Christian Heck

Kritik des gläsernen Gefechtsfeldes[1]

Was Sprachmodelle und Massendaten im Krieg bedeuten

Für die Kriegsführung 4.0 ist das gläserne Gefechtsfeld ausschlaggebend. Doch das Internet of Military Things *(IoMT) und* Battle Management Systeme *sind nicht nur militärisch, sondern auch aufgrund ihrer Operationslogik hochgradig kritikwürdige Instrumente. Der Trend zu immer mehr Komponenten des Maschinellen Lernens, die in diese Systeme implementiert werden, scheint derzeit unaufhaltbar. Doch KI ist entgegen der öffentlichen Meinung keine* Blackbox. *Sie besteht aus vielen* Whiteboxes*, in die wir hineinsehen können. Einzig sie zu erschließen, um ihre inneren Funktionsweisen zur maschinellen Bedeutungsgenerierung verstehen zu lernen, dazu sind wir noch nicht in der Lage. So gilt es, die grundsätzlichen Prämissen dieser Systeme adäquat zu kritisieren. Insbesondere die Bedeutung, die ihnen mittlerweile für kriegerisches Handeln zugemessen wird, muss umso mehr Anlass für Kritik sein, die in diesem Beitrag ausgeführt wird.*

Die letzte sogenannte *Revolution in Military Affairs*, die vernetzte Kriegsführung (*Network Centric Warfare* NCW) der späten 1990er-Jahre, in der technologische Innovationen gezielt zur Vernetzung vieler Informations- und Aufklärungssysteme miteinander gestaltet wurden, wurde im 2. Golfkrieg erstmals durch die US-Streitkräfte als Testfeld unter Beweis gestellt. Es entwickelte sich in der Folge eine systemische Denkweise von Militärtheoretiker:innen, dass durch ein integratives Verständnis von Einzelkomponenten exponentielle Leistungssteigerungen des Gesamtsystems in seiner militärischen Wirksamkeit erreicht werden können. Dieses frühe *Internet der Dinge* (IoT) verwob sich zunehmend mit dem Internet, das wir heute kennen: mit sozialen Netzwerken, IT-Monopolen und Clouds, die riesige Rechenzentren weltweit zu Big Data durch Deep Learning[2] und weitere neue technische kognitive Systeme verrechnen. Ein Auftakt ins neue Jahrtausend.

Ein Jahrtausend hybrider Konflikte und *gläserner Gefechtsfelder* mit Kampfhandlungen in unseren Städten, im Netz sowie auch im Weltraum. Das *Internet of Military Things* (IoMT) und die *Kriegsführung 4.0* (vgl. W&F 4/2019) entwickelten sich aus zahlreichen Spin-in- und Spin-off-Effekten, d. h. Innovationen aus Industrie, Wirtschaft und Gesellschaft, die vom Militär übernommen wurden und umgekehrt. Militärtechniken fanden auf diese Weise Eingang in unsere zivilen Räume. In öffentliche und private Räume, in denen wir uns bewegen und miteinander sprechen bzw. chatten.

Eine technisch erzeugte Wirklichkeit

Um *Multi Domain Operationen* (MDO) auszuführen, das heißt sich innerhalb dieser neu ausgeschriebenen Gefechtsfelder zurechtzufinden, Objekte und Situationen zu erkennen, sie adäquat einzuschätzen um daraufhin beste Gewissensentscheidungen zu treffen, benötigen Soldatinnen und Soldaten in Operationszentralen technische kognitive Systeme, sogenannte *Battle Management Systeme* (BMS), zu deutsch: digitale Führungssysteme. Diese Systeme dienen in erster Linie der Entscheidungsunterstützung, da die notwendigen Erkenntnisse innerhalb dieser Art der High-Tech-Kriegsführung nur durch die Unterstützung durch Technologien erlangt werden können. Diese digitalen Führungssysteme versprechen Einsatzschnelligkeit – und in der Vision von MDOs soll dies heißen, nicht nur möglichst schnell, d. h. in Echtzeit, sondern gar seiner Zeit voraus zu handeln. Diese innerhalb dieser Gefechtsfelder durch und durch technische Zeiteinheit ist jedoch nicht für den Menschen gemacht, sondern zur möglichst fehlerfreien Funktionstüchtigkeit der technischen Systeme selbst. So auch die Interpretation anfallender Datenströme, die im gläsernen Gefechtsfeld in Echtzeit in militärisch-technische Handlungen überführt werden muss. Auch diese ist nicht für den Menschen gemacht und er wird die undenkbare Masse an Daten nicht allein bewältigen können, sei er noch so gut ausgebildet. Auch hierfür braucht es technische kognitive Systeme auf aktuellem Stand und ein ausgefeiltes *Man Machine Teaming* (MMT).

Dies bedeutet jedoch auch, dass im gläsernen Gefechtsfeld kognitive Technologien das meiste, das man wahrnehmen und erkennen kann, auch erst herstellen, dass das gläserne Gefechtsfeld also eine technisch erzeugte Wirklichkeit ist. Unter anderem deshalb ist es in MDOs zwingend notwendig, zur eigenen technischen Handlung während der militärischen Operationen Distanz einzunehmen bzw. einnehmen zu können. Dies erfordert, Trennlinien zu setzen zwischen technischer und menschlicher kognitiver Leistung.

Da es bei militärischen Operationen fast immer um Leben und Tod geht, muss ethisch die letzte Entscheidung beim Menschen liegen. Zugleich müssen Führungskräfte, Beamt:innen und politische Entscheidungsträger:innen aber auch anerkennen, dass diese Entscheidungsfindung ohne technische kognitive Systeme nicht realisierbar ist. Sie müssen hierfür also mehr eine innere als eine analytische oder formale Grenze setzen. Eine Grenze zwischen dem Gewissen und der Entscheidung, basierend auf technischer Kognition.

Denn die spezielle Art dieser technischen Systeme räumt menschlichen Akteuren und nicht-menschlichen Artefakten eine gänzlich neue aktive politische Handlungskraft ein. Auch jenen, die in früheren Operationspraxen eine unwichtige, bisweilen gar keine, zumindest aber eine andere Rolle spielten: Dazu zählen unter anderem Betreiber:innen von Cloudplattformen, Rechenzentren und Satellitenanlagen, kriegspropagandistische Influencer:innen oder eben BMS.

Die Integration von großen Sprachmodellen

Erste Komponenten für MD-Operationen wurden bereits in den frühen 1980er Jahren in den Militärapparat implementiert. Hans-Jörg Kreowski erinnerte in seinem Artikel *Die militärische Seite der Digitalisierung* (Kreowski 2023) bspw. an die *Strategic*

Computing Initiative (SCI) aus dem Jahr 1983, die das US-Verteidigungsministerium bereits Jahrzehnte vor der Notwendigkeit von Multi-Domain-Operationen in gläsernen Gefechtsfeldern startete. Mit SCI sollten KI-Projekte entwickelt werden, bei denen neben dem Design autonomer Landfahrzeuge auch die Konzeption eines frühen BMS und eines Sprachassistenten für die Pilot:innen der Luftwaffe zur Aufgabe standen.

Die jüngste Generation von Software-Produkten, die aus diesem Ansatz heraus entwickelt wurden, ist im April diesen Jahres auf dem Markt erschienen. Im September 2023 unterzeichnete das erste Rüstungsunternehmen einen Vertrag mit dem Hersteller, dem US-amerikanischen Datenanalyse-Unternehmen *Palantir* (vgl. Palantir Technologies 2023a). Mit ihrer digitalen Plattform *AIP for Defense* (Palantir Technologies 2023b) werden große vortrainierte Sprachmodelle mit künstlichen neuronalen Einbettungen (LLMs) für militärische Operationen nutzbar gemacht.

Das Unternehmen selbst wurde 2004 gegründet und nahm in diesem Jahrtausend u. a. im „Krieg gegen den Terror" eine nicht zu unterschätzende Rolle ein. Es spezialisierte sich ziemlich schnell auf die Überwachung von Individuen und die Zusammenführung eigentlich getrennter Datenbestände und wurde somit ein wichtiger Akteur in hybriden und asymmetrischen Kriegsführungsstrategien.

Ihre Datenbankvisualisierungs- und Analyse-Software *Gotham* wurde unter anderem zuerst von der *Joint Improvised-Threat Defeat Organization* (JIDO) getestet, einer Einheit des US-Verteidigungsministeriums (DoD), die eingerichtet wurde, um einem neuen Phänomen in dieser Art des Krieges entgegenzuwirken: dem von Anschlägen mit improvisierten Sprengsätzen (*Improvised Explosive Devices* (IED)), mit Autobomben, Paketbomben, Selbstmordattentaten etc. Die CIA, die NSA und das FBI wurden ziemlich schnell zu Kunden von Palantir. Auch die Europäische Polizeibehörde Europol nutzt inzwischen Produkte von Palantir für die Datenauswertung. In der Bundesrepublik wird Palantirs *Gotham* u. a. in Bayern als *Verfahrensübergreifende Recherche- und Analyseplattform* (Vera) eingesetzt. Auch die Polizei in NRW hat Software von Palantir in Betrieb, mit der *Datenbanküber-greifenden Analyse- und Recherche-Software* (DAR). In Hessen wurde *Gotham* in dem System *HessenData* seit 2017 eingesetzt. Die Landesregierung in Hessen hatte die Anschaffung der Software freigegeben, doch der Einsatz wurde im Februar 2023 vom Bundesverfassungsgericht in Karlsruhe als verfassungswidrig eingestuft. Soweit bekannt, führt das System Daten aus sozialen Medien mit Einträgen in verschiedenen polizeilichen Datenbanken sowie Verbindungsdaten aus der Telefonüberwachung zusammen, um mögliche Straftäter:innen zu ermitteln. Zudem spielte Palantir eine nicht unerhebliche Rolle im Skandal um *Cambridge Analytica*. Das Unternehmen soll Facebook bei der Auswertung der illegal weitergegebenen Daten geholfen haben.

Palantirs neuestes Softwarepaket *AIP for Defense* wird nun in naher Zukunft Einsätze unterstützen, indem es u. a. feindliche Stellungen erkennt und durch eine Chatfunktion (ähnlich dem Interface der international viel diskutierten KI *ChatGPT*) Gegenmaßnahmen vorschlägt und gegebenenfalls autonom ausführt – wie z. B. das Starten einer Aufklärungsdrohne ins Zielgebiet. Doch nicht nur das Interface mit integrierter Chatfunktion erinnert an ChatGPT, auch das maschinelle Lernverfahren in AIP ist ein ähnliches wie bei OpenAIs Künstlicher Intelligenz hinter ChatGPT: das generative Sprachmodell GPT-3 (Generative Pretrained Transformer 3).

GPT-3 ist ein LLM, dessen 175 Milliarden Parameter auf Clouds trainiert werden, die über viele Rechenzentren verteilt sind und dessen Entwicklung derzeit an physische und auch ökologisch tragbare Grenzen stößt. Der Stromverbrauch für das Training entspricht dem von 3.000 europäischen Durchschnittshaushalten, eine Frage an ChatGPT benötigt 1.000 Mal mehr Strom als eine Suchanfrage bei Google und für jede Antwort, die man von dem Bot erhält, könnte man ein Smartphone bis zu 60 Mal aufladen. Vermutlich ist *AIP for Defense* eines der ersten *Battle-Management-Systeme*, die LLMs implementiert haben.

Wie maschinelle Bedeutung generiert wird

Bisher hatten LLMs wie eben GPT-3 bzw. GPT-4, LaMDA und PaLM von Google oder LlaMA von Meta in digitalen Führungssystemen eine eher kleine bis gar keine Rolle gespielt. Etablierteren Ansätzen, wenn auch nicht zwingend minder experimentellen bei der komputativen Verarbeitung natürlicher Sprache (NLP[3]), kann auf der anderen Seite schon seit Jahren eine Art Schlüsselrolle zugeschrieben werden.

Der Sozialpsychologe und Sozialwissenschaftler James W. Pennebaker fasste eines der Hauptargumente hierfür wie folgt zusammen: *„Die Worte, die wir im täglichen Leben verwenden, spiegeln wider, worauf wir achten, woran wir denken, was wir zu vermeiden versuchen, wie wir uns fühlen und wie wir unsere Welt organisieren und analysieren."* (Tausczik und Pennebaker 2010:25)

Die Entwicklung von *Word embeddings*[4], zu deutsch: Worteinbettungen, eröffnete hierfür einen gänzlich neuen Handlungsspielraum in der komputativen Sprach- und Netzwerkanalyse.

Zur Extraktion inhaltlicher Strukturen, Features, Organisationseinheiten etc. werden aus einer Vielzahl von generierten semantischen Beziehungen zwischen Wörtern und Sätzen symbolische Repräsentationen in Sprachmodellen mit künstlichen neuronalen Einbettungen errechnet. Eingebettet in digitale Führungssysteme, kommen die Wörter und Sätze, die es zu berechnen gilt, u. a. aus *Open-Source-Intelligence*-Datensätzen (Medienberichten, Social Media, wissenschaftlichen Arbeiten, öffentlich zugänglichen Statistiken etc.), aus Berichten von Geheimdiensten und des militärischen Nachrichtenwesens, Analysen von Sicherheitsbehörden, Erkenntnissen aus der signalerfassenden Aufklärung, der Bild- und Satellitenaufklärung, von Bots, Drohnen und anderen technischen kognitiven Systemen bzw. unbemenschten Fahrzeugen. Die „Lern"-Kriterien, nach denen maschinell Bedeutungen generiert werden, liegen in diesen Sprachmodellen verankert. Nach ihnen werden die Millionen von Gewichtungen, die an den einzelnen künstlichen Neuronen liegen, eingestellt. Eine undenkbare Masse an Informationen aus den unterschiedlichsten Quellen wird für diesen Prozess gesammelt. Sie wird übersetzt, selektiert, kategorisiert und mit strukturierten Daten angereichert, sprich, sie wird enkodiert. Sie wird maschinenlesbar gemacht (*machine readable*), um sie dann nach ihrer Verarbeitung wieder für den Menschen

lesbar zu machen (*human readable*). Es findet auf diese Weise eine maschinelle Vorinterpretation statt. Eine Menschenlesbarmachung von maschinell erlernten symbolischen Repräsentationen, nach vorgegebenen Regeln, die im Code verankert sind.

Als im Jahr 2013 das vortrainierte Sprachmodell *Word2vec* von einem Google-Forscherteam veröffentlicht wurde, galt dies als ein Durchbruch in den NLP-Forschungsgemeinden. Es wurde recht zügig in Technologien der inneren und äußeren Sicherheit implementiert und ist bis heute noch eines der gebräuchlichsten Modelle zur Generierung von Worteinbettungen mittels Deep Learning. Die von Word2vec erzeugten Worteinbettungen können einfach und bequem zur Weiterverarbeitung verwendet werden und zugleich konnte das KI-Modell Ergebnisse auf dem neuesten Stand der Technik (ca. 2013-15) liefern.

Die inneren Funktionsweisen vortrainierter Sprachmodelle mit künstlich neuronalen Einbettungen, egal ob LLMs oder Word-2vec, beruhen auf der Idee, dass im Gegensatz zur formalen Linguistik und zur Chomskyschen Tradition allein die Kontextinformation eine brauchbare Darstellung sprachlicher Elemente darstellt. Je nach Modell werden hier einzelne Wörter eines Korpus (Textes) als symbolische Repräsentationen in einem semantischen Vektorenraum (Wortraum) mit etwa 300 Dimensionen dargestellt. Zum Vergleich: in heutigen Transformer-Architekturen à la GPT-3 werden 1.536 Dimensionen pro Wort errechnet. Worteinbettungen repräsentieren also den innertextlichen Kontext eines Datensatzes, in dem das jeweilige Wort vorkommt.

Niemanden interessiert, wie es funktioniert, solange es funktioniert

Das Studium aktueller NLP-Forschungsarbeiten zeigt, dass trotz der weit verbreiteten Anwendung von künstlich neuronalen Worteinbettungen noch immer erstaunlich wenig über die Struktur und die Eigenschaften – und folglich auch die Konsequenzen – dieser Einbettungsräume bekannt ist. Dennoch werden zunehmende Teile von Welt durch sie in formale, computerlinguistisch verarbeitbare Informationen und Beschreibungsebenen (Morphologie, Syntax, Semantik, Aspekte der Pragmatik etc.), das heißt in symbolische Repräsentationen umgewandelt.

Auf diese Weise fließen aber auch immer wieder auftauchende, bisweilen diskriminierende Tendenzen bis hin zu Rassismen in digitale Führungssysteme mit ein. Sogenannte „Verzerrungen", die diesen textanalytischen Machine-Learning-Verfahren zwar nicht explizit, auch nicht vorsätzlich eingeschrieben werden, die aber dennoch im realweltlichen Gebrauch in Erscheinung treten. Denn trotz dieser bekannten und in den letzten 3-4 Jahren auch in der Öffentlichkeit verhandelten Defizite von KI-Sprachmodellen implementieren Firmen diese weiterhin in ihre Produkte und verkaufen sie an Sicherheitsbehörden und an das Militär.

Der Trend hin zu immer größer werdenden Modellen und immer mehr (unüberprüften) Trainingsdaten in den letzten Jahren führt auch zu einer immer schlechter werdenden Kontrollierbarkeit der inneren Funktionsweisen von technischen kognitiven Systemen. Funktionsweisen, durch die Minderheiten diskriminiert und gesellschaftliche Gruppen marginalisiert werden, und das auf eine Weise, die meist den Entwickler:innen selbst nicht bekannt bzw. bewusst, schlimmstenfalls egal ist (vgl. Bender et al. 2021).

Auch aus diesem Grund ist es für Soldatinnen und Soldaten in Operationszentralen oder in Ämtern und Firmen, die die analytische Ausarbeitung für Führungskräfte unterstützen, unabdingbar, sich ein grundlegendes Verständnis des inneren Aufbaus dieser technischen kognitiven Systeme anzueignen. Denn die Bedeutung der jeweiligen militärisch-technischen Handlung muss aus diesen Systemen heraus, also in deren inhärenter Pragmatik, erschlossen werden. Aus technischen kognitiven Systemen, die neben ihrer natürlichen Begrenztheit zugleich auch eine scheinbare Grenzenlosigkeit zu Tage fördern.

Eine Grenzenlosigkeit, die in ihrem Lern-Vermögen liegt, die uns derzeit auch bis zur Entgrenzung unseres menschlichen Denkens führt. Denn diese technischen Systeme können *„sowohl lernen, dass die Erde flach ist, als auch rund"* (Chomsky et al. 2023), so Noam Chomsky (linker Intellektueller, Anarchist und emeritierter Professor für Linguistik am Massachusetts Institute of Technology, MIT) im März 2023 in der *New York Times*. Sie können auch lernen, dass seit weit über einem Jahr ein von Russland geführter Angriffskrieg gegen die ukrainische Bevölkerung wütet. Im nächsten Moment können sie dies jedoch auch wieder verlernen, egal ob es der Wirklichkeit entspricht oder nicht. Was diese technischen kognitiven Systeme eben nicht können, ist, ihre innere Grenzenlosigkeit durch ethische Prinzipien einzuschränken. Eine Eigenschaft, die wir als „moralisches Denken" bezeichnen. Die technischen kognitiven Systeme sind, während sie prozessieren, getrennt von der Außenwelt. In ihren inneren Entscheidungsfunktionen liegen keine Modellierungen von dem, was Worte, was Dinge, was Taten und Ereignisse für uns in der Welt bedeuten. Dennoch werden sie über Leben und Tod von Soldatinnen und Soldaten, von Jugendlichen, von Großeltern, von Eltern und unseren Kindern algorithmisch mitentscheiden.

Christian Heck

Christian Heck ist künstlerisch-wissenschaftlicher Mitarbeiter für Ästhetik & neue Technologien / Experimentelle Informatik und Doktorand an der Kunsthochschule für Medien Köln. Seine Forschungs- und Arbeitsschwerpunkte liegen auf Friedensforschung, Ästhetische Praxis und Ethik der Künstlichen Intelligenz mit Fokus auf Generative Systeme, ADM, IT-Sicherheitstechnologien, Kampfdrohnen und autonome Waffensysteme. Er ist Mitglied im Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung (FIfF) e. V. und der Gesellschaft für Informatik (GI).

Referenzen

Bender EM, Gebru T, McMillan-Major A and Shmitchell S (2021) On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, S. 610-623.

Chomsky N, Roberts I, Watumull J (2023) The False Promise of ChatGPT. New York Times, 8.3.2023.

Kreowski HJ (2023) Die militärische Seite der Digitalisierung. IMI-Ausdruck 113, S. 25-27.

Mikolov T, Sutskever I, Chen K, Corrado G, Dean J (2013) Distributed Representations of Words and Phrases and their Compositionality. NIPS'13: Proceedings of the 26th International Conference on Neural Information Processing Systems – Volume 2, S. 3111–3119.

Palantir Technologies (2023a) Palantir Technologies Signs Partnership With Titan Defence Firm, Babcock. Pressemitteilung, 13.9.2023.

Palantir Technologies (2023b) AIP for Defense. Homepage, URL: palantir.com/aip/defense/.

Tausczik YR, Pennebaker JW (2010) The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. Journal of Language and Social Psychology 29 (1), S. 24-54.

Anmerkungen

1 *Nachdruck aus dem Jubiläumsheft 4/2023 der Zeitschrift Wissenschaft und Frieden mit freundlicher Genehmigung der W&F-Redaktion und des Autors.*

2 *Die Technologie des „Deep Learning" begann sich um die Jahrtausendwende zu entfalten. Es begann bald darauf die Zeit von Big Data (dem Anstieg der Datenmengen durch die Verbreitung der Internettechnologien), und es wurden erhebliche Fortschritte in den Computertechnologien (in der Rechenkapazität, GPUs und preiswerten Speichertechnologien) erzielt. Erst durch diese technische Infrastruktur wurde die Weiterentwicklung der Künstlichen Neuronalen Netze (KNN) hin zum Deep Learning im Forschungs- und vermehrt auch im Anwendungsbereich möglich. Von dieser Technologie sprechen wir heute in erster Linie, wenn von Künstlicher Intelligenz zu hören ist: der subsymbolischen Künstlichen Intelligenz.*

3 *Das Kürzel NLP steht für Natural Language Processing. Eine Mischwissenschaft, die anteilig aus der Computerlinguistik, den Computerwissenschaften und der Künstliche Intelligenz Forschung besteht. Sie ist eine Wissenschaft der algorithmischen Verarbeitung von Sprache, der Verarbeitung von Daten und des künstlich intelligenten Verhaltens zugleich.*

4 *Der Sammelbegriff „Word embeddings" steht für eine Reihe von Sprachmodellierungs- und Feature-Learning-Techniken in NLP, bei denen Wörter oder Phrasen aus dem Vokabular auf Vektoren mit reellen Zahlen abgebildet werden: z. B. eine globale Korpusstatistik (GloVe: Globale Vektoren für Wortdarstellung) oder eine Wortkontextdarstellung (Word2vec) (siehe Mikolov et al 2013).*