

Tobias Wilke: KI, DIE BOMBE. Zu Gegenwart und Geschichte einer Analogie

zfl. zflprojekte.de/zfl-blog/2023/10/19/tobias-wilke-ki-die-bombe-zu-gegenwart-und-geschichte-einer-analogie/

19. Oktober 2023

»A.I. or Nuclear Weapons: Can You Tell These Quotes Apart?« – so fragte die *New York Times* ihre Leser:innen am 10. Juni dieses Jahres. In Form eines metadiskursiven Ratespiels rückte die Zeitung damit eine Analogie in den Blick, die in den jüngsten Debatten um Künstliche Intelligenz zu erheblicher Prominenz gelangt ist. Wenn derzeit die Risiken neuester (Sprach-)Technologien beschworen werden, lässt der Vergleich mit dem Vernichtungspotenzial von Atomwaffen nicht lange auf sich warten. Dies gilt für die in Print- und Onlinemedien ausgetragene Diskussion, ist aber auch innerhalb der wissenschaftlichen Community mit ihren Spezialöffentlichkeiten zu beobachten.[1] Warnungen vor den unabsehbaren Folgen der KI-Entwicklung gleichen in ihrer Drastik und teils bis aufs Wort den Mahnrufen und Appellen, mit denen in der Nachkriegszeit auf die damals neue Gefahr eines drohenden globalen Nuklearkonflikts reagiert wurde. Ob sich die von der *New York Times* anonymisiert präsentierten Aussagen wie »If we go ahead on this, everyone will die« oder »We are drifting toward a catastrophe beyond comparison«[2] auf unsere Gegenwart oder auf die historische Situation der 1950er und 1960er Jahre beziehen, ist daher tatsächlich nicht immer ohne Weiteres auszumachen. Und erst kürzlich berichtete die Zeitung von einem Mitarbeiter des von Google betriebenen DeepMind-Forschungslabors, der einen Vortrag zum maschinellen Lernen mit einem Zitat aus George Orwells Essay *You and the Atomic Bomb* (1945) eröffnet hatte – ein weiteres Beispiel dafür, wie sehr die zwischen Künstlicher Intelligenz und Nuklearwaffen gezogenen Parallelen zu einem anscheinend unverzichtbaren Topos in der Auseinandersetzung mit neuronalen Netzen und Large Language Models (LLMs) geworden sind.[3]

Erklären lässt sich diese Ubiquität zunächst damit, dass die Atombombenreferenz in der Regel als Teil von Aufmerksamkeitsstrategien fungiert, die auf die Dringlichkeit von Sicherheitsstandards, Risikomanagement und Kontrollmechanismen für die KI-Entwicklung hinweisen sollen. Insofern solche Warnungen und Forderungen jedoch praktisch immer mit der Betonung des schier »grenzenlosen« Problemlösungspotenzials von Künstlicher Intelligenz einhergehen bzw. sich eben daraus allererst begründen, befeuern sie zusätzlich den gegenwärtigen Hype um diese Technologien. Dabei fällt auf, dass es gerade die Protagonisten der Big-Tech-Konzerne sind, die in den Chor der vielzähligen Warnrufe und Regulierungsappelle einstimmen bzw. darin sogar den Ton angeben. Als Ende Mai etwa das Center for AI Safety ein offizielles *Statement on AI Risk* herausgab, in dem erklärt wurde, »[m]itigating the risk of extinction from AI should be a global priority alongside other society-scale risks such as pandemics and nuclear war«, zählten die CEOs von OpenAi und Google DeepMind, Sam Altman und Demis Hassabis, sowie der Microsoft-Gründer Bill Gates zu den

prominentesten Unterzeichnern. Und in einem schon im April veröffentlichten Memorandum zum Thema *Governance of Superintelligence* schlugen die ChatGPT-Entwickler von OpenAI selbst die Einrichtung einer KI-Kontrollinstanz nach dem Vorbild der Internationalen Atomenergiebehörde (IAEA) vor.

Diese Beispiele lassen erkennen, dass in den Diskussionen der vergangenen Monate durchaus verschieden geartete Vorstellungen von einer atombombenhaften Gefährdung der Menschheit durch KI miteinander konkurrieren. Am konkretesten gestalten sich die Inbezugsetzungen dort, wo auf die künftige Integration von Künstlicher Intelligenz *in* die Steuerung nuklearer Waffensysteme sowie auf eine zu erwartende Automatisierung von militärischen Entscheidungsprozessen insgesamt rekurriert wird. »Never give artificial intelligence the nuclear codes« lautete etwa die Überschrift eines Artikels in der Juni-Ausgabe des Magazins *The Atlantic*, der mit einem (computergenerierten) Bild eines aus binärem Zahlencode geformten Atompilzes illustriert wurde. Und auch in den Kontexten von Security Studies und internationaler Politik debattiert man mittlerweile intensiv über die Möglichkeit eines globalen »AI arms race«, das die Gefahren des atomaren Wettrüstens nach dem Zweiten Weltkrieg noch weit in den Schatten stellen würde.[4]

Während solche Szenarien also zu antizipieren versuchen, wie sich Künstliche Intelligenz ganz buchstäblich in eine Bombe verwandeln könnte, kreist ein anderer, sehr viel stärker spekulativer Strang der Auseinandersetzung um ein aus Filmen wie *Terminator* und *The Matrix* bekanntes Motiv: Die dort noch fiktional verhandelte Auslöschung der Menschheit durch außer Kontrolle geratene »superintelligente« Maschinen wird nun unter dem Stichwort der sogenannten »rogue AI«[5] – also einer abtrünnigen, zerstörerischen Variante gegenwärtiger Computertechnologien – in den Raum bald schon möglicher Realitäten verlegt. Unter Rückgriff auf das Imaginationsarchiv der Science-Fiction (die ihrerseits ja gerade im Zeitalter der ersten Atombomben und Digitalrechner breite Popularität erlangte) bilden sich somit prognostische Aussagen darüber heraus, wie Künstliche Intelligenz faktisch zu einer Bedrohung im planetarischen Maßstab avancieren könnte. Entscheidend ist dabei die Annahme, dass sie sich bereits in naher Zukunft – und im Gegensatz zu »klassischen« Atomwaffen – einer nachträglichen Einhegung durch menschliche Akteure auf irreversible Weise entziehen dürfte.

Auf der einen Seite also die Warnung vor Künstlicher Intelligenz *als* (Teil der) Bombe, auf der anderen Seite die Befürchtung, dass KI letztlich *wie* die Bombe, oder sogar eher noch als diese, zu einer Quelle von totaler physischer Vernichtung zu werden vermag: Zwischen diesen extremen Polen hat es in den letzten Monaten eine Fülle von äußerst diffusen Beschwörungen der potenziell destruktiven Kräfte von *generative artificial intelligence* gegeben. In einer der frühesten prominenten Reaktionen auf ChatGPT forderten im März 2023 zahlreiche Vertreter:innen aus US-amerikanischer Wissenschaft, Wirtschaft und Politik in einem offenen Brief ein mindestens sechsmonatiges Moratorium in der KI-Entwicklung. Denn Systeme, die das Problemlösungsvermögen menschlicher Intelligenz besäßen oder gar überträfen, brächten »profound risks to society and humanity« mit sich, darunter die

mögliche Auflösung zivilisatorischer Institutionen, soziale Desintegration und die Verdrängung von menschlicher *agency* aus den Sphären von Ökonomie, Arbeit, Politik und Kultur. Die angesprochenen Gefahren bleiben hier – wie auch in vielen anderen Einlassungen zur Künstlichen Intelligenz – eher vage. (Die automatisierte Produktion und Verbreitung von Fake News und Propaganda durch Chatbots und eine daraus folgende Überflutung gesellschaftlicher Kommunikationskanäle ist noch einer der deutlicher benannten katastrophalen Effekte.) Doch gibt die Vorstellung einer zwischen privaten Konzernen aus ökonomischen Interessen ausgetragenen Konkurrenz um immer leistungsfähigere KI-Systeme eine mindestens implizite historische Referenz zur Dynamik des atomaren Wettrüstens zu erkennen. Zu befürchten bzw. eigentlich schon im Gange sei, so die Autor:innen des Briefs, »an out-of-control race to develop and deploy ever more powerful digital minds that no one – not even their creators – can understand, predict, or reliably control«.

In ihrer kurzen Vorbemerkung weisen die Autoren des eingangs genannten *New York Times*-Ratespiels darauf hin, dass der von ihnen zum Quizgegenstand gemachte Vergleich durchaus kein gänzlich neues Phänomen, sondern bereits seit Jahren verwendet worden sei. Dabei beziehen sie sich namentlich auf Tesla-Gründer und X-Besitzer Elon Musk, einen der derzeit wohl vernehmbarsten Advokaten einer umfassenden KI-Regulierung, der schon am 3.8.2014 in einem Tweet warnte, »[w]e need to be super careful with AI. Potentially more dangerous than nukes« – um sich dann wenig später selbst an der Gründung von Open AI zu beteiligen.[6] Hervorzuheben ist allerdings, dass nicht erst die beschleunigte Entwicklung von ›Superintelligenz‹ im frühen 21. Jahrhundert zu solchen Vergleichen Anlass gab. Diese bilden vielmehr eine Konstante in der diskursiven Auseinandersetzung mit digitalen Technologien, was sich insbesondere an zwei bereits länger zurückliegenden formativen Phasen des Informationszeitalters beobachten lässt: dem Digitalisierungsschub der 1980er und frühen 1990er Jahre, der durch die Verbreitung erschwinglicher PCs und die Freigabe des World Wide Web (1993) für die öffentliche Nutzung gekennzeichnet war, sowie den Nachkriegsjahrzehnten, die sowohl durch die Anfänge des *mainframe computing* als auch durch die eskalierende atomare Konfrontation zwischen Ost- und Westmächten geprägt wurden.

»**La bombe informatique/Die Informatikbombe**«: Unter diesem Titel strahlte der deutsch-französische Kultursender ARTE im Herbst 1995 ein Fernsehgespräch zwischen dem Philosophen Paul Virilio und dem Medienwissenschaftler Friedrich Kittler aus.[7] Geführt gleichsam auf halber Strecke zwischen Kittlers frühen Studien zur Mediengeschichte wie *Grammophon Film Typewriter* (1986) und Virilios später publiziertem Buch *The Information Bomb* (1999) legt diese Diskussion Zeugnis davon ab, wie der Prozess globaler digitaler Vernetzung erstmals nachhaltig ins Zentrum der geistes- und kulturwissenschaftlichen Theoriebildung zu rücken begann. Für Virilio, den Theoretiker der Geschwindigkeit und ihrer

militärischen wie medientechnischen Hintergründe, stellte sich das Heraufziehen einer weltweit synchronisierten Informationsgesellschaft als ultimatives Stadium historischer Beschleunigung mit einem Unfallpotenzial von nie gesehenen Ausmaßen dar:

»[T]his is what Einstein called, very judiciously, ›the second bomb‹. The first bomb was the atomic bomb, the second one is the information bomb, that is, the bomb that throws us into ›real time‹.«[8]

Ausgehend von einer – vermutlich apokryphen[9] – Einstein-Äußerung machte Virilio so eine Reihe von bildlichen Analogien zwischen nuklearer und digitaler Katastrophe geltend: Instantane Datenübertragung und unbegrenzte Interaktivität würden letztlich zu einer »Überhitzung« gesellschaftlicher Kommunikation, zu einer Art von sozialer »Kernspaltung« mit entsprechend destruktiven Folgen führen müssen; geboten seien daher Kontrollen und internationale Präventivmaßnahmen nach dem Vorbild der »nuklearen Abschreckung«.[10]

Kittler hingegen richtete seinen Blick in für ihn charakteristischer Weise auf die historischen Entwicklungszusammenhänge, die Computertechnologie und Atomwaffen seit ihren Anfängen miteinander verbunden haben. Dies klingt bedeutend weniger alarmistisch als Virilios Statements und kommt im Ton eher lakonisch daher. Doch galt das Kommen der Apokalypse auch Kittler als ausgemacht. »Vor dem Ende, geht etwas zu Ende«, hieß es in diesem Sinne schon in der Einleitung zu *Grammophon Film Typewriter* (das just im Jahr des Reaktorunfalls von Tschernobyl erschienen war).[11] Ehe das nukleare Armageddon eintritt, so sollte dies besagen, vermag die zu militärischen Zwecken entwickelte Datenübertragung per Glasfaserkabel noch den zivilen Funktionsbereich sämtlicher (Unterhaltungs-)Medien zu revolutionieren – und deren Nutzer:innen genau dadurch von der Einschreibung der bevorstehenden Katastrophe in die digitale Technik abzulenken. Denn die Verkabelung von Rechnern, wie sie zunächst im Arpanet des US-Militärs erprobt wurde, hatte ursprünglich dazu gedient, Waffen- und Kommunikationssysteme gegen die elektromagnetischen Auswirkungen möglicher Nuklearschläge zu immunisieren. Für Kittler leitete sich daraus eine materiell verbürgte Komplizenschaft zwischen Digitalisierung und atomarer Aufrüstung her, die hinter den Datenflüssen alltäglicher Computernutzung jedoch verborgen bleibe: »Glasfaserkabel übertragen eben jede denkbare Information außer der einen, die zählt – der Bombe.«[12]

Virilio und Kittler waren noch nicht mit Künstlicher Intelligenz in ihren heute aktuell gewordenen Erscheinungsformen befasst. Doch weisen ihre Reflexionen auf bestimmte Kontinuitäten hin, die sich auch in der forcierten Digitalisierung des frühen 21. Jahrhunderts fortsetzen sollten. So lässt sich eine direkte Entwicklungslinie vom (geschlossenen) Arpanet der 1970er und 80er Jahre über das (offene) Internet der 1990er und 2000er zu den (teils offenen, teils geschlossenen) LLMs der Gegenwart ziehen, die ohne ihre aus dem Netz bezogenen sprachlichen Trainingsdaten schlechthin undenkbar wären. Kittlers Hinweis indes, »that both computers and atomic bombs are an outcome of the Second World War«, [13] erlaubt es zugleich, diese historischen Verbindungen bis an die Anfänge des

Computerzeitalters zurück zu verlängern. Eine zentrale Rolle kommt in diesem Kontext dem Mathematiker John von Neumann zu, der während des Kriegs zeitweilig am Los Alamos National Laboratory arbeitete und dort Berechnungen für das Design der ersten Atombombe anfertigte; wenig später entwickelte er im Rahmen des Manhattan-Projekts der US-Regierung ein neues Schaltungskonzept für elektronische Rechner, die sogenannte Von-Neumann-Architektur, die zum bis heute gängigen Bau- und Funktionsplan von Computern avancieren sollte.

Zum Komplex dieser institutionell-genealogischen Verflechtungen gehört aber auch, dass sich die rasch aufkommende Frage nach der Lernfähigkeit und quasimenschlichen Intelligenz solcher Rechen- bzw. Denkmaschinen[14] sehr bald mit Reflexionen über die möglichen Folgen der Atomtechnologie und das nach dem Krieg einsetzende Wettrüsten zwischen den USA und der UdSSR zu verbinden begann. So zu verfolgen etwa bei Norbert Wiener, der die erste Ausgabe seines Buches *Cybernetics: Or Control and Communication in the Animal and the Machine* (1948) mit einem kurzen, an von Neumann anknüpfenden Ausblick auf die Möglichkeit beschloss, lernfähige Schachcomputer zu konstruieren. In der zweiten Auflage 1961 fügte er diesen Überlegungen dann ein neu verfasstes Kapitel (*On Learning and Self-Reproducing Machines*) hinzu, in dem er seine Perspektive auf den möglichen Einsatz solcher ›intelligenten‹ Maschinen im Kontext militärischer Konfliktszenarien, vor allem des »new and as yet untried war with atomic weapons«, [15] ausweitete. Kurz vor dem kritischen Höhepunkt der atomaren Konfrontation angesichts der Kubakrise (1962) sah Wiener somit schon die Automatisierung von Waffensystemen und kriegsstrategischen Entscheidungsprozessen mittels Künstlicher Intelligenz voraus, wie sie gegenwärtig unter neuen technischen Vorzeichen diskutiert wird. Und er hob dabei verschiedene mit dieser Entwicklung verbundene Gefahren hervor, die auch in der aktuellen Debatte um KI zentrale Bedeutung (wieder-)erlangt haben: das Problem eines »unguarded use of learning machines«[16] zum Beispiel, das aus einer unkontrollierbaren Eigendynamik solcher Systeme resultiert, sowie die noch weiter reichende Möglichkeit, dass sich diese Systeme aufgrund ihrer »new and real agencies«[17] irgendwann sogar gezielt gegen ihre Schöpfer wenden könnten.

Weniger auf das destruktive Potenzial von Künstlicher Intelligenz gerichtet, doch dafür umso radikaler in der Einschätzung ihrer zivilisatorischen Konsequenzen war wiederum eine Diagnose, die der deutsche Wissenschaftsphilosoph Max Bense bereits 1955 formulierte:

»Nicht die Erfindung der Atombombe ist das entscheidende technische Ereignis unserer Epoche, sondern die Konstruktion der großen mathematischen Maschinen, die man [...] gelegentlich auch Denkmaschinen genannt hat. Diese Feststellung begründet sich auf der Tatsache, daß die Technik mit ihnen einen neuen Aufgabenbereich, fast möchte man sagen: einen neuen Sinn gewonnen hat.«[18]

Zehn Jahre nach den ersten Atombombenabwürfen über Japan und inmitten der Eskalationsphase des Kalten Kriegs war diese Einschätzung gewiss (noch) nicht konsensfähig und ersichtlich (auch) als Provokation intendiert. Dennoch war es Bense mit seinem vergleichenden Befund durchaus ernst. Während die Atombombe vor allem eine physische Bedrohung von außen darstellte, waren die ›Denkmaschinen‹ dazu angetan, alle Bereiche des sozialen und geistigen Lebens zu durchdringen und in ihrem innersten Kern zu verändern; und eben dies rechtfertigte es, sie als die eigentliche Grundlage »einer neuen Stufe der Technischen Welt«^[19] anzusehen. Benses halb diagnostische, halb prognostische Einschätzung stellt sich im Rückblick als hellsichtige Charakterisierung des (kommenden) Digitalzeitalters dar – ist es doch gerade das mittlerweile eingetretene und sich aktuell stark beschleunigende Vordringen von Künstlicher Intelligenz in die unterschiedlichsten Kommunikations-, Informations- und Kulturtechniken, das zuletzt auch für eine topische Rückkehr der Atombombe in die öffentliche Diskussion gesorgt hat.

Was aber hat eine heutige KI wie ChatGPT selbst zur Konjunktur dieser vergleichenden Bezugnahmen zu sagen? Stellt man dem Chatbot diese Frage (und was läge in der aktuellen Situation näher?), so erzeugt das System eine Liste von generischen Aussagen über ähnliche Risiken, unabsehbare Konsequenzen und ethische Implikationen beider Technologien, ergänzt um den relativierenden Hinweis, es handle sich um »keine perfekte Analogie«.^[20] Dieser Einschränkung mag ohne Weiteres zuzustimmen sein. Entscheidend jedoch ist etwas anderes: Die diskursive Rekurrenz des Vergleichs – und seine eben darauf beruhende Erfass- und Reproduzierbarkeit durch statistisch verfahrenende LLMs – lässt ideologische *constraints* sichtbar werden, denen die Bildung von Aussagen über Künstliche Intelligenz unterliegt – wobei ›ideologisch‹ hier im Sinne einer Regulierung des Denk- und Sagbaren durch bestimmte Wahrscheinlichkeiten zu verstehen ist.^[21] Eben diese Wahrscheinlichkeiten aber sind es wiederum, die auch und gerade in Zukunft eine Analyse ihrer historischen Bedingtheit notwendig machen.

Der Literaturwissenschaftler Tobias Wilke arbeitet am ZfL auf einer Heisenberg-Stelle an dem Projekt »Digitale Sprache. Linguistik, Kommunikationsforschung und Poetik im frühen Informationszeitalter«.

[1] Für die journalistische Debatte im deutschen Sprachraum vgl. exemplarisch Alexander Grau: »KI-Moratorium? Künstliche Intelligenz ist die Atombombe«, in: *Cicero Online*, 15.4.2023, und Ursula Scheer: »Ist KI so gefährlich wie die Atombombe?«, in: *Frankfurter Allgemeine Zeitung*, 24.5.2023. Für den wissenschaftlichen Diskussionskontext vgl. u.a. Dan Hendrycks/Mantas Mazeika/Thomas Woodside: *An Overview of Catastrophic AI Risks*, in: arXiv, 9.10.2023, S. 4.

[2] Ian Prasad Philbrick/Tom Wright-Piersanti: »A.I. or Nuclear Weapons: Can You Tell These Quotes Apart?«, in: *The New York Times*, 10.6.2023.

- [3] Vgl. Siobhan Roberts: »AI Is Coming for Mathematics, Too«, in: *The New York Times*, 2.7.2023. Orwells Essay erschien am 19.10.1945 in der Londoner Zeitung *Tribune*.
- [4] So warnte unlängst u.a. Charles Oppenheimer, Enkel des Atombomben-Entwicklers J. Robert Oppenheimer, vor einer entsprechenden Konkurrenz zwischen China und den USA. Vgl. Charles Oppenheimer: »To Avoid an AI ›Arms Race‹, the World Needs to Expand Scientific Collaboration«, in: *Bulletin of Atomic Scientists*, 12.4.2023. Als einschlägige Publikation aus dem Bereich der Security Studies vgl. auch James Johnson: *AI Bomb: Nuclear Strategy and Risk in the Digital Age*, Oxford 2023.
- [5] Zu diesem Begriff vgl. Hendrycks(Mazeika/Woodside: *Overview* (Anm. 1)
- [6] Vier Jahre später (11.4.2018) war Musk sich angeblich sicher: »Mark my words – A.I. is far more dangerous than nukes.«
- [7] Ein Video der Sendung ist online zugänglich. Eine schriftliche englische Übersetzung von Auszügen des Gesprächs erschien später unter dem (leicht abgewandelten) Titel »The Information Bomb. A Conversation«, in: *Angelaki. Journal of the Theoretical Humanities* 4.2 (1999), S. 81–90.
- [8] Ebd., S. 82.
- [9] Vgl. dazu Roger Stahl: »Weaponizing Speech«, in: *Quarterly Journal of Speech* 102.4 (2016), S. 376–395, hier S. 378.
- [10] Vgl. Paul Virilio: *The Information Bomb*, New York 1999, S. 108.
- [11] Friedrich Kittler: *Grammophon Film Typewriter*, Berlin 1986, S. 7.
- [12] Ebd.
- [13] »The Information Bomb. A Conversation« (Anm. 7), S. 82.
- [14] Prägend hierzu Alan Turing: »Computing Machinery and Intelligence«, in: *Mind* 59.236 (1950), S. 433–460.
- [15] Norbert Wiener: *Cybernetics, Or Communication and Control in the Animal and the Machine*, Boston ²1961, S. 242.
- [16] Ebd.
- [17] Ebd., S. 244.
- [18] Max Bense: »Vorwort«, in: Louis Couffignal: *Denkmaschinen*, übers. v. Elisabeth Walther und Max Bense, Stuttgart 1955, S. 7.

[19] Ebd., S. 7f.

[20] ChatGPT (May 24 Version), Response auf die Frage »Why is AI so often compared to the atomic bomb?«, generiert am 5.7.2023. Hervorzuheben ist, dass sich die Datenbasis von ChatGPT bislang nur bis Ende 2021 erstreckt und damit (noch) nicht den Zeitraum erfasst, in dem die Freischaltung des Systems selbst maßgeblich zur Renaissance der Atombomben-Referenz beigetragen hat.

[21] Vgl. dazu Leif Weatherby: »ChatGPT is an Ideology Machine«, in: *Jacobin*, 17.4.2023.

VORGESCHLAGENE ZITIERWEISE: Tobias Wilke: KI, die Bombe. Zu Gegenwart und Geschichte einer Analogie, in: ZfL Blog, 19.10.2023, [<https://www.zflprojekte.de/zfl-blog/2023/10/19/tobias-wilke-ki-die-bombe-zu-gegenwart-und-geschichte-einer-analogie/>]. DOI: <https://doi.org/10.13151/zfl-blog/20231019-01>