Schattauer

Digital, automatisch – unbehaglich, gefährlich

Korrespondenzadresse
Prof. Dr. Dr. Manfred Spitzer
Universität Ulm
Leimgrubenweg 12–14
87054 Ulm, Deutschland

Bibliografie
Nervenheilkunde 2021; 40: 588–597
DOI 10.1055/a-1467-0665
ISSN 0722-1541
© 2021. Thieme. All rights reserved.
Georg Thieme Verlag KG, Rüdigerstraße 14,
70469 Stuttgart, Germany

Schon mehrfach war an dieser Stelle von Fortschritten in der Informationstechnik die Rede, bei denen es um die Resultate der Anwendung künstlicher (d. h. im Computer simulierter) neuronaler Netzwerke auf reale Probleme ging. Diese Arbeiten sind zuweilen spektakulär und werden nicht zuletzt deswegen in den bedeutendsten wissenschaftlichen Fachblättern publiziert.

Vor einem halben Jahrzehnt beispielsweise hatte ein solches Netzwerk an Beispielen gelernt, das komplexeste Brettspiel der Welt – Go mit seinen 10^{171} möglichen Kombinationen[1] – zu spielen. Es war also kein Programm geschrieben worden, das Go spielen konnte; vielmehr lernte ein Computer, der künstliche Neuronen und Synapsen simulierte (man nannte das System Alpha-Go), durch das Nachspielen publizierter Go-Partien dieses Spiel und gewann danach vier von fünf Spielen gegen Lee Sedol, den damals weltbesten Go-Spieler, wie das Fachblatt *Nature* berichtete [26]. Nur eineinhalb Jahre später erschien – ebenfalls in *Nature* – eine Arbeit zu einem neueren neuronalen Netzwerk (genannt Alpha-Go zero), das gleichsam bei „Null" mit dem Lernen begonnen hatte (daher der Name) und von Anfang an nur gegen sich selbst spielte. Innerhalb von 3 Tagen durchlief es 3000 Jahre Go-Geschichte (d. h. entdeckte beispielsweise Eröffnungen und Endspiele, für deren Entdeckung die Go-spielenden Menschen 3 Jahrtausende gebraucht hatten) und spielte dann 37 Tage weiter gegen sich selbst auf eine Weise, die für Go-Experten immer unverständlicher wurde, d. h. von ihnen nicht mehr nachvollzogen werden konnte. Danach ließ man Alpha-Go zero 100-mal gegen Alpha-Go (also dasjenige neuronale Netzwerk, das zuvor den weltbesten Spieler besiegt hatte) spielen, wobei Alpha-Go zero 100 Mal gewann [26]. Seither versuchen die weltbesten Go-Spieler, von Alpha-Go zero zu lernen, wie man dieses Spiel besser spielt.

Spiele waren jedoch nur der Anfang. Ebenfalls im Jahr 2017 wurde – wieder in *Nature* – publiziert, dass ein mit entsprechend trainierten neuronalen Netzwerken im Internet verbundenes Smartphone Hautkrebs so gut diagnostizieren kann wie ein Dermatologe. Neuronale Netzwerke finden mittlerweile neue Synthesewege in der Chemie, haben das Problem der Faltung von Eiweißkörpern, völlig neue Erkenntnisse in der Radiologie und vieles mehr an Problemen in der Medizin schon gelöst oder sind dabei, wo es ja oft um komplexe Zusammenhänge geht, die man anhand von sehr vielen Daten herausfinden möchte. Dies befeuert seit einigen Jahren gerade in der Medizin eine Goldgräberstimmung mit Hoffnungen auf weitere ungeahnte Fortschritte [1]. Bei all diesen Fortschritten des Einsatzes von automatisch lernenden künstlichen neuronalen Netzen gibt es jedoch immer das gleiche Problem: Nachdem diese sehr große Datenmengen verarbeitet und gelernt haben, können sie etwas – aber niemand weiß, warum. Es bleibt unklar, auf welche Weise sie ihre Kunststücke vollbringen. Denn in ihnen steckt kein Programm, das man versuchen könnte, rational zu rekonstruieren und damit zu verstehen. Vielmehr enthält ein solches trainiertes neuronales Netzwerk neben seiner *allgemeinen* Struktur – meist in Schichten angeordneten Neuronen mit bestimmten Funktionsregeln – aufgrund seines Trainings mit wirklichen Daten Milliarden von einzelnen konkreten Zahlen, welche die Stärke der Milliarden von Verbindungen (Synapsenstärken) zwischen den simulierten Neuronen angeben. In diesem riesigen Zahlenhaufen steckt das „Können" des neuronalen Netzwerks. Zu verstehen gibt es hier nichts. Von „Wissen" kann und sollte man daher nicht sprechen. Wenn Computer lernen und danach etwas können, ist niemand hinterher schlauer.

Gerade einmal 5 Jahre nach dem eingangs erwähnten Go-Turnier gibt es neue Nachrichten zu weiteren von neuronalen Netzwerken vollbrachten Kunststücken, die ahnen lassen, was vielleicht noch alles kommen könnte. Ich rede nicht von mittlerweile erstaunlich guter automatischer Übersetzung von einer Sprache in eine andere, von automatischer Gesichtserkennung und Personenidentifizierung in Videoüberwachungssystemen oder von Alexa, Siri und Cartana, dem auf Zuruf aktiven digitalen Dienstpersonal der Firmen Amazon, Apple und Microsoft in Millionen von Haushalten der westlichen Welt. – An all das haben wir uns längst ebenso gewöhnt wie an die frustranen Kämpfe mit automatisierten Telefonsystemen, mit denen wir am anderen Ende der Leitung beim Wählen einer Kundendienstnummer konfrontiert sind. Die 3 im Folgenden kurz dargestellten neuen Entwicklungen aus völlig verschiedenen Anwendungsgebieten der digitalen künstlichen Intelligenz lösen vielmehr Unbehagen[2] oder gar Angst aus, denn man muss sich an manche Gedanken und Tatsachen erst gewöhnen, und an andere möchte man sich am liebsten gar nicht gewöhnen.

1 Die Anzahl der beim Schach möglichen Züge liegt dagegen bei „nur" 10^{80}, was – um diese Zahl besser einordnen zu können – etwa der Anzahl der Atome im gesamten Universum entspricht.

2 Vgl. [29] für weitere Beispiele solchen digitalen Unbehagens.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Beispiel 1: Raumplanung auf Computerchips

Im Juni 2021 wurde im Fachblatt *Nature* publiziert, dass man bei Google erstmals mit Hilfe von lernenden neuronalen Netzwerken das Design von neuen Computerchips optimieren konnte [23]. Dieses Design erfolgt in 2 Schritten: Zunächst werden die Funktion und die hierfür notwendigen Komponenten in einer Art logischem Schaltplan zu einem Gesamtsystem verknüpft und das Ergebnis festgehalten. Daran schließt sich eine zweite Phase des Chipdesigns an, die „Floorplanning" genannt wird. In diesem Prozess wird auf der Grundlage des fertigen logischen Schaltplans eines neuen Chips das konkrete räumliche Layout der Millionen von Komponenten auf dem Silikonträger festgelegt. Es geht also hierbei um die optimale Lösung des praktischen räumlichen Problems, wo genau man welche Komponenten auf dem Chip platziert. Dieses Problem kennt jeder, allerdings in etwas einfacherer Gestalt: Wie stelle ich die Möbel in meinem Zimmer so auf, dass alles praktisch dort steht, wo es stehen sollte, damit man sich wohlfühlt und nicht ständig irgendetwas im Wege ist? – Daher rührt sein Name: „Raumplanung". Wie eine gut geschnittene und clever möblierte Wohnung wesentlich besser funktioniert als eine, in der auf Funktionalität nicht geachtet wurde, hat ein effizientes Layout großen Einfluss auf die Funktion eines Computerchips wie beispielsweise dessen Geschwindigkeit oder Stromverbrauch. Bislang war diese Raumplanung – wie in der Architektur auch – eine manuelle und sehr zeitaufwändige Arbeit für ganze Teams von Chip-Designern, die zwar bei der Arbeit von Computern unterstützt wurden, aber dennoch hierfür Monate brauchten. Nicht zuletzt deswegen dauerte bislang die Entwicklung einer neuen Generation von Computerchips mindestens 2-3 Jahre.

Am 10. Juni 2021 wurde im Fachblatt *Nature* eine Arbeit publiziert, die beschreibt, wie man das Problem des Chip-Layout-Designs mittels lernender neuronaler Netzwerke gelöst hat. Es gibt, ähnlich wie beim Brettspiel Go, nahezu unendlich viele Möglichkeiten, Funktionseinheiten auf einer kleinen begrenzten Oberfläche anzuordnen. Verglichen mit der Anzahl der Komponenten heutiger Computerchips (Millionen bis Milliarden) und der Tatsache, dass diese Komponenten in komplexer Weise mit Leitungen verbunden sein müssen, die so einfachen wie schwierigen Bedingungen gehorchen müssen (wie beispielsweise der, dass sie sich nicht kreuzen dürfen), nehmen sich die 181 schwarzen und 180 weißen Steine auf einem Brett mit 361 Positionen – ohne jegliche Verbindungskabel! – bescheiden übersichtlich aus: Den Autoren zufolge gibt es bei 1000 Komponenten auf 1000 möglichen Orten „etwa 10^{2500}" Möglichkeiten [23]. Prinzipiell geht es aber um das gleiche Problem: Wo kommt was hin, um eine Menge von festgelegten Randbedingungen am besten zu erfüllen?

Wie schon beim Brettspiel Go ist es unmöglich, alle möglichen Kombinationen zu berechnen: Dafür reichen weder die Atome noch das Alter des gesamten Universums. Man muss das Problem daher in Teilprobleme zerlegen und diese optimieren. Auch das erfordert sehr viel (mathematisches) Geschick, bewirkt jedoch Fortschritte. Ähnlich wie Alpha-Go zero fing man mit einer leeren Fläche an, auf der Komponenten zunächst zufällig platziert wurden. Von einem separaten Algorithmus wurden diese „Lösungen" des Problems der optimalen Platzierung auf Leistung und Effizienz geprüft und diese Bewertung an das erste Netzwerk zurückgemeldet (man spricht von *reinforcement-learning*). So lernte das Netzwerk nach und nach, welche Anordnungen effektiv waren, und baute auf vergangenen Fehlern und Erfolgen auf. Am Anfang produzierte es zunächst miserable Layouts, aber nach Tausenden von (Lern-) Durchgängen wurden seine Vorschläge immer besser und nach weniger als 6 Stunden Lernen waren seine Vorschläge gleich gut wie oder besser als die Lösungen von Menschen.

Der Witz an der Geschichte steht im letzten Satz der Publikation am Ende der „Conclusion", ganz unvermittelt: „Our method has been used in production to design the next generation of Google TPU"[3] [23]. Man muss dazu wissen: Google baut nicht irgendwelche Chips für irgendwelche Computer, sondern entwickelt vor allem Komponenten für die inhouse produzierte Hardware, auf der seit einigen Jahren Spiele und neuronale Netzwerke besonders effektiv laufen: Googles *Tensor Processing Unit (TPU)*. Das Herzstück dieser Einheit, die wie eine Festplatte aussieht und in herkömmlichen PCs in die dafür vorgesehenen Slots eingeschoben werden kann, ist der mit der beschriebenen Technik neu entwickelte Chip. Mit anderen Worten: Neuronale Netzwerke halfen erstmals beim Design von Hardware, auf der man neuronale Netze – noch besser – lernen lassen kann. In einer am 19.6.2021 im *New Scientist* publizierten Nachricht kann man lesen: „Es ist denkbar, dass dieser neue, von der künstlichen Intelligenz entworfene Chip in der Zukunft verwendet wird, um seinen Nachfolger zu entwerfen, und dieser Nachfolger würde wiederum verwendet werden, um seinen eigenen Nachfolger zu entwerfen"[4] [28]. In diesem Bericht wird auch die Senior-Autorin der Arbeit sinngemäß wie folgt zitiert: „Unsere Arbeitsgruppe ist der Meinung, dass derselbe Ansatz mit neuronalen Netzwerken auf weitere zeitaufwändige Phasen des Chipdesigns angewendet werden kann, wodurch sich die gesamte Designzeit von Jahren auf Tage verkürzen ließe. Das Unternehmen (Google) ist bestrebt, schnell zu iterieren, da selbst kleine Verbesserungen bei der Geschwindigkeit oder dem Stromverbrauch bei dem riesigen Ausmaß der globalen Nutzung neuronaler Netzwerke einen enormen Unterschied (bei der Nachhaltigkeit von maschinellem Lernen im Hinblick auf den Klimawandel) bewirken können"[5] [28].

Beispiel 2: Digitaler Debattierer

Im Mai 2021 wurde in *Nature* eine Arbeit publiziert, in der ein debattierender Roboter vorgestellt wurde, der nach Meinung des Publikums einer Debattier-Veranstaltung seine Sache etwa so gut machte wie ein menschlicher Debattier-Experte [27]. Schon vor etwa 3 Jahren stellte Google sein Projekt Duplex vor, einen digitalen Assistenten, der eigenständig telefonieren kann [18]. Man ruft seinen Frisör oder das Restaurant an, und am anderen Ende der Lei-

3 „Unsere Methode wurde für das Design der nächsten Generation der Google TPU bei deren Produktion eingesetzt" (Übersetzung durch den Autor).

4 "It is conceivable that this new AI-designed chip will be used in the future to design its successor, and that successor would in turn be used to design its own replacement."

5 "The team believes that the same neural network approach can be applied to the various other time-consuming stages of chip design, slashing the overall design time from years to days. The company aims to iterate quickly because even small improvements in speed or power consumption can make an enormous difference at the vast scale it operates at."

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

KASTEN 1

Dialogbeispiele mit dem weltweit ersten „Chatbot“ ELIZA

Mensch: „Ich habe ein Problem mit meinem Vater.“
ELIZA: „Erzählen Sie mir mehr über Ihre Familie!“
Diese Sequenz klingt natürlich; die Grenzen der Nutzung einfacher Synonyme zeigt jedoch der folgende Dialog:
Mensch: „Krieg ist der Vater aller Dinge.“
ELIZA: „Erzählen Sie mir mehr über Ihre Familie!“
Wenn das Programm kein Wort im Wörterbuch findet, war es programmiert, z. B. eine der folgenden Antworten zu produzieren:
ELIZA: „Das habe ich noch nicht verstanden, können Sie mir das erklären.“
oder
ELIZA: „Davon verstehe ich nichts, lassen Sie uns über ein anderes Thema reden.“
Falls es mehrfach stecken blieb, konnte es auch neu beginnen, ganz wie man sich das damals bei einem Analytiker vorstellte:
ELIZA: „Erzählen Sie mir von Ihrer Mutter.“

tung ist ein Computer. „Die Vorführung (auf der Entwicklerkonferenz am 8.5.2018) gelang so überzeugend, dass man sich Sorgen machen muss“, schrieb der *Spiegel* einen Tag später [16]. Aber das System funktioniert nur in einem ganz bestimmten Kontext und für ganz bestimmte Aufgaben. Lange zuvor, im Jahr 1966, hatte der deutsch-US-amerikanische Informatiker und Gesellschaftskritiker Joseph Weizenbaum (1923–2008) das Computer-Programm ELIZA veröffentlicht, mit dem er die Verarbeitung natürlicher Sprache durch einen Computer demonstrieren wollte. Eliza, der Prototyp für moderne Chatbots, simulierte das Gespräch mit einem Psychologen (geführt am Rechner über Bildschirm und Tastatur). Das Programm durchsucht mit Hilfe eines strukturierten Wörterbuchs einen eingegebenen Satz nach bedeutungstragenden Wörtern, findet Synonyme und enthält vor allem eine Sammlung von Sätzen zu verschiedenen Themengebieten, bei denen es sich auch um Fragen oder Ermunterungen zum Weiterreden handeln kann. Nicht immer klappte das reibungsfrei (▶ **Kasten 1**).

Weizenbaum war entsetzt darüber, dass viele Menschen nicht bemerkten, dass sie mit einer Maschine kommunizierten und im Dialog intimste Details von sich preisgaben – was ihn zu einem der weltweit bekanntesten Kritiker der gedankenlosen Computergläubigkeit machte. Er hätte sich über die *Nature*-Arbeit zum *Project Debater* wahrscheinlich ziemlich geärgert, die von Wissenschaftlern der Firma IBM an deren Forschungszentrum im israelischen Haifa bereits im Jahr 2012 gestartet und deren Ergebnisse kürzlich publiziert wurden. Das Format einer Debatte bestand in einer vereinfachten Version des britischen parlamentarischen Debattenstils, wie man sie auch von Debattierwettbewerben[6] kennt. Nach einer 4-minütigen Eröffnungsrede von jeder Seite folgen Antworten von bis zu 4 Minuten Länge jeder Seite (Replik), und danach gibt es zum Abschluss eine 2-minütige Zusammenfassung jeder Seite.

Anstatt auf ein Wörterbuch konnte die künstliche Intelligenz von *Project Debater* auf eine nach semantischen Gehalten vorverarbeitete Datenbank von 400 Millionen (!) Zeitungsartikeln zurückgreifen. Argumente für und gegen den zu vertretenen Standpunkt wurden automatisch extrahiert und organisiert; damit wurden selbsttätig Sätze sowie die Argumentationsstruktur der Statements organisiert. So diente diese Suche auch der Vorbereitung der Antworten, die u. a. darin bestand, dass die tatsächlich vom Menschen, gegen den debattiert wurde, gegebene Antwort mit den vom System gelieferten Gegenargumenten abgeglichen wurde, um die beste Replik zu liefern. Project Debater enthielt neben Spracherkennungs- und Transskriptionssoftware, die eine Live-Debatte ermöglichten, vor allem die 4 Komponenten (1) „Argument Mining“ (mit den Unterfunktionen „Claim detection“, „Evidence detection“, „Stance detection“ und „Topic expansion“), (2) „Argument Knowledge-Base“ (mit der Unterfunktion „Argument Knowledge-Base for Rebuttal“), (3) „Rebuttal System“ (mit 6 Unterfunktionen[7]) und (4) „Debate construction“. Diese vierte Komponente wiederum hatte die Unterkomponenten (Funktionen) „Element filtering“, „Repetition removal“, „Clustering“, „Theme extraction“, „Rephrasing“ und „Speech generation“, die im insgesamt 117-seitigen Supplement auf den Seiten 8 bis 43 ausführlich beschrieben werden. Nach jahrelangem Training und Fine-Tuning der Komponenten, Optimierungszyklen und wiederholt ausgetragenen Debatten zu den verschiedensten Themen wurden 3 Debatten öffentlich vor Publikum ausgetragen. Sie behandelten die Themen „öffentliche Finanzierung der Kindergärten – Pro und Kontra“ (▶ **Kasten 2**), „sollten wir mehr Telemedizin verwenden – Pro und Kontra“ sowie „öffentliche Finanzierung der Raumfahrt – Pro und Kontra.“

Ein Problem der Evaluation der Güte von Debattiersoftware ist, dass im Gegensatz zu Spielen bei einer Debatte nicht immer eindeutig klar ist, wer gewonnen hat. Die Autoren lösten dieses Problem durch das Format der öffentlichen Debatten: Vorher und nachher wurden die Zuschauer nach ihrer Meinung zum Thema (Pro oder contra?) befragt. Zudem wurden die Zuschauer danach gefragt, ob sie mehr von *Project Debater* oder mehr vom argumentierenden Menschen gelernt hatten. Im Hinblick auf die Meinungen der Zuschauer, wer gewonnen hatte, lag *Project Debater* nahezu gleichauf mit dem jeweiligen menschlichen Gegenüber. Bei den Antworten auf die Frage, wer informativer redete, lag *Project Debater* jeweils vorne (Suppl. S. 101). In einem Kommentar in der gleichen Ausgabe von *Nature* [1] wird darauf hingewiesen, dass mit der vorliegenden Arbeit ja erst der Anfang gemacht ist. Mit anderen Worten: Die Systeme werden künftig noch besser werden. – „Dann quasseln eben irgendwann neben den Politikern, Moderatoren und Kommentatoren noch einige Computer mit, was soll’s?“, könnte man sagen. Was aber wird sein, wenn die Debatten-AI beispielswei-

6 In Deutschland gibt es seit knapp 20 Jahren den Schülerwettbewerb *Jugend debattiert*, der unter der Schirmherrschaft des Bundespräsidenten durchgeführt und von großen gemeinnützigen Stiftungen (Hertie, Robert Bosch, Heinz Nixdorf, Mercator) unterstützt wird. Es ist das größte privat-öffentlich finanzierte Projekt zur sprachlichen und politischen Bildung in Deutschland.

7 “Key-Concepts Identification“, „Claim Leads and Counter-Evidence Detection“, „ Sentiment Leads“, „Claim Leads from the Argument Knowledge-Base“, „iDebate-Based Rebuttal“ [„a manually curated database of motions, containing a list of arguments for and against each motion“; Suppl., S. 36] und „Detecting Opponent’s Main Themes.”

Spitzer M. Editorial Nervenheilkunde 2021; 40: 588–597 | © 2021. Thieme. All rights reserved.

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

KASTEN 2

Eröffnungsrede

Eröffnungsrede von *Project Debater* bei der am 19. Februar 2019 um die Frage nach der Finanzierung von Kindergärten aus öffentlichen Mitteln (Pro-Position) gegen den Debattier-Experten Harish Natarajan geführten Debatte (Suppl. S. 72–74). Im Supplement sind alle 3 Debatten vollständig nachzulesen. Die Autoren legen Wert auf die Feststellung, dass dieses spezifische Thema nicht zum spezifischen Trainingsmaterial bei der Optimierung der Komponenten gehörte. Der Text wurde absichtlich nicht übersetzt, weil dadurch die sprachlichen Eigenheiten des von der künstlichen Intelligenz produzierten Texts zumindest teilweise verloren gehen würden.

Project Debater – Opening Speech

„Greetings, Harish. I have heard you hold the world record in debate competition wins against humans. But I suspect you've never debated a machine. Welcome to the future.

I will argue that we should subsidize preschools.

We are going to talk about financial issues, but not only about them. In the current status-quo, we accept that the question of subsidies goes beyond money, and touches on social, political and moral issues.

When we subsidize preschools and the like, we are making good use of government money, because they carry benefits for society as a whole. It is our duty to support them . Subsidies are an important policy instrument. They provide governments with the means through which to pursue industrial development and ensure the livelihoods of their citizens.

There are two issues I will elaborate on now. I will start by explaining why preschool is an important investment. I will also say a few words about poverty. And I will end by discussing some other issues that show the positive aspects of preschools.

Regarding investment. Nature-based preschools are powerful interpretive programs as well as lucrative business decisions. As I mentioned, preschool is an important investment. For decades, research has demonstrated that high-quality preschool is one of the best investments of public dollars, resulting in children who fare better on tests and have more successful lives than those without the same access.

Secondly, a few words about poverty. While I can not experience poverty directly, and have no complaints concerning my own standards of living, I still have the following to share. Regarding poverty, research clearly shows that a good preschool can help kids overcome the disadvantages often associated with poverty. The OECD has recommended that governments subsidise pre-primary education to boost performance in poorer areas. A statistical summary of studies from 1960 and 2013 by the National Institute for Early Education Research found that high-quality preschool can create long-term academic and social benefits for individuals and society "far exceeding costs.". The Centers for Disease Control and Prevention reports that universal full-day preschool creates significant economic savings in health care as well as decreased crime, welfare dependence and child abuse [7]. Former prime minister Gough Whitlam said in 1973 that preschool is "the greatest single aid in removing or modifying the inequalities of background, environment, family income or family nationality".

Now to an additional, final issue. A study by the Melbourne institute of applied economic and social research shows that attendance at preschool has a significant positive impact on later NAPLAN outcomes, particularly in the domains of numeracy, reading and spelling. The results of a new study of over 1,000 identical and fraternal twins, published in Psychological Science, a journal of the Association for Psychological Science, confirm that preschool programs are a good idea. Here is a study from New Jersey that is worth noting. In New Jersey, the follow-up to the Abbott Preschool Program study continues to find that high-quality preschool programs increase achievement in language, arts and literacy, math, and science through fourth and fifth grade.

I hope I relayed the message that we should subsidize preschools.

You will possibly hear my opponent talk today about different priorities in subsidies. He might say that subsidies are needed, but not for preschools. I would like to ask you, Mr. Natarajan, if you agree in principle, why don't we examine the evidence and the data and decide accordingly?

Thank you for listening."

se auch die Art, wie wir Menschen ticken, lernt, also beispielsweise Überlegungen dazu anstellt, was der eine denkt, wie der andere denkt (also eine „theory of mind" hat [5, 33]), die impliziten Annahmen bestimmter Standpunkte durch Ausmessung semantischer Abstände von Webseiten mittels Google besser kennt, als das ein Mensch jemals könnte, und beides benutzt, um durch winzig kleine Argumentations- und daran anknüpfende Verhaltensschritte (man spricht auf neudeutsch von „Nudging") sein Gegenüber – von diesem vollkommen unbemerkt – in eine bestimmte Richtung lenkt? Der eingangs erwähnte Fall des Go-Spielens zeigt, dass Maschinen in bestimmten Bereichen schon heute deutlich mehr lernen „können" als Menschen – und dadurch auch Zusammenhänge finden können, die uns Menschen verborgen sind. Das erste ausgeführte Beispiel zeigt, dass eine selbsttätig lernende Maschine dann, wenn sie erst einmal eine Aufgabe so gut bewältigt wie Menschen, sehr rasch besser ist als Menschen. Und wer schützt uns vor Debattierern, die – egal wofür sie argumentieren – nach unserem eigenen Eindruck immer Recht behalten?

Falsche Nachrichten waren schon immer Teil von medialen Manipulationsversuchen [17], Emotionen machen uns nachweislich anfälliger dafür [20], aber erst durch Möglichkeiten der digitalen Kommunikation und insbesondere Social Media sorgt künstliche Intelligenz heute dafür, dass sich falsche Nachrichten rascher und weiter verbreiten als wahre [32]. „Angesichts solcher Fähigkeiten ist es möglich, dass ein Computer irgendwann überzeugende Aussagen mit besseren rhetorischen Fähigkeiten und geschickterer emotionaler Ansteckung produzieren könnte – beides ist bekanntlich effektiver als Tatsachen und Logik –, um Aufmerksamkeit zu erlangen und die Leute von falschen Behauptungen zu überzeu-

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

gen“[8], kann man im Kommentar des Fachblatts *Nature* lesen [1]. Dort wird sogar darauf verwiesen, dass der frühere US-Präsident Donald Trump vorgemacht hat, wie es geht, und dass wir uns davor hüten müssen, dass Maschinen gebaut werden, die dies nachmachen. Da beruhigt es wenig, wenn AI-Spezialisten einschließlich des Erstautors der Arbeit zu *Project Debater* versichern, dass „Menschen immer das Recht haben sollten, zu wissen, ob sie mit einer Maschine interagieren – was natürlich auch das Recht einschließt, zu wissen, ob eine Maschine versucht, sie zu überreden. Ebenso wichtig ist die Sicherstellung, dass die Person oder Organisation hinter der Maschine gefunden und zur Verantwortung gezogen werden kann, falls Menschen zu Schaden kommen“[9].

Auch das, was man sonst noch im *Nature*-Kommentar nachlesen kann, bereitet mir großes Unbehagen: „Sofern sich diese Forschung weiterentwickelt, besteht auch die dringende Notwendigkeit von Richtlinien, wenn nicht sogar Vorschriften, zur Transparenz in der künstlichen Intelligenz – damit die Menschen wenigstens wissen, ob sie mit Menschen oder Maschinen zu tun haben. Noch mehr Aufsicht muss sein, wenn computerisierte Debattierer eines Tages manipulative Fähigkeiten entwickeln. […] Ebenso wird künstliche Intelligenz sicherlich für Unternehmen attraktiv sein, die Menschen zum Kauf ihrer Produkte überreden wollen. Das ist ein weiterer Grund, notfalls durch Regulierung einen Weg zu finden, um Transparenz zu gewährleisten und potenzielle Schäden zu reduzieren. Man könnte Algorithmen auch ähnlichen Studien unterziehen, wie sie für neue Medikamente üblich sind, bevor sie für den öffentlichen Gebrauch zugelassen werden können. […] Die Regierung wird bereits unterminiert, wenn Politiker auf überzeugende, aber unehrliche Argumente zurückgreifen. Noch schlimmer wäre es, wenn der Sieg bei den nächsten Wahlen davon abhängt, wer den besten Algorithmus hat“[10] [1]. Das schreiben die für freien Markt sehr offenen und gegenüber staatlichen Eingriffen sehr skeptischen Briten über die Gefahren durch heutige Politiker und künftige debattierende Computer. Sie müssen es wissen, denn sie haben das Parlament und Debatten erfunden.

Beispiel 3: Autonome Waffen

Auf der Webseite des US-amerikanischen Militär-Forschungsinstituts *Defense Advanced Research Projects Agency* (DARPA) findet sich unter „Our Research“ (Unsere Forschung) eine sehr kurze Beschreibung des Projekts *System-of-Systems Enhanced Small Unit* (SESU) (zu Deutsch etwa „durch ein System von Systemen verstärkte kleine Einheit“). Sie lautet wie folgt: „Zukünftige US-Landstreitkräfte werden immer wahrscheinlicher mit einem Gegner konfrontiert sein, der in Größe und Bewaffnung überwältigend überlegen ist und über gewaltige Anti-Access/Area Denial (A2/AD) Fähigkeiten verfügt. SESU zielt darauf ab, System-of-Systems-Fähigkeiten (SoS) bereitzustellen, die es einer kleinen Einheit (etwa 200–300 Soldaten samt der entsprechenden materiellen Ausstattung und begrenzter Rückendeckung) ermöglichen, die gegnerischen A2/AD- und Manövrierfähigkeiten zu zerstören, zu stören, zu beeinträchtigen bzw. zu verzögern, um gemeinsam mit anderen Militäreinheiten (Luftwaffe, Marine) Multidomänen-Operationen zu jeder Zeit und an jedem Ort zu ermöglichen. Das SESU-Programm wird sich auf 2 technische Bereiche konzentrieren: 1) Adaptive Führung, die eine verteilte Planung der Militärmission und Anti-A2/AD-Missionen ermöglicht, indem sie eine große Anzahl kostengünstiger, autonomer, unbemannter Luft- und Bodenplattformen nutzt, die mit geeigneten Sensoren und Effektoren ausgestattet sind; und 2) Innovative Sensoren und Effektoren, die zusammenarbeiten können, um Aspekte der A2/AD-Fähigkeiten eines Gegners zu zerstören, zu stören, zu beeinträchtigen und/oder zu verzögern“[11] [34].

Was hier teilweise etwas kryptisch formuliert ist – „kostengünstige, autonome, unbemannte Luft- und Bodenplattformen“ bedeutet im Klartext „Drohnen und Kampfroboter“ – ist die Beschreibung künftiger Kriegführung: Um selbst bei Übermacht und Lufthoheit des Gegners zu gewinnen, spielen Soldaten nicht nur beim Schießen, sondern auch beim Planen und Entscheiden nur noch eine untergeordnete Rolle; denn „adaptive Führung“ und „verteilte Planung“ bedeutet, dass die Kampfmaschinen nicht ferngesteuert operieren, sondern autonom und damit im konkreten Kampf selbst mittels künstlicher Intelligenz Entscheidungen treffen – von taktischen „Spielzügen“ bis zum „Schlagen“ gegnerischer Figuren, um es vorsichtig mit dem Vokabular des Schachspiels auszudrü-

8 “Given such capabilities, it’s possible that a computer might one day create persuasive language with stronger oratorical ability and recourse to emotive appeals – both of which are known to be more effective than facts and logic in gaining attention and winning converts, especially for false claims.”

9 “[…] humans should always have the right to know whether they are interacting with a machine – which would surely include the right to know whether a machine is seeking to persuade them. It is equally important to make sure that the person or organization behind the machine can be traced and held responsible in the event that people are harmed.“ Im Hinblick auf den Erstautor der Studie fährt der Text wie folgt fort: „Project Debater’s principal investigator, Noam Slonim, says that IBM implements a policy of transparency for its AI research, for example making the training data and algorithms openly available.”

10 Auch dieses längere Zitat sei dem anglophilen Leser im Original nicht vorenthalten: ”As this research develops, it’s also a reminder of the urgent need for guidelines, if not regulations, on transparency in AI – at the very least, so that people know whether they are interacting with a human or a machine. AI debaters might one day develop manipulative skills, further strengthening the need for oversight. […] Equally, AI will surely look attractive to those companies looking to persuade people to buy their products. This is another reason to find a way, through regulation if necessary, to ensure transparency and reduce potential harms. AI algorithms could also be required to undergo trials akin to those required for new drugs, before they can be approved for public use. Government is already undermined when politicians resort to compelling but dishonest arguments. It could be worse still if victory at the polls is influenced by who has the best algorithm.”

11 „Future U. S. land forces are increasingly likely to face an adversary force that is overwhelmingly superior in size and armament with formidable anti-access/area denial (A2/AD) capabilities. SESU seeks to deliver system-of-systems (SoS) capabilities that could enable a small unit (~200–300 soldiers, corresponding materiel footprint, and limited rear-echelon support) to destroy, disrupt, degrade, and/or delay the adversary’s A2/AD and maneuver capabilities in order to enable joint and coalition multi-domain operations at appropriate times and locations. The SESU program will focus on two technical areas: 1) Adaptive command and control (C2) that enables distributed mission planning and counter-A2/AD missions leveraging a large number of low-cost, autonomous, unmanned air and ground platforms equipped with appropriate sensors and effectors; and 2) Innovative sensors and effectors that can cooperate to destroy, disrupt, degrade, and/or delay aspects of an adversary’s A2/AD capabilities.“

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

cken. Gelernt wird das wie unter den Beispielen 1 und 2 beschrieben durch vielfaches Durchspielen von Militärmissionen, wobei die Systeme (ähnlich wie Alpha-Go zero, der Chip-Raumplaner oder der Debattier-Roboter) durch Erfahrung selbsttätig lernen. Nach einem im Fachblatt *New Scientist* publizierten Bericht beschreibt ein Wissenschaftler der US Army die Resultate wie folgt: „Es ist sehr interessant zu beobachten, wie die künstliche Intelligenz von sich aus einige sehr knifflige und interessante Taktiken entdeckt [...] Man denkt oft: ‚Oh wow, das ist ziemlich clever. Wie hat die KI das herausgefunden?'"[12] [12] – Man fühlt sich unweigerlich an den Spielzug 37 von Alpha-Go im zweiten Spiel beim Go-Turnier gegen Lee Sedol im März 2016 erinnert, der zunächst als „Fehler" betrachtet wurde, nach dem von Alpha-Go gewonnenen Spiel von Kommentatoren jedoch als „besonders kreativ", „völlig unvorhergesehen" und „einfach genial" bezeichnet wurde [21].

Ergänzt wird das Forschungsprojekt SESU durch das DARPA-Forschungsprojekt OFFSET *(OFFensive Swarm-Enabled Tactics;* zu Deutsch etwa „Offensive schwarmgestützte Taktiken"), bei dem es um die Entwicklung von Systemen von mehr als 250 Drohnen und am Boden operierenden Kampfrobotern geht, „um verschiedene Missionen in komplexen städtischen Umgebungen zu erfüllen. Durch die Nutzung und Kombination neuer Technologien im Bereich der Schwarmautonomie und der Zusammenarbeit zwischen Mensch und Schwarm soll das Programm die schnelle Entwicklung und den Einsatz bahnbrechender Fähigkeiten ermöglichen. (Das Programm) zielt darauf ab, die Werkzeuge zur Verfügung zu stellen, um schnell Schwarmtaktiken zu generieren, diese Schwarmtaktiken auf ihre Effektivität hin zu evaluieren und die besten Schwarmtaktiken in Feldoperationen zu integrieren. Um diese Ziele zu erreichen, wird OFFSET ein aktives Ökosystem für die Entwicklung von Schwarmtaktiken und eine unterstützende offene Systemarchitektur entwickeln" (▶ **Kasten 3**).

Im Rahmen dieses Projekts sollen „ungefähr alle 6 Monate [...] Live-Experimente mit verschiedenen unbemannten Luft- und Bodenplattformen durchgeführt werden." Dabei soll die Komplexität sowohl der Schwarmarchitektur als auch der Schwarmtaktik zunehmen, was über immer neue Variationen „technologischer und operativer Testvariablen [...] wie z. B. Schwarmgröße, räumlicher Umfang der Operationen und Missionsdauer" erfolgen soll. Die Benutzer würden die Schwarmschnittstelle nutzen, um die besten der virtuellen Taktiken in der realen Welt zu testen und ihre unbemannten Plattformen interaktiv und nahezu in Echtzeit mit Taktik-Updates zu versorgen" [4]. Man könnte das Ganze zunächst für Science Fiktion halten. Aber es gibt dies alles wirklich! Und obgleich in Europa durchaus kritische Stimmen gegenüber autonomen Waffensystem erhoben wurden, hat der Philosoph Thomas Metzinger in seiner Funktion als Mitglied der EU-*Kommission Ethik Richtlinien für Künstliche Intelligenz* in einem in der Zeitschrift *Forschung und Lehre* erschienenen Interview sehr klar herausgestellt, dass die Lobby stärker war als die Vernunft: „Das Europäische Parlament hat zwar deutlich gemacht, dass Europa das nicht will, aber im Gesetzentwurf wurde das Thema ganz lapidar mit einem Satz behandelt" [22].

12 „It's very interesting to watch how the AI discovers, on its own, some very tricky and interesting tactics [...] Often you say, 'Oh whoa, that's pretty smart. How did it figure out that one?'"

KASTEN 3

Systemarchitektur

Systemarchitektur des DARPA-*Projekts OFFSET* (zitiert aus [4]), dem Autor der entsprechenden Webseite der DARPA. Auch dieser Text wurde absichtlich nicht übersetzt, um die Ähnlichkeit der Diktion mit Texten über Computerspiele zur Freizeitgestaltung für Jugendliche (z. B. in der Werbung) deutlich werden zu lassen.

Das System „Offensive Schwarmgestützte Taktiken" soll die folgenden 3 Komponenten enthalten:

„• An advanced human-swarm interface to enable users to monitor and direct potentially hundreds of unmanned platforms simultaneously in real time. The program intends to leverage rapidly emerging immersive and intuitive interactive technologies (e. g., augmented and virtual reality, voice-, gesture-, and touch-based) to create a novel command interface with immersive situational awareness and decision presentation capabilities. The interface would also incorporate a swarm interaction grammar, enabling "freestyle" design of swarm tactics that allow dynamic action and reaction based on real-time conditions in the field.

• A real-time, networked virtual environment that would support a physics-based, swarm tactics game. In the game, players would use the interface to rapidly explore, evolve, and evaluate swarm tactics to see which would potentially work best, whether using various unmanned systems in the real world or exploring innovative synthetic technologies in the virtual one. Users could submit swarm tactics and track their performance from test rounds on a leaderboard, to encourage both competition and collaboration.

• A community-driven swarm tactics exchange. This curated, limited access program portal would house tools to help participants design swarm tactics by composing collective behaviors, swarm algorithms, and even other existing tactics. It would provide these key ingredients to an extensible architecture for end-user-generated swarm tactics and help create a lasting community to innovate and cultivate the most effective tactics, with the potential to integrate third-party tactics and future users."

Das Problem stellt sich als systematisch sehr schwierig heraus, wie kurz gezeigt sein soll. Zunächst einmal ist auch und gerade der Krieg eine Sache des Menschen, wie Robert Bunker in einer vom ihm für die Vereinigung der US-Armee verfassten und als „eher explorativ als autoritativ" (S. 2) bezeichneten Arbeit vom Mai 2020 gleich zu Beginn feststellt. „Richtig verstandene militärische Führung befähigt zur dezentralen Entscheidungsfindung auf untergeordneter Ebene, um die Ausführung der Operation entsprechend der jeweiligen Situation zu gewährleisten. Sie ist ausschließlich für die Führung und Kontrolle von Mensch zu Mensch gedacht. Wie der Krieg selbst, ist es ein inhärent „menschliches Unterfangen ... kein mechanischer Prozess, der von Maschinen (oder) Berechnungen präzise gesteuert werden kann." [7]. Systeme, die für ihre Entscheidungsprozesse maschinelle Algorithmen verwenden, stehen in direktem Widerspruch zu den emotionalen und moralischen Komponenten des menschlichen Denkens. Menschen empfinden

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

Liebe, Angst, Kameradschaft und Hass – Maschinen tun das nicht. Ebenso wenig verstehen sie Ehre, Integrität oder Selbstaufopferung. Wie lässt sich angesichts dieses Konflikts der Einsatz von autonomen Maschinen mit den Prinzipien militärischer Führung vereinbaren?“[13] ([3] frei übersetzt vom Autor).

Unter expliziter Berufung auf die vor über 150 Jahren von der preußischen Armee eingeführte, sehr effiziente dezentrale Führungsstrukturen und auf die Geschwindigkeit des deutschen Blitzkriegs gegen Frankreich im Jahr 1940, geht Bunker dann der Frage nach, wie diese Form der Führung mit der Beteiligung künstlicher Intelligenz an Entscheidungsprozessen funktionieren kann. Sein Ergebnis: Weil Menschen wesentlich langsamer sind als Maschinen, wird immer derjenige gewinnen, der sich bei der Kontrolle der Maschinen am meisten zurückhält. Anders gewendet: Autonome Waffensysteme funktionieren am besten, wenn der Mensch sich nicht einmischt und damit die Kontrolle – über Leben und Tod – den Maschinen überlässt. „Wird es versäumt, die mit den Fortschritten in der Militärtechnologie und den sie unterstützenden Betriebskonzeptmodellen – vor allem C2 (Steuerung und Kontrolle) – notwendigen Veränderungen zu implementieren, sind dieselben Ergebnisse zu erwarten, wie als die Ritterschaft unter dem Sperrfeuer von Musketenschüssen starb, vorrückende Wellen von Infanterie durch Maschinengewehrfeuer niedergemäht wurden oder moderne gepanzerte und mechanisierte Einsatzkräfte zunehmend durch präzisionsgelenkte Munition dezimiert werden. Dies wäre ein katastrophales Schicksal – eines, das US-Soldaten niemals erleben dürfen,“[14] sind die letzten Sätze in der Schlussfolgerung von Bunker [3]. “You’re damned if you use them, and damned if you don’t,” wird Bunker kommentierend im *New Scientist* zitiert [10].

Im August 2020 wurde bekannt, dass ein System mit künstlicher Intelligenz einen erfahrenen F-16-Kampfpiloten in 5 simulierten Luftkämpfen (Dogfights) besiegte. Mittlerweile wurde klar, dass die US-Armee gar nicht daran denkt, irgendetwas zu versäumen. Im Rahmen des beschriebenen Forschungsprogramms SESU wurden Kämpfe zwischen Einheiten von ein paar hundert Soldaten simuliert, die mit Schwärmen von autonomen Drohnen und Robotern zusammenarbeiten. Dabei ergab sich empirisch, dass die Tendenz des Menschen, sich in den Entscheidungsprozess der Roboter einzumischen, die Einheit deutlich verlangsamte. „Wenn man ein Bild des Ziels übertragen muss, den Menschen darauf schauen lassen und darauf warten muss, dass der Mensch den „Feuer“-Knopf drückt, […] ist das bei Maschinengeschwindigkeit eine Ewigkeit. […] Wenn wir die KI auf menschliche Geschwindigkeit verlangsamen, werden wir verlieren“, berichtete Sydney Freedberg Jr. im Herbst 2020 auf der *Army Futures Command Konferenz* [19].

Hinzu kommt, dass die Kommunikationsverbindung zwischen allen Beteiligten nie ausfallen darf, was unter Gefechtsbedingungen schwer zu garantieren ist. „Die Übertragung von Videos erfordert eine drahtlose Verbindung mit hoher Bandbreite und großer Reichweite, die schon schwer aufrechtzuerhalten ist, wenn Sie zu Hause auf Zoom sind, geschweige denn auf dem Schlachtfeld, wo der Feind Ihre Signale stört, ihr Netzwerk hackt und jeden Sender bombardiert, den er aufspüren kann“, kann man in Freedbergs Bericht lesen. Wenn die Soldaten oder die Maschinen also immer erst fragen müssen, bevor sie feuern, hat man die Schlacht schon verloren. Das Argument der Militärs für eine Kontrolle durch künstliche Intelligenz lautet ganz allgemein wie folgt: „Wenn KI-gesteuerte Waffen die von Menschen gesteuerten Waffen besiegen können, dann wird derjenige, der die KI hat, gewinnen, und wer sie nicht einsetzt, hat die Niederlage schon akzeptiert“ [13].

Zudem ergaben Millionen von Simulationen mit autonomen Waffensystemen, dass diese – wie Alpha-Go zero – lernen können. Hätte man beim Lernen von Alpha-Go zero nur Spielzüge erlaubt, die Menschen nachvollziehen können, wäre das System nie schlauer geworden als Menschen. Gerade dadurch, dass man den Menschen aus dem Lernen der Maschine herausgehalten hatte, fand man kreative Lösungen durch die Maschine, auf die Menschen noch nicht gekommen waren, weil die Zeit eines Menschenlebens begrenzt ist. Spielt man also nur lange genug „War Games“ durch, dann erfinden autonome Systeme selbstständig völlig neue Taktiken – mit dem oben bereits beschriebenen „Wow-Effekt“. Das Problem ist jedoch dann das Gleiche wie beim Go: Man *kann* dann zwar besser Krieg führen, *weiß* aber nicht warum. Dieses Problem der fehlenden Erklärbarkeit von Taktiken plagt mittlerweile nicht nur die Gemeinde der Go-Spieler, sondern auch weltweit die Militärs, sofern sie sich autonomer Waffen bedienen möchten. Die einen sagen nun, dass man das Problem nicht zu ernst nehmen sollte, weil wir schon heute in vielen Fällen künstlicher Intelligenz vertrauen, ohne sie zu verstehen. Die anderen halten dagegen, dass man sich kaum an den Gedanken gewöhnen kann, dass eine Maschine über Leben und Tod entscheidet, deren Entscheidungsprozesse wir – ganz *prinzipiell* – nicht nachvollziehen können. Der US-General John Murray, Leiter des *Army Futures Command*, sagte in einem Webinar im *Center for Strategic & International Studies*, dass Menschen zu langsam und daher nicht in der Lage sind, Schwärme von feindlichen Drohnen zu bekämpfen. Schon deshalb müsse man künstliche Intelligenz anstatt menschlicher Kontrolle bei solchen Angriffen nutzen.

Was aber geschieht, wenn – wie die USA ja gerade selber simulieren und optimieren – Drohnen und Kampfroboter zusammen mit Soldaten angreifen? Kann, darf oder muss man da die menschliche Kontrolle der Verteidigung an die künstliche Intelligenz abgeben? „Die Regeln für die menschliche Kontrolle über künstliche Intelligenz müssen möglicherweise gelockert werden“, sagte der US-General [13]. Vielleicht braucht man auch gar nicht mehr über all das nachzudenken, weil die Realität schon weiter ist. Im vergangenen August wurden in den USA (60 km südlich von Seattle) Drohnen

13 “Mission command, properly understood, empowers subordinate decisionmaking and decentralized execution appropriate to any given situation. It is solely meant for human-to-human command and control. Like war itself, it is an inherently “human endeavor … not a mechanical process that can be precisely controlled by machines [or] calculations”. Systems that use machine algorithms for their decisionmaking processes are in direct variance to the emotive- and moral-seeking components of human cognition. Humans experience love, fear, camaraderie and hate – machines do not. Nor do they understand honor, integrity or self-sacrifice. Faced with this conflict, how can the deployment of machines work in concert with the Army’s command and control?“

14 “Failure to change with advances in military technology and the CONOPS [concept of operations] supporting them – most importantly C2 [command and control] – will yield the same result as the knighthood of chivalry dying under barrages of musket shot, advancing waves of infantry being cut down by machine gun fire and contemporary armored and mechanized forces being increasingly decimated by precision-guided munitions. This would be a catastrophic fate – one that U. S. soldiers must never be allowed to experience.”

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

und Kampfroboter in einem Manöver eingesetzt, in dem es darum ging, in Häusern versteckte Terroristen aufzufinden [15]. Bereits im Jahr 2019 wurde bekannt, dass die Türkei im Syrien-Konflikt Drohnen einsetzt, die zunächst manuell gesteuert wurden, um eine Sprengladung im Ziel zur Explosion zu bringen, und diese Drohnen zu autonomen Waffensystemen weiterentwickelt [10]. Nach einem Bericht des Sicherheitsrats der Vereinten Nationen (UN) haben militärische Drohnen bereits im vergangenen Jahr in Libyen erstmals Menschen angegriffen. In dem 548-seitigen Report vom März 2021 kann man auf S. 17 nachlesen [31]: „Logistikkonvois und sich zurückziehende Haftar Affiliated Forces (HAF) wurden anschließend von unbemannten Kampfflugzeugen oder den tödlichen autonomen Waffensystemen wie die STM Kargu-2 Drohne gejagt [...] Diese Waffensysteme waren so programmiert, dass sie Ziele angreifen, ohne dass eine Datenverbindung zwischen Bediener und Munition erforderlich ist.“[15]

Schließlich nutzte Israel (Israel Defense Forces, IDF) Mitte Mai 2021 erstmals einen ganzen Schwarm von Drohnen, gesteuert von KI und einem einzigen Menschen, neben dem noch ein Kommandeur stand, um wichtige Entscheidungen zu fällen. Die in Israel produzierte Thor-Drohne wiegt 9 kg, fliegt nahezu lautlos und wurde in Kombination mit anderen Drohnen vor allem zur Aufklärung und Zielidentifikation eingesetzt. Jede einzelne Drohne fliegt weitgehend autonom und der Schwarm, der als Gesamtsystem operiert, funktioniert auch dann noch, wenn einzelne Mitglieder ausgefallen sind. Der Einsatz des Schwarms wurde von (menschlichen) Bodentruppen unterstützt und richtete sich – offenbar sehr erfolgreich – gegen die Raketenangriffe der Hamas aus dem Gaza-Streifen. „Die Berichte deuten darauf hin, dass die IDF das erste Militär sind, das über einen einsatzfähigen Drohnenschwarm verfügt“, wird der Analyst Zak Kallenborn vom *Nationalen Konsortium für die Untersuchung von Terrorismus und Terrorismus-Abwehr* in Maryland vom *New Scientist* zitiert[16] [14]. Das israelische Militär plant die Ausrüstung weiterer Infanterie-Einheiten mit Drohnenschwärmen, wie ein Sprecher des IDF bekannt gab [14]. Es ist daher kaum anzunehmen, dass die Militärs weltweit freiwillig auf die neuen Fähigkeiten verzichten.

Die 3 Beispiele zeigen, dass die Fortschritte der digitalen Informationstechnik uns in jedem Fall zu denken geben sollten. Es sollte eigentlich eine breite Diskussion über deren Risiken und Nebenwirkungen in den unterschiedlichsten Anwendungsbereichen geben, aber selbst in der Europäischen Union scheint man sich eher wegzuducken als diesen Ball aufzunehmen. Das dürfen wir nicht tun. Maschinen, die sich selbst immer schneller immer besser machen, sich also selbst optimieren, bereiten Kopfschmerzen; Debattier-Roboter sicherlich auch, aber wenn wir nicht aufpassen, gefährden sie zudem die Demokratie.

Atomare, biologisch und chemische Waffensysteme (ABC-Waffen) gibt es seit Jahrzehnten, sie sind jedoch geächtet und Verträge sollen dafür sorgen, dass sie niemals eingesetzt werden. Autonome Drohnen gibt es erst seit wenigen Jahren. Sie funktionieren nur dann wirklich gut, wenn sie autonom (also ohne Eingreifen des Menschen) und damit sehr schnell über Leben und Tod entscheiden. Dies tun sie nach Regeln, die Menschen ganz prinzipiell nicht mehr nachvollziehen können (und genau deswegen funktionieren sie so gut), weil es bei Milliarden von Zahlen, die Synapsengewichte respräsentieren, nichts nachzuvollziehen gibt. Wie unter Beispiel 3 erwähnt, betrachtet das Militär selbst seine Entscheidungen als von bestimmten Werten (Liebe, Angst, Kameradschaft, Hass, Ehre, Integrität und Selbstaufopferung [3]) geprägt, die mit der spezifisch menschlichen Sichtweise auf die Welt untrennbar verknüpft sind. Dies mit prinzipieller Notwendigkeit zu übergehen, nur damit das Militär neue Waffensysteme verwenden kann, sollte man sich sehr genau überlegen. Und es sollte vor allem von allen darüber diskutiert werden.

15 “Logistics convoys and retreating HAF were subsequently hunted down and remotely engaged by the unmanned combat aerial vehicles or the lethal autonomous weapons systems such as the STM *Kargu*-2 (see annex 30) and other loitering munitions. The lethal autonomous weapons systems were programmed to attack targets without requiring data connectivity between the operator and the munition” [31].

16 “The reports suggest the IDF is the first military to field an operational drone swarm capability,” says analyst Zak Kallenborn at the *National Consortium for the Study of Terrorism and Responses* to Terrorism in Maryland.

Literatur

[1] Anonymus. Rise of AI debaters highlights need for transparency. Nature 2021; 592: 166

[2] Anonymus. The shape of things to come. DeepMind’s latest AI could spark a medical revolution. New Scientist 2020; 248 (3311): 5

[3] Bunker RJ. Mission Command and Armed Robotic Systems Command and Control: A Human and Machine Assessment. The Association of the United States Army (AUSA). Land Warfare Paper 2020; 132

[4] Chung T. OFFensive Swarm-Enabled Tactics (OFFSET). Defense Advanced Research Projects Agency (DARPA) 2021 https://www.darpa.mil/program/offensive-swarm-enabled-tactics

[5] Cominelli L, Mazzei D, De Rossi DE. SEAI: Social Emotional Artificial Intelligence Based on Damasio’s Theory of Mind. Front Robot AI 2018; 5: 6. doi: 10.3389/frobt.2018.00006

[6] Crane L, Wilson C, Whyte C et al. The essential guide to the algorithms that run yourlife. New Scientist 2021; 3339: 34–39

[7] Department of the Army (DA), Army Doctrine Publication (ADP) 6–0, *Mission Command: Command and Control of Army Forces*. Washington, DC: U. S. Government Printing Office, July 2019, 1–1

[8] Esteva A, Kuprel B, Novoa RA et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 547: 115–118

[9] Freedberg SJ Jr. Let Your Robots Off The Leash – Or Lose. Breaking Defense, 19.11.2020 https://breakingdefense.com/2020/11/let-your-robots-off-the-leash-or-lose-ai-experts/; abgerufen am 4.7.2021

[10] Hambling D. Autonomous killer drones set to be used by Turkey in Syria. New Scientist 2019; 3249 https://www.newscientist.com/article/2217171-autonomous-killer-drones-set-to-be-used-by-turkey-in-syria/#ixzz6ziY6fPAF; abgerufen am 1.7.2021

[11] Hambling D. Military robots perform worse when humans won’t stop interrupting them. New Scientist 4.12.2020 https://www.newscientist.com/article/2261842-military-robots-perform-worse-when-humans-wont-stop-interrupting-them/; abgerufen am 1.7.

[12] Hambling D. Drone Swarms Are Getting Too Fast For Humans To Fight, U. S. General Warns. Forbes 2021 https://www.forbes.com/sites/davidhambling/2021/01/27/drone-swarms-are-getting-too-fast-for-humans-too-fight-us-general-warns/?sh=4b33e39e372c; abgerufen am 1.7.2021

[13] Hambling D. Israel used world’s first AI-guided combat drone swarm in Gaza attacks. New Scientist 2021 https://www.newscientist.com/

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.

article/2282656-israel-used-worlds-first-ai-guided-combat-drone-swarm-in-gaza-attacks/; abgerufen am 1.7.2021

[14] Hambling D. Drones may have attacked humans fully autonomously for the first time. New Scientist 2021 https://www.newscientist.com/article/2278852-drones-may-have-attacked-humans-fully-autonomously-for-the-first-time/; abgerufen am 1.7.2021

[15] Knightt W. The Pentagon Inches Toward Letting AI Control Weapons 2021 https://www.wired.com/story/pentagon-inches-toward-letting-ai-control-weapons/; abgerufen am 1.7.2021

[16] Kremp M. Google Duplex ist gruselig gut. Spiegel 9.5.2018 https://www.spiegel.de/netzwelt/web/google-duplex-auf-der-i-o-gruselig-gute-kuenstliche-intelligenz-a-1206938.html; abgerufen am 2.7.2021

[17] Lazer DMJ, Baum MA, Benkler Y et al. The science of fake news. Science 2018; 359: 1094–1096

[18] Leviathan Y, Matias Y. Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone. Google AI Blog 2018 https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html; abgerufen am 2.7.2021

[19] Machemer T. Human Interruption Slows Down Military Robots in Simulations. Smithonian Magazine. 7. Dezember 2020 https://www.smithsonianmag.com/smart-news/human-interruption-slows-down-military-robots-simulations-180976472/; abgerufen am 1.7.2021

[20] Martel C, Pennycook G, Rand DG. Reliance on emotion promotes belief in fake news. Cogn Research 2020; 5: 47. doi.org/10.1186/s41235-020-00252-3

[21] Menick J. Move 37, Artificial Intelligence, Randomness, and Creativity. 2016 https://www.johnmenick.com/writing/move-37-alpha-go-deep-mind.html; abgerufen am 4.7.2021

[22] Metzinger T. Geist aus der Flasche. Ist der Kampf um einen ethischen Einsatz Künstlicher Intelligenz schon verloren? (Interview). Forschung und Lehre 2021; 28 (7): 548–549

[23] Mirhoseini A, Goldie A, Yazgan M et al. A graph placement methodology for fast chip design. Nature 2021; 594: 207–212

[24] Sætra HS. First, They Came for the Old and Demented: Care and Relations in the Age of Artificial Intelligence and Social Robots. Human Arenas 2020. doi.org/10.1007/s42087-020-00125-7

[25] Silver D et al. Mastering the game of Go with deep neural networks and tree search. Nature 2016; 529: 484489

[26] Silver D et al. Mastering the game of Go without human knowledge. Nature 2017; 550: 345–359

[27] Slonim N, Bilu Y, Alzate C. An autonomous debating system. Nature 2021; 591: 379–384

[28] Sparkes M. Google is using AI to design processors that run AI more efficiently. New Scientist 2021, 3339; https://www.newscientist.com/article/2280321-google-is-using-ai-to-design-processors-that-run-ai-more-efficiently/; abgerufen am 1.7.2021

[29] Spitzer M. Digitales Unbehagen. München: mvg Verlag; 2020

[30] Spitzer M. Psychiatrie im dritten Jahrzehnt des 21. Jahrhunderts. Nervenheilkunde 2021; 40: 6–12

[31] UN Security Councel. Report S/2021/229. Final report of the Panel of Experts on Libya established pursuant to Security Council resolution 1973 (2011), pp 1–548 https://undocs.org/S/2021/229; abgerufen am 3.7.2021

[32] Vosoughi A, Roy D, Aral S. The spread of true and false news online. Science 2018; 359: 1146–1151

[33] Winfield AFT. Experiments in Artificial Theory of Mind: From Safety to Story-Telling. Front Robot AI 2018; 5: 75. doi: 10.3389/frobt.2018.00075

[34] Zablocky P. System-of-Systems Enhanced Small Unit (SESU). Defense Advanced Research Projects Agency (DARPA) 2021; https://www.darpa.mil/program/system-of-systems-enhanced-small-unit#:~:text=SESU%20seeks%20to%20deliver%20system,joint%20and%20coalition%20multi%2Ddomain; abgerufen am 4.7.2021

[35] Zitser J. A rogue killer drone 'hunted down' a human target without being instructed to, UN report says. Business Insider 30.5.2021 https://www.yahoo.com/news/rogue-killer-drone-hunted-down-110624193.html?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_%E2%80%A6%201/6&guccounter=2; abgerufen am 1.7.2021

Dieses Dokument wurde zum persönlichen Gebrauch heruntergeladen. Vervielfältigung nur mit Zustimmung des Verlages.