erschienen in der FIfF-Kommunikation,
herausgegeben von FIfF e.V. - ISSN 0938-3476
www.fiff.de

Nico Lück

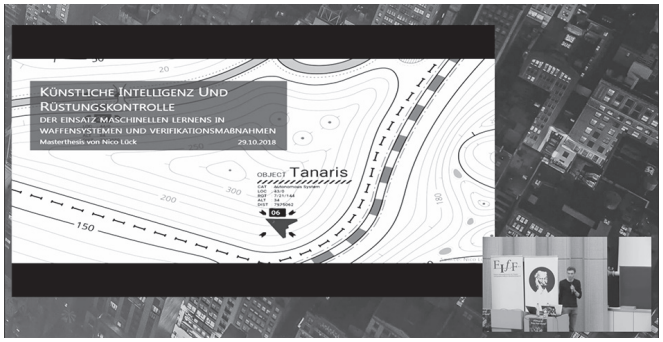
Künstliche Intelligenz in Waffensystemen als Herausforderung für die Rüstungskontrolle

3. Preis

Hinweis: *Dieser Beitrag ist eine stark gekürzte Fassung der Masterarbeit des Autors. Eine ausführliche Fassung mit umfangreichem Literaturkorpus findet sich als eine diesjährige Veröffentlichung durch das Leibniz-Institut Hessische Stiftung Friedens- und Konfliktforschung oder als hinterlegte Masterarbeit an der Goethe-Universität Frankfurt.*

Bei zunehmender Rechenleistung birgt Künstliche Intelligenz (KI) ein großes militärisches Potential. Als rüstungstechnologische Innovation streben Regierungen durch den Einsatz von KI in Waffensystemen eine technologische Überlegenheit und damit einen strategischen Vorteil gegenüber anderen Staaten an. Für die Rüstungskontrolle stellt dieser Umstand ein Risiko dar.

Das Risiko zeigt sich in Waffensystemen, in denen KI absichtlich mit zerstörerischen Aufgaben beauftragt werden kann oder eigenständig einen zerstörerischen Weg zu einem vordefinierten Ziel wählt. Um dieses Risiko zu minimieren, steht Rüstungskontrolle vor der Herausforderung, den Einsatz von KI in Waffensystemen gänzlich zu verbieten. Doch KI entzieht sich den herkömmlichen Ansatzpunkten traditioneller Rüstungskontrolle und minimiert gleichzeitig einen wichtigen deeskalierenden Faktor auf dem Schlachtfeld: die Langsamkeit des Menschen.



Künstliche Intelligenz und maschinelles Lernen

Der Begriff KI ist wegen seiner unpräzisen und tautologischen Definitionen umstritten. Doch für den Zweck dieses Artikels reicht eine Annäherung: Die meisten Definition nennen übereinstimmend zwei Kerneigenschaften (1) die Lösung hoch komplexer Aufgaben und (2) die Anpassungsfähigkeit gegenüber der Umwelt. Besonders die Fähigkeit zu Lernen ermöglicht besonders gute Leistungen in diesen Eigenschaften. Denn in der Vergangenheit interpretierten Computerprogramme Sachverhalte anhand von durch ProgrammiererInnen vorgegebene Regeln. Das sogenannte *maschinelle Lernen* ermöglicht dem Programm hingegen, selbstständig Regeln auf Grundlage der erkannten Muster in Datensätzen zu generieren. Damit können Programme die komplexe Wirklichkeit besser abbilden und sind nicht von vorgegebenen Lösungswegen abhängig. Die dem Menschen überlegene Leistungsfähigkeit gilt jedoch bisher nur für KI, die für spezifische Anwendungen konstruiert wurde (*narrow AI* oder *weak AI*). Eine generell intelligente Maschine, welche den Menschen als Vorbild hat, existiert bisher und wahrscheinlich für die nächsten Jahrzehnte nicht. Dementsprechend geht es in diesem Artikel um die Folgen von anwendungsspezifischer lernender KI für die Rüstungskontrolle. Durch diese Spezifizierung des Begriffs

werden herkömmliche Computerprogramme sowie futuristische KI-Konzepte ausgeschlossen.

Ziele und Maßnahmen der Rüstungskontrolle

Die neusten Entwicklungen im Bereich KI treffen die Rüstungskontrolle in Krisenzeiten: Militär-technologischer Fortschritt, Vertragsbrüche und fehlender politischer Wille lassen das Vertrauen in das multilaterale Instrument sinken. Doch Vertrauen ist ein wichtiger Faktor in einem internationalen System, in der Nationalstaaten immanenten Sicherheitsdilemmata gegenüberstehen: Was schreckt ab? Wie viel ist ausreichend? Was passiert, wenn meine Abschreckung fehlschlägt? Eine Antwort für Regierungen ist es, rüstungstechnologische Innovation zu nutzen, um eine technologische Überlegenheit und damit einen strategischen Vorteil gegenüber anderen Staaten zu erlangen oder auszugleichen. Doch damit dies in keinen Rüstungswettlauf eskaliert, schwächt die Rüstungskontrolle diese Dilemmata, indem sie militärische Kapazitäten und Potentiale reguliert. Diese Regulierungen können auf das Rüstungsvolumen, die geographische Stationierung, den Bereitschaftsgrad oder auch die Nichtverbreitung, Abrüstung oder Abschaffung gewisser Waffenkategorien zielen. Die Ziele der Rüstungskontrolle und der nationalen Militärstrategie sollten damit dieselben sein: Einem gewaltsamen Konflikt ausweichen und eine grundlegende Abschreckung und Sicherheit gewährleisten.

Eine globale Bedeutung erreichte die Rüstungskontrolle erst in der zweiten Hälfte des 19. Jahrhunderts mit dem humanitären Völkerrecht, welches zunächst den Umgang mit Verwundeten, Kriegsgefangenen und der Zivilbevölkerung im gegnerischen Land regelte. Die zahlreichen bi- und multilateralen Rüstungskontrollverträge des 20. Jahrhunderts zeigen, dass konventionelle Waffen, Massenvernichtungswaffen oder selbst die Trägersysteme von Waffen beschränkt wurden und immer noch werden. KI würde sich als Software erstmals in diese Liste einreihen.

Das Kernelement moderner Waffensysteme

Wie bei allen IT-Systemen findet sich auch bei Waffensystemen eine Trennung zwischen Hard- und Software. Zwar verbessern Entwicklungen an der Hardware die physischen Handlungsfähigkeiten und die Feuerkraft von Waffensystemen, doch ist das Entwicklungspotential schlicht durch die Physik begrenzt. Die Verbesserung der Software hat hingegen noch erheblich höheres Potential, da es keine natürlichen Grenzen im Agieren in einer komplexen Umwelt und Reagieren auf den Gegner gibt. Dabei hilft insbesondere die Lernfähigkeit, um eine technologische Überlegenheit und damit eine Überlegenheit auf dem Schlachtfeld zu erlangen.

Anwendungsspezifische KI übernimmt bereits in Waffensystemen als Steuerungs- oder Assistenzeinheit Aufgaben wie Navigation, Zielerkennung und -identifikation sowie Angriffsplanung und -ausführung. Sei es die Drohne *Taranis* des britisches Herstellers *BAE Systems*, welche in einem neu hinzugefügten Flugmodus Routen entwirft und eigenständig nach Zielen sucht, bis das übergeordnete Missionsziel erreicht ist. Oder es sei der Geschützturm *Super aEgis II* des Herstellers *DoDaam Systems* genannt, welcher selbstständig Ziele identifiziert, anvisiert, verfolgt und die Entscheidung zum Feuern trifft. Aufgrund von Befürchtungen der Kunden, das Systeme könne Fehler machen, kann der Grad der Autonomie eingestellt werden. Als letztes Beispiel sei die KI *Alpha* beschrieben, welche bisher nur in Simulationen aktiv gegen US-Air-Force-Piloten angetreten ist und dabei zugleich Geschossen auswich, auf mehrere feindliche Ziele feuerte, sich an koordinierten Manövern mit befreundeten Piloten beteiligte sowie feindliche Taktiken registrierte, lernte und reagierte. Die KI *Alpha* und eine Ankündigung des Rüstungskonzerns *Kalaschnikow*, maschinelles Lernen in Waffensystem nutzen zu wollen, zeigen, dass die Fähigkeit in zukünftige Systeme integriert wird. Ob bereits in operierenden Waffensystemen maschinelles Lernen eingesetzt wird, ist Berichten über die Systeme nicht zu entnehmen. Es zeigt sich aber auch, dass sich die ersten Anwendungen im digitalen Raum, im Luftraum oder in statischen Verteidigungssystemen finden, da die Umgebung, in der die KI operieren muss, vergleichsweise weniger komplex als im Boden- oder Häuserkampf ist. Die Entwicklungspotentiale sollten jedoch nicht darüber hinwegtäuschen, dass KI das befähigende Element dieser Waffensysteme ist und damit der zu regulierende Faktor.

Kontrollerschwerende Eigenschaften und neue Ansatzpunkte

Wenn nun KI als das zu kontrollierende Element in Waffensystemen gilt, dann stellt sich die Frage, welche Eigenschaften von KI für eine qualitative oder quantitative Begrenzung genutzt werden können. Die traditionelle Rüstungskontrolle begrenzt die Anzahl an militärischen Fahrzeugen und Waffensystemen unter anderem aufgrund des äußerlich sichtbaren Erscheinungsbildes. Doch wie bei chemischen, biologischen oder radiologischen Stoffen existiert auch bei Software keine physische Manifestierung. Der Mensch braucht technische Hilfsmittel, um diese Stoffe bzw. die Software wahrzunehmen. Das äußere Erscheinungsbild kann demnach kein Ansatzpunkt sein.

Eine weitere Möglichkeit ist die Beschränkung anhand der inneren Funktionsweise. Antipersonenminen oder Streumunition werden beispielsweise aufgrund der mangelnden Unterscheidung zwischen Kombattanten und Zivilisten verboten. Doch die innere Funktionsweise von KI ist intransparent und nicht nachzuvollziehen. Zwar könnte der Quellcode analysiert werden, doch schon dieser Schritt wird erschwert, da der aus Systemen extrahierte Code in Maschinensprache vorliegen würde und eine Rekonstruktion zu Hochsprache schwierig bis unmöglich wäre. Doch selbst wenn der Quellcode vorliegen würde,

gäbe es bei der Nutzung von maschinellem Lernen die Hürde eines nicht nachvollziehbaren Lernmodells. Die Lernmethode der *Deep Neural Networks* liefert keine Begründung für die berechneten Ergebnisse und die Komplexität des Modells macht es unmöglich, Entscheidungen im Voraus zu determinieren, wie es bei der simplen Funktionsweise von Antipersonenminen oder Streumunition der Fall ist. KI bzw. maschinelles Lernen ist inhärent intransparent und kann daher nicht aufgrund der inneren Funktionsweise beschränkt werden.

Rüstungskontrolle kann auch die äußerlich sichtbaren Fähigkeiten einer Waffe zur Beschränkung nutzen. Beispielsweise verbietet der Kernwaffenteststop-Vertrag, sobald er in Kraft tritt, die Durchführung von Kernwaffenexplosionen für zivile oder militärische Zwecke. An dieser Stelle ist nicht die Funktionsweise oder der Sprengkörper an sich verboten, sondern die Explosionsfähigkeit. Die Fähigkeiten einer KI können allerdings flexibel hinzugefügt oder entfernt werden. Wie bei herkömmlicher Software können über Updates oder eine offene Softwarearchitektur Sicherheitslücken geschlossen oder Funktionen hinzugefügt werden. So ist dies beispielweise schon bei dem F-35 Kampflugzeug des US-amerikanischen Militärs möglich. An das importierte Flugzeug kann die israelische Armee eigene Waffen anbringen und das System über eine App-Integration entsprechend erweitern. Somit könnten bei Inspektionen kritische Funktionen kurzzeitig hinzugefügt oder entfernt werden. Damit entfällt für die Rüstungskontrolle nicht nur ein Ansatzpunkt, sondern es macht auch die Debatte um das Verbot von autonomen Waffen nichtig, wenn Autonomie eine flexibel erweiterbare Fähigkeit ist.

Als letzten Ansatzpunkt für Rüstungskontrolle sei hier die Kontrolle anhand der Äußerlichkeiten des Trägersystems genannt. Es werden beispielsweise nicht die nuklearen Sprengköpfe, sondern die Trägerraketen für kurze und mittlere Reichweite quantitativ begrenzt. Als Kernelement ist KI in diesem Fall dem Sprengkopf gleichzusetzen und damit sind Drohnen, Roboter und andere Systeme das Trägersystem. Durch einheitliche Standards könnte dieselbe KI in unterschiedlichen Waffensystemen die Navigation, Zielerkennung oder Feuerentscheidung übernehmen. Demnach kann der Kern moderner Waffensysteme, KI, zu keinem festen Waffensystem assoziiert werden. Eine Limitierung eines Waffensystems würde nur den Transfer auf ein anderes System bewirken.

Welche Ansatzpunkte bleiben nun übrig? Nachdem Rüstungskontrolle weder an dem äußerlich sichtbaren Erscheinungsbild, der inneren Funktionsweise, den äußerlich sichtbaren Fähigkeiten oder dem Trägersystem ansetzen kann, bleibt nur die Kontrolle in der Entwicklung oder im Einsatz. Für eine Kontrolle in der Entwicklung existiert das bisher nicht umgesetzte Konzept der präventiven Rüstungskontrolle. Sie soll militärisch nutzbare Technologie, Stoffe oder dergleichen bereits bei Entwicklung oder Erprobung verbieten und damit die Umsetzung in Waffensysteme zu verhindern. Denkbar wäre beispielsweise ein Register für militärische Forschung und Entwicklung, welches Risken frühzeitig erkennen lässt. Sollte KI im Einsatz kontrolliert werden, so müssten strategische und taktische Ziele oder Handlungsmöglichkeiten des Systems beschränkt werden. Das von Jürgen Altmann vorgeschlagene Konzept der *Glass Box* könnte hier eine mögliche Option sein. Die Glass Box soll in allen Systemen integriert sein und alle Handlungen und Entscheidungen in einem Protokoll aufzeichnen. Das Protokoll selbst bleibt in Kontrolle des Staates, in dessen Besitz das Waffensystem ist. Doch ein Hash des Protokolls wird zentral bei einer internationalen Organisation gespeichert. Ein Hash ist eine Prüfsumme, die die Integrität bzw. Unveränderbarkeit des Protokolls garantiert. Bei Verdacht auf illegale Kampfhandlungen könnte das Protokoll angefordert werden und durch den Hash wäre erkenntlich, ob das Protokoll manipuliert wurde. Es existieren zwar vereinzelt Lösungsvorschläge von Wissenschaftlern, doch das frühe Stadium der Konzepte zeigt, dass die Akteure in der Rüstungskontrolle die Problematik nicht erkannt haben und in Zukunft vor der Herausforderung stehen werden, alternative Konzepte zu erproben und vor allem politisch durchzusetzen.

Kriseninstabilität

Es wird in drei Stufen unterschieden, in welcher Form ein Mensch in den Entscheidungsprozess einer Waffe eingebunden werden kann: (1) *human in the loop* (semiautonom): Ein Mensch kann am Entscheidungsprozess teilhaben und bspw. dem Waffensystem ein Ziel vorgeben und einen Angriff befehlen. (2) *human on the loop* (autonom unter menschlicher Aufsicht): Das Waffensystem wählt das Ziel und führt einen Angriff eigenständig aus, wird jedoch von einem Menschen überwacht, der jederzeit intervenieren könnte. (3) *human out of the loop* (vollautonom): Das Waffensystem wählt das Ziel und führt einen Angriff eigenständig aus, ohne dass ein Mensch es überwachen oder intervenieren könnte. Eine KI mit hoch entwickelten Wahrnehmungs-, Lern-, und Schlussfolgerungsfähigkeiten kann Stufe 3 erreichen. Doch die schnellen autonomen Entscheidungen gefährden ein Ziel der Rüstungskontrolle – die Stabilität in einem Krisenfall.

Der Mensch ist langsam. Er muss Sprache ausformulieren, Gestik interpretieren oder emotionale Eindrücke verarbeiten. Genau auf diesen menschlichen Faktor zielt die Rüstungskontrolle, um Stabilität im Krisenfall zu verbessern. Beispielsweise dürfen Raketen nicht dauerhaft abschussbereit sein oder Sprengköpfe müssen separat von Raketen gelagert werden. Denn in dieser Zeit bietet sich dem Menschen drei deeskalierende Handlungsmöglichkeiten: (1) Validierung der maschinellen Meldung oder Empfehlung (Es existieren einige Fälle im Kalten Krieg, in denen Menschen technischen Fehlalarm identifizierten und eine Eskalation verhinderten), (2) Kommunikation mit dem Gegner für Verhandlungen oder Klärungen und (3) Abwägung der moralisch und rechtlichen Implikationen. Die Rüstungskontrolle nutzt diesen Faktor nicht nur, sondern stärkt ihn mithilfe von vertrauensbildenden Maßnahmen (z. B. Informationsaustausch, gemeinsame Militärübungen, Austauschprogramme oder direkte Kommunikationskanäle). Doch mit selbstständig agierenden Waffensystemen beschleunigt sich das Geschehen. Der Fall *Flash Crash* aus dem

Jahr 2010 an der New Yorker Börse zeigt dies eindrücklich. Durch Marktmanipulation entstand dort eine Abwärtsspirale von Verkäufen, die von Computerprogrammen in einer Geschwindigkeit getätigt wurden, sodass Menschen nicht rechtzeitig eingreifen konnten. Eine ähnlich gegenseitig eskalierende Spirale kann auch im Zusammentreffen von KI-gesteuerten Waffensystemen geschehen. Vorstellbar ist ein Szenario, in dem zwei Drohnen in Konflikt geraten und selbständig vernetzte Waffensysteme als Unterstützung anfordern. Diese Systeme würden in eskalierende Kampfhandlungen eintreten, noch bevor ein Mensch die Situation erfasst und reagiert hätte.

Dass der Faktor Mensch auf dem Schlachtfeld minimiert wird, ist aus Perspektive der Rüstungskontrolle eine alarmierende Entwicklung. Zwar bieten eigenständig entscheidende Waffensysteme einen strategischen Vorteil auf dem Schlachtfeld, jedoch sollte das Bewusstsein zurückkehren, dass die Langsamkeit des Menschen eine gute Sache sein kann.

Rüstungskontrolle im Zugzwang

Es lässt sich feststellen, dass das Konzept der Rüstungskontrolle in einer Krise steckt und parallel eine neue Technologie aufkommt, deren komplexe und anpassungsfähige Datenanalysen neue technische Potentiale für Waffensysteme eröffnen. Daher wird KI in Zukunft das Kernelement moderner Waffensysteme sein. Sie ermöglicht unter anderem höhere Grade an Autonomie von Waffensystemen und minimiert damit den deeskalierenden Charakter des Menschen. Doch auch wenn auf der multinationalen Ebene der Rüstungskontrolle erkannt wird, dass KI das kontrollwürdige Element von Waffensystemen ist, dann bleibt der Mangel an möglichen Ansatzpunkten zur Regulierung. Da diese Entwicklung bereits heute absehbar ist, müssen Akteure der Rüstungskontrolle aktiv neue Konzept erproben, die die KI-gesteuerten Waffensysteme bereits in der Entwicklung oder im Einsatz kontrollieren.

FIfF e. V. – Rainer Rehak: Laudatio für den Sonderpreis

Jörg Pohle: Datenschutz und Technikgestaltung

Dissertation an der Humboldt-Universität zu Berlin

Kommen wir jetzt zu der letzten Verleihung eines Weizenbaum-Studienpreises für heute. Wie Ihr vielleicht bemerkt habt: Von den vier vergebenen, die eingangs erwähnt worden sind, haben wir jetzt den dritten, den zweiten und dem ersten vergeben. Es wird jetzt noch einer vergeben, und zwar gab es dazu verschiedene Ansichten innerhalb der Jury, so dass wir uns entschieden haben, dass wir diese Arbeit auszeichnen, einfach ohne Zahl.

Es geht um eine Dissertation mit dem Thema *Datenschutz und Technikgestaltung. Die Geschichte und Theorie des Datenschutzes aus informatischer Sicht und Folgerungen für die Technikgestaltung* im Fach Informatik an der Humboldt-Universität zu Berlin, an den damaligen Diplom-Informatiker Jörg Pohle; mittlerweile Dr. Jörg Pohle.

Die Art der Arbeit spiegelte sich nach unserer Ansicht gut in unserer internen Diskussion wider, und zwar hat sich das kondensieren lassen auf die Frage: Was ist Informatik? Wo verlaufen die Grenzen? Bezeichnet man Informatik auf der einen Seite eher als technischen Ansatz in der formalisierten Welt, oder auf der anderen Seite, wie zum Beispiel Wolfgang Coy für eine Theorie der Informatik auch proklamiert hat, dass Informatik sich auch weit über die formalen Bestandteile ihrer selbst bewegen muss, sonst kann es auch keine ernst zu nehmende Informatik sein. Genau diese Diskussion hatten wir auch, bei der Diskussion über die Arbeit von Jörg Pohle.

Gerade Datenschutz und Technikgestaltung berühren viele Disziplinen und haben viele Elemente, von Rechtswissenschaften über sozialwissenschaftliche Ansichten hinaus, natürlich zu den informatischen Aspekten, und diese Arbeit an sich ist mit ihren 314 Seiten das sichtbare Ergebnis von sehr viel harter Arbeit, eine immense Arbeit, die sich dadurch auszeichnet, dass sie Diskussionen über den Datenschutz nachzeichnet und Fragen behandelt wie: Was schützt der Datenschutz überhaupt? Was soll er tun, was tut er? Und sie zeichnet auch nach, wie sich Diskussionen über die Zeit hinweg verändert haben, und möchte – so haben wir das gelesen – auch gleichzeitig sagen, dass viele der Diskussionen, die wir heutzutage haben, schon geführt worden sind und – wie man einer süffisanten Fußnote auch entnehmen kann – schon mit Ergebnissen, die heute allerdings wieder vergessen worden sind.

Man könnte auch sagen, die Arbeit, die wir darin auch sehen und so wertvoll bepreisen wollen, ist nachzuzeichnen, wie sich Ideen verändert haben und vielleicht unter anderem Namen noch einmal aufkamen, und dann zu erkennen, das sind gleiche oder ähnliche Überlegungen, wie sie schon einmal getroffen worden sind, in neuem Gewand. Diese Ähnlichkeiten herzustellen, verlangt natürlich einen weiträumigen Überblick über die Materie, und den kann man sowohl der Arbeit als auch dem Literaturverzeichnis ansehen.

In der Diskussion haben wir uns auch entschieden zu sagen, eventuell wird da nicht jede oder jeder Lesende den Überlegungen zustimmen, aber wer sich überhaupt sinnvoll in den Diskussionen um den Datenschutz, Datenschutztheorie, Datenschutzrecht und auch Datensicherheit bewegen oder positionieren will, muss nicht allen Punkte in der Arbeit zustimmen, aber muss sie auf jeden Fall gelesen haben.

Das heißt, an dieser Stelle schon einmal ein Aufruf, wer in diese Richtung denkt und arbeitet – wie immer bei den Weizenbaum-Studienpreisen – ist das eine klare Leseempfehlung.

Das lässt sich auch sehr schön daran ablesen, dass der erste Satz der Zusammenfassung lautet: „Ziel der vorliegenden Arbeit ist es, die historische Konstruktion des Datenschutzproblems, des Datenschutzes als seiner Lösung sowie die Architektur seiner rechtlichen Implementation aufzudecken und einer kritischen Revision aus informatischer Sicht zu unterziehen, um da-