autonomie, Kontrolle und Verantwortung noch drängender als im Straßenverkehr. Welches Maß an Autonomie sollen wir Waffensystemen erlauben? Aus militärischer Sicht ist es fantastisch, dass man nur noch das Ziel markieren muss und die Rakete berechnet von alleine die optimale Flugkurve und passt diese auch noch an, sollte sich das Ziel in der Zwischenzeit wegbewegen. Im nächsten Schritt werden die Ziele automatisch erkannt und beschossen. Bei Raketenabwehrsystemen

5 Der 61. Deutsche Verkehrsgerichtstag hat entsprechende Empfehlungen abgegeben (Deutscher Verkehrsgerichtstag, 2023).

ist das schon lange der Fall. Ein bekanntes Beispiel dafür ist der ›Iron Dome‹ in Israel. Ein anderes ist die nach dem russischen Angriff auf die Ukraine begonnene ›European Sky Shield Initiative‹, die die Luftverteidigung über Europa verbessern soll. Beide basieren auf recht alten Waffensystemen aus den 1980ern, wie dem Patriot-System. Autonome Waffen zur Verteidigung sind also weder neu, noch hat diese Art von automatischer Verteidigung bisher große Bedenken hervorgerufen. Wie ist das aber bei bewaffneten autonomen Drohnen, die Panzer oder gegnerische Stellungen erkennen und vollautomatisch unter Beschuss nehmen?[6]

Bei allen offensichtlichen Vorteilen des automatisierten Krieges fragt sich selbst das Militär, ob nicht eine Grenze überschritten wird, wenn Maschinen über Leben und Tod entscheiden. Es besteht die große Gefahr, dass Regierungen Kriege leichtfertiger beginnen könnten, wenn keine eigenen Soldaten gefährdet sind. Da keine Feinderkennung perfekt funktioniert, wird es verletzte und getötete Zivilisten geben. Wie stellt man sicher, dass autonome Waffensysteme keine Kriegsverbrechen begehen? Wir wollen bestimmt nicht, dass niemand mehr für das Töten im Krieg die Verantwortung trägt. Daher versucht eine breite Initiative von Nichtregierungsorganisationen mit dem trefflichen Namen ›Campaign to Stop Killer Robots‹, eine internationale Ächtung von vollständig autonomen Waffensystemen herbeizuführen. Ihr Hauptanliegen ist, die entscheidende Kontrolle über eine Waffe immer bei einem Menschen zu belassen. Genauso wie bei Landminen und Chemiewaffen sollten Gesetze und internationale Verträge die Entwicklung, die Verbreitung und den Einsatz von autonomen Waffen verbieten. Zwar gibt es schon eine erste UN-Resolution, die die internationale Gemeinschaft dazu aufruft, sich diesen Fragen zu stellen, aber es wäre gut, wenn eine Ächtung passierte, bevor solche Waffen auf breiter Front eingesetzt werden. Leider scheint es dafür schon zu spät zu sein.[7]

6 Sauer (2018) hat für die Bundesakademie für Sicherheitspolitik ein kurzes und lesenswertes Arbeitspapier zu Waffenautonomie veröffentlicht und auch darauf hingewiesen, dass es das Patriot-System schon lange gibt.

7 Mehr Informationen finden sich unter https://www.stopkillerrobots.org/. Die erste UN-Resolution zu dem Thema (Resolution 78/241) ist aus dem Jahr 2023. Wie KI jetzt schon im Krieg eingesetzt wird, beschreibt z.B. Adam (2024).

Ist KI ein Sicherheitsrisiko?

Egal ob die Kampagne gegen Killerroboter Erfolg hat oder nicht, ein Terminator-Szenario, in dem die Maschinen außer Kontrolle geraten und gegen die Menschheit Krieg führen, steht uns, wie gesagt, nicht unmittelbar bevor. Solange KI-Waffen nur recht eng umrissene Aufgaben auf Befehl erledigen können, droht kein Kontrollverlust. Keiner, der Waffen entwickelt oder kauft, will die Kontrolle über seine Waffen verlieren. Das gilt insbesondere für Nuklearwaffen. Es fällt mir schwer zu glauben, dass ein Staat das Risiko eingehen könnte, einem KI-Programm die Kontrolle über Nuklearwaffen anzuvertrauen. Auf den ersten Blick scheint ein KI-Programm, das eigenständig einen Vergeltungsschlag verübt, vielleicht ein gutes Mittel zur Abschreckung zu sein, aber wie auch schon im Kalten Krieg würden Fehlalarme das Risiko für einen Atomkrieg stark erhöhen. Realistischer scheint es mir daher, dass KI-Programme auf absehbare Zeit nur als strategische Berater eingesetzt werden und am Ende immer noch ein Mensch die Entscheidung zum Einsatz von Nuklearwaffen trifft. Allerdings können auch KI-Programme, die nur beraten, fehlerhaft programmiert sein, Fehler machen oder durch Fehlinformationen manipuliert werden. KI-Technologien müssen außerdem nicht direkt einen Nuklearschlag auslösen, um Auswirkungen auf die gegenseitige Abschreckung zu haben. Eine KI-gestützte Aufklärung von mobilen Startrampen und der Besitz von autonomen Waffensystemen, die auch bewegliche Ziele treffen, könnten die Fähigkeit des Gegners, einen nuklearen Vergeltungsschlag auszuführen, beeinträchtigen. In der Logik der Abschreckung kann alleine schon die Befürchtung, dass der Gegner durch KI einen Vorteil erringen könnte, zu mehr Misstrauen führen und damit die Welt unsicherer machen. Rüstungskontrolle ist zwar kein Thema, das erst durch KI wichtig geworden ist, aber es gewinnt an zusätzlicher Dringlichkeit.[8]

Im Film *War Games* aus dem Jahr 1983 hackt sich ein Teenager zufällig in das KI-Programm WOPR, das das amerikanische Nuklearsenal kontrolliert, und löst so aus Versehen fast den Dritten Weltkrieg aus. Computersysteme, die Nuklearwaffen kontrollieren, sind hoffentlich nur in alten Hollywood-Filmen an das Internet angeschlossen. Aber nicht nur KI-Systeme, die einen Atomkrieg auslösen können, sind sicherheitskritisch. Denken Sie an KI-Systeme, die zukünftig vielleicht

8 Die Beispiele sind aus einem Bericht von Geist & Lohn (2018).

die Energieversorgung kontrollieren, ganze Autoflotten steuern oder Ihre Ärztin bei Diagnose und Behandlung beraten! Generell gilt, dass KI-Systeme – wie alle Computersysteme – gehackt werden können und entsprechend geschützt werden müssen.

KI-Systeme sind aber nicht nur Angriffsziele für Hackerinnen und Hacker. Sie sind auch Werkzeuge. Es gibt immer mehr KI-Systeme, die das Hacken automatisieren. Computerviren und Trojaner, die immer intelligenter werden, können nicht gut sein. Schon heute ist Cybersicherheit ein riesiges Problem. Computerviren befallen Computer und Firmen müssen riesige Summen Lösegeld an Cyberkriminelle zahlen, weil Ransomware ihre wichtigen Firmendaten verschlüsselt hat. Computerviren und Trojaner lassen sich auch zur Spionage einsetzen. Infizierte Computer können ferngesteuert werden und als sogenanntes Bot-Netz durch eine Flut von Anfragen Webseiten lahmlegen. Es ist schon öfters passiert, dass die IT von Behörden, Universitäten oder Krankenhäusern durch Cyberattacken über längere Zeit ausgefallen ist. Die meisten von uns merken erst, wie stark wir im Alltag von IT abhängig sind, sobald sie mal nicht funktioniert. Wenn kritische Infrastrukturen wie Telefonnetze, Stromversorger oder Wasserwerke gehackt werden, kann dies riesigen Schaden anrichten. Daher befinden wir uns inmitten eines Wettrüstens zwischen KI-Programmen zum Cyberangriff und KI-Programmen zur Cyberabwehr.

Selbst wenn wir autonome KI-Systeme von einem direkten Zugriff auf Nuklearwaffen fern halten, die Entwicklung von autonomen Waffensystemen bedeutet immer, dass Computer in irgendeiner Form Zugriff auf Waffen bekommen. Militärische Computersysteme sind miteinander vernetzt und Waffensysteme werden über Computer ferngesteuert. Diese Systeme sind hoffentlich extrem gut gegen Cyberangriffe geschützt. Aber KI-Programme könnten sich in Zukunft selber programmieren und weiterentwickeln und dadurch immer intelligenter werden. Daher warnen manche KI-Expertinnen und Experten lautstark vor der theoretischen Möglichkeit, dass ein zukünftiges superintelligentes Computerprogramm uns die Kontrolle über unsere Waffen entreißen könnte. Selbst wenn wir einen Notausschalter einbauen, so argumentieren sie, könnte das KI-Programm uns austricksen und alle unsere Sicherheitsvorkehrungen umgehen.

Aus dieser viel beschworenen theoretischen Möglichkeit eines vollkommenen Kontrollverlustes folgt aber keineswegs, dass er eine wirkliche – oder gar die größte – Gefahr darstellt, die von KI ausgeht. Ich

halte KI für ein Sicherheitsrisiko, ich glaube aber nicht, dass ein Terminator-Szenario in absehbarer Zukunft eintreten wird. Trotz aller Fortschritte in der KI-Entwicklung bleibt Skynet momentan reine Science-Fiction. Ohne Skynet kein Terminator-Szenario. Warum warnen einige Kolleginnen und Kollegen also vor einer Auslöschung der Menschheit durch KI? Im Gegensatz zu mir, glauben sie nicht, dass wir uns über die Intelligenz der Maschinen täuschen. Sie denken stattdessen, es ist unausweichlich, dass superintelligente Computerprogramme uns schon bald überflügeln. Aber was soll das überhaupt heißen?