

⇒ **Nicole Kunkel**

## Autoregulative Waffensysteme. Automatisierung als friedensethische Herausforderung – ein Werkstattbericht

⇒ Einleitung

Sogenannte autonome Waffensysteme – oder besser Waffensysteme mit autoregulativen Funktionen – sind Gegenstand aktueller politischer und gesellschaftlicher Debatten. Vor allem auf internationaler Ebene wird im Rahmen der UN über die Entwicklung und den möglichen Einsatz von Automatisierungsprozessen in Waffensystemen verhandelt. In meiner Promotion beschäftige ich mich seit 2018 mit den friedensethischen Herausforderungen, die solche Systeme mit sich bringen, und möchte im Folgenden einen Einblick in meine noch nicht abgeschlossenen Überlegungen zu diesem Thema bieten. Dazu werde ich zunächst vorstellen, wovon genau die Rede ist, wenn von autoregulativen Funktionen in Waffensystemen gesprochen wird. Dann werde ich deutlich machen, warum ich den Terminus *Autoregulation* der Rede von Autonomie vorziehe, und damit zugleich ethische Einsichten verbinden. Im Zuge dessen sollen auch die bisher vorgebrachten ethischen Argumente und der damit verbundene politische Prozess dargestellt werden. Zum Abschluss werde ich das Thema in einen größeren friedensethischen Rahmen einbinden und einige ethische Diskussionspunkte benennen.

⇒ 1 Autoregulative Funktionen in Waffensystemen

Autoregulative Funktionen spielen in der derzeitigen technischen Entwicklung, vor allem im Bereich von Robotik, eine zentrale Rolle. Gemeinsam ist solchen, dass sie ohne menschliche Intervention und

Kontrolle in einem bestimmten Bereich auskommen sollen. Für den Verkehr betrifft dies beispielsweise die Entwicklung selbstfahrender – autoregulativer – Fahrzeuge, während für die Pflege der Einsatz autoregulativer Pflegeroboter diskutiert wird

---

**Nicole Kunkel**, 1984 in Löbau, Dipl.-Theologin, Studium in Leipzig, Jerusalem und Berlin, Doktorandin am Lehrstuhl für Ethik und Hermeneutik der Theologischen Fakultät der Humboldt-Universität zu Berlin.

GND: 1088338763

---

DOI: [10.18156/eug-2-2021-art-6](https://doi.org/10.18156/eug-2-2021-art-6)

(vgl. hierzu etwa: Loh 2019, 22-29; oder Misselhorn 2018, 136ff). Auch Waffensysteme sind von diesem Prozess fortschreitender Automatisierung betroffen. Die grundlegende Funktionsweise eines solchen Waffensystems ist alles andere als komplex und lässt sich an der *Loitering Munition IAI Harpy* veranschaulichen: Ist diese Rakete erst einmal gestartet, schwebt sie solange über feindlichem Luftraum, bis sie eine spezielle, feindliche Radiosignatur – in aller Regel von einem System zur Raketenabwehr – empfängt. Bei Empfang stürzt sich die Waffe auf das Ziel und zerstört es dabei.<sup>1</sup> Eine solche Waffe bringt jedoch bei allen möglichen militärischen Vorzügen, die sie bieten mag, zugleich etliche Probleme mit sich. Etwa kann die Waffe nicht ohne Weiteres beurteilen, ob die Radiosignatur von einem militärischen Gelände, einem Wohnhaus oder sogar einem Krankenhaus gesendet wird. Alle drei Szenarien jedoch würden nach geltendem internationalem Recht sehr unterschiedlich beurteilt werden. Handelt es sich bei ersterem nämlich um ein im Krieg legitimes Ziel, so stellen die beiden anderen – sicherlich in Abstufungen – kein solch legitimes Ziel dar.

Um eine etwas andere Anwendung von Autoregulation geht es in dem Militärprojekt FCAS (*Future Combat Air System*). In diesem gemeinsamen Vorhaben Frankreichs, Deutschlands und Spaniens sollen autoregulative Funktionen eine zentrale Rolle spielen. Der Mitentwickler Wolfgang Koch beschreibt FCAS mit folgenden Worten:

Konflikte werden künftig automatisierter ausgetragen als je zuvor. Im Rahmen des Future Combat Air System FCAS sollen unbemannte, künstlich intelligente und technisch autonome Luftfahrzeuge bemannte Kampffjets neuester Generation als *loyale wing men* begleiten. Bei Angriffen schützen sie den Piloten und lenken von ihm ab. In Kampfmissionen fliegen sie als *remote carrier* weit voraus, klären koordiniert auf und bekämpfen gegnerische Ziele. Auch andere Komponenten der Luftverteidigung werden durch FCAS als System-of-Systems vernetzt: der Eurofighter, die Militärtransporter, Lenkflugkörper oder AWACS [Airborne Early Warning and Control Systems, d.h. fliegende Radarsysteme, Anmerkung N.K.]. Damit ist klar,

(1) Alle Informationen sind der offiziellen Website der Firma IAI entnommen. Vgl.: <https://www.iai.co.il/p/harpy>.

dass Künstliche Intelligenz und Technische Autonomie bei FCAS eine Schlüsselrolle spielen müssen (Koch 2019, 8).

Es geht bei FCAS entsprechend nicht darum, eine einzelne Waffe mit Autoregulation auszustatten, wie etwa bei *IAI Harpy*, sondern es ist vielmehr das Ziel, ein sogenanntes *System of Systems* zu betreiben. Gemeint ist damit das Zusammenspiel verschiedener bemannter und unbemannter Geräte, die durch ein Netzwerk in Echtzeit miteinander in Kontakt stehen.<sup>2</sup> Dass sich solche Systeme selbstständig und ohne Bezugnahme auf menschliche Entscheidungen regulieren, ergibt vor dem Hintergrund hochkomplexer und sehr schneller Kommunikationsprozesse innerhalb des Systems durchaus Sinn. In einer Argumentation wie der oben benannten scheint es im Grunde genommen keine andere Wahl zu geben, wenn solche Systeme entwickelt werden sollen.

Allerdings sagt Wolfgang Koch wenig Konkretes darüber aus, in welcher Form »Künstliche Intelligenz« und »Autonomie« genutzt werden sollen und lässt damit die Frage offen, welche Funktionen denn eigentlich von Automatisierung betroffen sind.<sup>3</sup> Deswegen spielt die Frage nach der Definition von maschineller Autonomie bzw. Autoregulation im politischen und akademischen Diskurs auch eine zentrale Rolle. Die erste Frage lautet folglich: Welche Funktionen sollen automatisiert werden? So führt etwa der automatisierte Start- und Landeprozess bei (militärischen) Drohnen nicht zwangsläufig zu politischen und ethischen Problemen. Das, was bei der Automatisierung jedoch durchaus öffentlich kontrovers diskutiert wird, ist die Automatisierung der Tötungsfunktion. Ich schließe mich aus diesem Grund in meiner Betrachtung der Definition des Internationalen Roten Kreuzes an. Diese lautet folgendermaßen:

Autonomous Weapon Systems are defined as any weapon system with autonomy in the critical functions of target selection and target engagement. That is, a weapon system that can select (i.e. detect and identify) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention (ICRC 2016).

(2) Bei einem Zusammenspiel der gleichen Geräte lautet der *terminus technicus* Schwarm. Handelt es sich um unterschiedliche Geräte, wird von einem *System of Systems* gesprochen.

(3) Ich setzte den Terminus »Künstliche Intelligenz« hier in Anführungszeichen, da es sich ebenfalls um einen Anthropomorphismus handelt, der sich nicht ohne Weiteres auf Maschinen übertragen lässt (etwa Charbonnier 2019).

Zwei Dinge sind in dieser Definition hervorzuheben: Erstens operiert die Maschine ohne menschliche Interventionsmöglichkeit in Echtzeit und zweitens geht es dabei um kritische Funktionen wie das Töten von Menschen.

Damit stellt sich zunächst die Frage, was eigentlich genau unter maschineller Autonomie verstanden wird.

#### ⇒ 2 Autoregulation statt maschineller Autonomie

Maschinelle Autonomie kann als das Ende der Automatisierungsskala beschrieben werden. Das heißt zunächst, dass es bei der Zusammenarbeit von Mensch und Maschine sehr unterschiedliche Grade an Automatisierung geben kann. Beginnend mit Prozessen, in denen der Mensch die Maschine vollständig kontrolliert, etwa beim klassischen Autofahren, über verschiedene Zwischengrade, in denen Mensch und Maschine stärker gemeinsam agieren, etwa bei Assistenzsystemen, reicht dies bis hin zu Vorgängen, in denen der Mensch – in Bezug auf eine bestimmte Funktion – keine technisch vorgesehene Interventionsmöglichkeit mehr hat. Erst im letzten Fall, also da, wo keine Interventionsmöglichkeit mehr vorhanden ist, liegt maschinelle »Autonomie«, bzw. Autoregulation vor.

Das Wort Autonomie selbst stammt aus dem Griechischen: Zusammengesetzt aus αὐτός (*autos*) und νόμος (*nomos*), bezeichnet es die Fähigkeit, sich selbst Gesetze zu geben. Zentral für die Rede von Autonomie sind in Philosophie und Ethik die Schriften Kants, der Autonomie im Wesentlichen als die Fähigkeit des Menschen, sich selbst *moralische* Gesetze zu geben, definiert (Kant 1977, 74-75). Soll dieser Begriff aber auf maschinelle Systeme übertragen werden, stellt sich die Frage, ob dies inhaltlich zutrifft. Um das zu klären, ist zunächst zu prüfen, an welcher Stelle die Fähigkeiten von Mensch und Maschine voneinander abweichen. Noel Sharkey, Informatiker und Professor für künstliche Intelligenz und Robotik an der Universität Sheffield, etwa weist auf die wesentlichen Unterschiede in der Informationserfassung und -verarbeitung von Mensch und Maschine hin: Maschinen seien seiner Meinung nach wesentlich besser darin, Zahlen zu berechnen, große Datensätze zu durchsuchen, schnell auf Kontrollaufgaben zu reagieren und wiederkehrende Aufgaben sorgfältig auszuführen. Die Fähigkeiten des Menschen jedoch liegen stärker im Bereich des abwägenden und kreativen Denkens sowie im Bereich

der Meta-Kognition (Sharkey 2016, 27).<sup>4</sup> Das heißt zugleich, dass Computer und Menschen über sehr unterschiedliche Fähigkeiten verfügen, sodass sich grundsätzlich festhalten lässt: Aufgaben, die Vorausschau, Meta-Kognition und Kreativität erfordern, sind im Arbeitsbereich des Menschen gut aufgehoben, während Berechnungen und Mustererkennung besser vom Computersystem umgesetzt werden können. Beide Entitäten gemeinsam können so gewinnbringend miteinander arbeiten und die jeweiligen Defizite ausgleichen: Die mangelnde Rechenleistung des Menschen durch den Computer, die fehlende Umsicht der Maschine durch die kognitiven Fähigkeiten des Menschen. Allein: Eine solche Zusammenarbeit geschieht stets unter der Maßgabe, dass der Mensch die Ziele einer gemeinsamen Aktion vorgibt. Das hat deswegen Sinn, weil nur der Mensch über die notwendigen kognitiven Fähigkeiten dazu verfügt: Mag die Maschine im Einzelfall den Weg zum Ziel auf jeden Fall schneller und vielleicht sogar zweckdienlicher erfassen – eine sinnvolle Zielvorgabe für komplexe Operationen kann sie derzeit nicht liefern.<sup>5</sup>

Nehmen wir mit diesen Einsichten noch einmal auf die Frage nach maschineller Autonomie Bezug, dann geht es darum, ob die Maschine in der Lage ist, sich selbst (moralische) Gesetze zu geben – und diese Frage ist schlichtweg zu verneinen. Dies geschieht in der philosophischen Debatte über maschinelle Autonomie in der Regel auch. Wer also über maschinelle Autonomie redet, macht zumeist erst einmal deutlich, dass nicht das philosophische Verständnis gemeint ist (etwa Leveringhaus 2016, 32-33; Grünwald/Kehl 2020, 36). Dieses Vorgehen halte ich jedoch aus mindestens zwei Gründen für problematisch: Zum einen ist fraglich, ob der Versuch einer solchen Übertragung tatsächlich gelingt, oder ob dabei nicht vielmehr die philosophische Idee trotz aller Umdefinitionen erhalten bleibt, einfach weil im

(4) Noel Sharkey erwähnt in diesem Zusammenhang auch die Fähigkeit des Menschen, induktiv zu denken. Allerdings ahmen neuere algorithmische Systeme wie neuronale Netze diese Prozesse durchaus in Ansätzen nach. Ohne sagen zu wollen, dass es sich hierbei um dieselben Prozesse handelt, möchte ich aber zumindest auf die neueren Ansätze hinweisen. Ich danke Laurence Lerch und Hannah Bleher für diesen Hinweis.

(5) Dies drückt sich etwa im Unterschied zwischen judgement und reckoning aus, eine Unterscheidung, die Brian Cantwell Smith in die Debatte eingetragen hat. Das Beurteilungsvermögen (judgement) von Menschen bezieht die (soziale) Welt mit ein – so kann sich ein Mensch innerhalb der materiellen und immateriellen Welt, die ihn umgibt, zurechtfinden. Wenn eine Maschine andererseits aufgrund von Algorithmen ihren Weltzugang berechnet (reckoning), so fehlt ihr dieser ganzheitliche Weltzugang. Ihre eigenen Zielsetzungen werden deswegen nicht auf dieselbe Art und Weise die umgebende Welt mit einbeziehen können (Smith 2019). Ich danke Florian Höhne für diesen Literaturtipp.

Wort eine Bedeutung mitschwingt, die sich nicht ohne Weiteres davon trennen lässt. Die Folge wäre, dass die Implikationen von Autonomie doch auf Maschinen übertragen werden und so eine (falsche) Anthropomorphisierung von Technik vorgenommen wird. In diesem Fall unterstellt der Mensch der Maschine, über Autonomie – im philosophischen Sinne – zu verfügen. Dann verliert der Begriff also seine ethischen Implikationen trotz Ansage gerade nicht. Das wiederum würde aber dazu führen, dass Entscheidungen an Maschinen abgegeben werden, die von der Maschine nicht sinnvoll getroffen werden können (Fuchs 2020)<sup>6</sup> – etwa vorausschauende oder kreative Entscheidungen oder Entscheidungen, die Meta-Kognition erfordern. Zum anderen könnte es auch sein, dass der Begriff der Autonomie seine philosophische Spitze verliert, dann aber auf den Menschen rückübertragen wird – mit Folgen für das Menschenbild, das im Ergebnis maschinisiert wird (Koch 2016, Charbonnier 2019, 80-81). Deswegen lautet mein Vorschlag, den Begriff der Autonomie durch den der Autoregulation zu ersetzen.<sup>7</sup> Dieser Terminus stammt ursprünglich aus der Kybernetik und meint die eigenständige Regulierung eines technischen Systems. Klassisches Beispiel hierfür ist das Thermostat, bei dem das System die Raumtemperatur entsprechend vorgegebener Parameter selbstständig anpasst (Heylighen/Joslynn 2001, 12). Diese sehr grundlegende Bedeutung ist dann zu erweitern um die Fähigkeit zur dynamischen Anpassung an sich verändernde Umwelteinflüsse – also die selbstständige Regulierung in der realen Umwelt. Auch wenn diese Umdeutung das Verständnis innerhalb der Kybernetik übersteigt, sehe ich den Vorteil darin, dass die moralischen Implikationen, die sich mit dem Begriff der Autonomie von vornherein verbinden, vermieden werden – die Anpassungsfähigkeit des Algorithmus bleibt damit auf einer technischen Ebene, ohne dass die immensen Möglichkeiten und Chancen verleugnet werden.<sup>8</sup>

(6) Mein Dank geht an Tobias Friesen für den Literaturhinweis.

(7) Ich entnehme diesen Begriff einem Beitrag von Lucy Suchman und Jutta Weber, die zunächst zwischen biologischen und später technischen Entwicklungen als »self-regulation«, und dem philosophischen Terminus »autonomy« unterscheiden (Suchman/Weber 2016, 79-80).

(8) Die hier vorgeschlagene Umdeutung innerhalb des Begriffs der Autoregulation ist sicher nicht unproblematisch, zumal der Begriff innerhalb der naturwissenschaftlichen Fächer bereits vorgeprägt ist. Dennoch sehe ich den Vorzug einer stärker technisch orientierten Redeweise als so zentral an, dass ich ihn zumindest für meine eigenen Beiträge nutzen werde, um sog. maschinelle Autonomie zu beschreiben.

Damit kann zugleich der inhaltliche Unterschied auch sprachlich markiert werden: Zu unterscheiden sind autonome und autoregulative Entitäten dann vorrangig in Bezug auf ihre Fähigkeit zur eigenen Zielsetzung. So halten Cliff Joslynn und Francis Heylighen in Bezug auf kybernetische Systeme fest: »Artificial systems, such as thermostats and automatic pilots, are not autonomous: their primary goals are constructed in them by their designers« (Heylighen/Joslynn 2001, 12). Während autonome Akteure also in der Lage sind, sich selbst Ziele zu setzen, und zwar, mit Verweis auf die philosophische Dimension gerade auch moralische Ziele, so sind autoregulative Entitäten zwar fähig, selbstständig Ziele zu verfolgen, diese jedoch werden vom System nicht selbst gesetzt.

Damit im Zusammenhang ist des Weiteren auf das Problem der nicht-intendierten Nebeneffekte hinzuweisen, auf das Nick Bostrom anschaulich mit dem Gedankenexperiment über das Büroklammer-System aufmerksam gemacht hat: Bostrom stellt sich dabei ein System vor, dessen einziges Ziel es ist, Büroklammern herzustellen. Aufgrund seiner Konstruktion ist es jeder Einflussnahme von außen entzogen, also autoregulativ. Es könnte nun sein, dass die Maschine anfängt, wortwörtlich alles auf dieser Welt zu Büroklammern zu verarbeiten, inklusive Mensch und Umwelt. Das Problem des Systems nämlich ist, dass es keine Vorstellung davon hat, was Leben ist. Umso mehr könnte die selbstständige Zielsetzung von selbstlernenden Systemen hochproblematische Konsequenzen nach sich ziehen. Was, beispielsweise, wenn ein solches System »lernt«, dass Menschen das Problem sind, etwa bei der Anwendung von Gewalt, und sich gegen uns richtet? Deswegen, so meine These, braucht es stets einen autonomen Akteur, der das System kontrolliert – eine Tatsache, auf die der Begriff der Autoregulation bereits terminologisch hinweist. Damit möchte ich an dieser Stelle ein erstes Fazit ziehen: *Da auch die autoregulative Maschine auf einen autonomen Akteur angewiesen bleibt, muss menschliche Kontrolle in ausreichendem Maße gewährleistet werden, denn nur der Mensch ist in der Lage, autonom zu entscheiden und die moralischen Implikationen dieser Entscheidungen zu reflektieren.* 3

⇒ 3 Der Stand der politischen Debatte: *meaningful human control*

Die Frage danach, ob und wenn ja, in welchem Maße menschliche Kontrolle von Waffensystemen mit autoregulativen Funktionen notwendig und sinnvoll ist, wird politisch seit 2014 auf Ebene der UN

diskutiert, ohne dass bisher eine substantielle Einigung erzielt werden konnte. Ich möchte an dieser Stelle nicht auf die politischen Argumente und Akteure eingehen – die Dringlichkeit einer politischen Regulierung versteht sich von selbst.<sup>9</sup> Ich werde mich in meiner Darstellung stattdessen auf die vorgebrachten ethischen Argumente konzentrieren, die im derzeitigen Diskurs eher am Rande vorkommen, zumeist eingebunden in politische und juristische Argumentationen. Dabei lassen sich im Wesentlichen vier Argumentationsstränge extrapolieren, die bisher eine Rolle gespielt haben. Zunächst wurden die Vorteile eines solchen Systems betont, auf die dann mit dem Argument einer Verantwortungslücke, sowie dem der menschlichen Handlungsfähigkeit und der Menschenwürde reagiert wurde.<sup>10</sup>

Zum ersten: Besonders vom Roboteringenieur Ronald C. Arkin wurde und wird argumentiert, dass Autoregulation in Waffensystemen helfen könne, Schädigungen auf beiden Seiten des Konflikts zu verringern, weil keine emotionalen Regungen wie Wut, Frust noch Angst sie ablenken. Dazu gehöre auch, dass solche Waffen nicht schon aus sicherer Distanz schießen müssten, etwa aus Furcht um ihr eigenes Leben. Sie könnten eine potenzielle Gefahr auf deutlich nähere Distanz herankommen lassen und so ihre Bedrohung gleichsam besser bewerten (Arkin 2009). Dass das Leben und nicht zuletzt die psychische Gesundheit von Soldaten und Drohnenpiloten aus den eigenen Reihen geschont würden, liegt auf der Hand, da diese weder physisch noch psychisch auf dem Schlachtfeld auftauchen (Galliot 2016, 85). Gegen diese Haltung wenden sich Ethiker:innen, die auf eine Verantwortungslücke hinweisen. So führt der Philosoph und Ethiker Robert Sparrow an, dass Autoregulation notwendigerweise zu einer Verantwortungslücke führe, weil letztlich unklar bleibe, wer für einen möglichen Fehler zur Rechenschaft zu ziehen sei: Programmierer:in, Kommandant:in oder Soldat:in, der:ie den Einsatzbefehl gegeben hat, könnten zwar rechtlich belangt werden, würden im Ernstfall jedoch für Handlungen zur Rechenschaft gezogen werden, die außerhalb ihrer Kontrolle liegen (Sparrow 2007). In Abgrenzung dazu brachte der Philosoph und Ethiker Alex Leveringhaus die Notwendigkeit menschlicher Handlungsfähigkeit in die Debatte ein: Während eine Maschine

(9) Für weitere Informationen zu diesem Thema verweise ich auf die Seiten des International Panel on the Regulation of Autonomous Weapons: [www.irpaw.org](http://www.irpaw.org), sowie auf die Seiten der UN: <https://meetings.unoda.org/meeting/ccw-gge-2021/>.

(10) Erstmals wurde auf die Möglichkeit einer Verantwortungslücke bereits 2007 von Robert Sparrow hingewiesen. Die ersten beiden von mir aufgegriffenen Wellen beschreibt Alex Leveringhaus (2016).

stets den programmierten Pfaden folgen werde, habe ein Mensch die Möglichkeit, anders zu reagieren und Milde walten zu lassen. Leveringhaus schreibt dazu:

Unless re-programmed, the machine will engage the targeted person upon detection. Killing a person, however, is a truly existential choice that each soldier needs to justify before his own conscience. Sometimes it can be desirable not to pull the trigger, even if this means that an otherwise legitimate target survives. Mercy and pity may, in certain circumstances, be the right guide to action (Leveringhaus 2016, 92).

Zudem wird im gegenwärtigen Diskurs immer wieder auf die Menschenwürde der Zielperson verwiesen: Weil die Maschine den anvisierten Menschen zu einem Datenpunkt reduziere, werde ihre Menschenwürde verletzt. Folglich sei die Tötung durch Maschinen moralisch unzulässig (Heyns 2016; Rosert/Sauer 2019).

Neben diesen ethischen Argumenten wird eine Reihe von (sicherheits-)politischen Fragestellungen diskutiert, wie etwa die Möglichkeit eines neuen Wettrüstens im Bereich künstlicher Intelligenz (Altmann/Sauer 2017) oder aber die juristische Frage, ob das internationale Rechtssystem, insbesondere das humanitäre Völkerrecht ausreichend sei, um Autoregulation in Waffensystemen zu regulieren (Geiss/Lahmann 2017, 378-383, 399). Insgesamt jedoch zeichnet sich die Wichtigkeit menschlicher Kontrolle ab, etwa in Form von *meaningful human control*. Dieser Begriff wurde von der NGO Article 36 in die Debatte eingebracht und in diesem Zusammenhang von Heather M. Roff und Richard Moyes genauer bestimmt. Ihrer Definition nach bedeute

meaningful human control (...) 1. That a machine applying force and operating without any human control whatsoever is broadly considered unacceptable. 2. That a human simply pressing a ›fire‹ button in response to indications from a computer, without cognitive clarity or awa-

recess, is not sufficient to be considered ›human control‹ in a substantive sense (Roff/Moyes 2016, 1).<sup>11</sup>

Die Autor:innen nehmen so zunächst eine negative Abgrenzung vor, welche die Anwendung von Autoregulation im oben genannten Sinne deutlich einschränkt, da bei Autoregulation ja gerade keine menschliche Kontrolle in Echtzeit mehr vorgesehen ist. Diese Definition ist zwar auf internationaler Bühne nicht unumstritten, findet jedoch juristisch, politisch und ethisch großen Anklang (Amoroso/Tamburrini 2021). Die Notwendigkeit menschlicher Einflussnahme auf das System ist zu unterstreichen mit dem Verweis darauf, dass auch Maschinen keineswegs unparteiische Entscheidungen treffen und sich so kulturelle Vorurteile bis hin zu Fehlern und nicht-intendierten Nebeneffekten in die Beurteilung einschleichen können (O’Neil 2017; Noble 2019). Um solche Probleme zu vermeiden, ist es unumgänglich, dass der Mensch die Maschine kontrolliert – unter Bedingungen, die genügend Handlungsspielraum lassen.

#### ⇒ 4 Anbindung an die Friedensethik

Alle bisher angesprochenen Diskussionen, politische wie ethische, bewegen sich im Rahmen eines sicherheitspolitischen Zugriffs, der in der Regel auf Grundlage der Lehre vom gerechten Krieg argumentiert. Diese ursprünglich im antiken Recht verortete Idee wurde durch die lateinisch-christliche Tradition transformiert und fand im 20. Jahrhundert einen Platz im Völkerrecht der Vereinten Nationen (Jensen 2015). Zwar kann diese Idee als ein »Krieg um des Friedens willen« rekonstruiert werden, wie dies beispielsweise Jessica Jensen in ihrer juristischen Analyse aufzeigt, doch haben die christlichen Kirchen in den letzten drei Jahrzehnten eindringlich darauf hingewiesen, dass eine friedensethische Neuorientierung nötig sei, die stärker noch den Frieden fokussiert (Hoppe/Werkner 2017). Unter dem Leitspruch *Si vis pacem para pacem* hat die Evangelische Kirche in Deutschland 2007 mit ihrer Friedensdenkschrift diese Neuorientierung vorgenommen, indem der äußerste Vorrang ziviler Konfliktbearbeitungsmechanismen und eines friedlich ausgerichteten internationalen Rechtssystems betont wird, bei dem die Kriterien zur Beurteilung des gerechten Krieges nur innerhalb der Ethik rechtswahrender Gewalt vorgestellt

(11) Dieser Terminus wurde in den letzten Jahren auch auf andere Ebenen des Einsatzes von künstlich intelligenten Systemen übertragen (Santoni de Sio/van den Hoven 2018), beispielsweise selbstfahrende Autos (Heikoop 2020).

werden.<sup>12</sup> Diese jedoch kommen nur als äußerstes Mittel in Frage, in Grenzsituationen und um den Rechtszustand zu wahren. Der Einsatz von Waffengewalt wird also überhaupt nur erwogen, um eine Rechts-situation (wieder-)herzustellen, aus der dann wiederum gerechter Friede erwachsen kann (EKD 2007, 66). Die herkömmlichen Kriterien der Tradition des gerechten Krieges – also sowohl das Recht zum Krieg, *ius ad bellum* als auch das Recht im Krieg, also *ius in bello* – bleiben hier als Prüfkriterien bestehen. Konkret bedeutet dies, dass in Erlaubnisgrund, Autorisierung, Absicht, der Gewalteinsatz als äußerstes Mittel, die Verhältnismäßigkeit seiner Folgen, sowie die Verhältnismäßigkeit der Mittel und die Anwendung des Unterscheidungsprinzips zu berücksichtigen sind (EKD 2007, 68-69). Damit ist zugleich der Schritt in Richtung eines bedingten Pazifismus getan. Ein solcher bewegt sich zwischen einer absolut pazifistischen Position, die einem jeglichen Gewalteinsatz ablehnend gegenübersteht, und einem gerechten Krieg, welcher Krieg als moralisch legitimes Mittel ansieht, so er denn in der richtigen Weise geführt wird.

In meiner Arbeit frage ich nach der ethischen Legitimation autoregulativer Funktionen in Waffensystemen innerhalb eines solchen bedingten Pazifismus. Dies verändert in erster Linie die Fragerichtung: Anstelle der Problematik, ob Autoregulation in kritischen Funktionen in Einklang mit geltendem Recht und ethischen Normen genutzt werden kann, tritt die Frage, ob und inwiefern solche Systeme einem als gerecht qualifizierten Frieden zuträglich sind. Um einen solchen Frieden zu definieren, benennt die Friedensdenkschrift vier Dimensionen, »die *Vermeidung von Gewaltanwendung*, die *Förderung von Freiheit und kultureller Vielfalt* sowie (...) den *Abbau von Not*« (EKD 2007, 54). Damit ist zugleich gesagt, dass Frieden zwar die Abwesenheit von Gewalt beinhaltet, aber in seiner positiven Ausformung zugleich darüber hinausgeht und als prozesshaftes Geschehen vorzustellen ist, hin zu einer immer stärkeren Ausbildung von Strukturen, die gerechten Frieden auf Dauer zu stellen vermögen.

Ist aber die Etablierung nachhaltiger Friedensprozesse das erste und einzige Ziel eines Gewalteinsatzes, setzt dies voraus, dass die konfligierenden Parteien sich so verhalten, dass durch ihre Kriegsführung zugleich die Basis für einen nachhaltigen Frieden gelegt wird. Die

(12) Ich verwende mit Reuter den Begriff der rechtswahrenden Gewalt anstelle rechtserhaltender Gewalt, wie es in der Denkschrift heißt. Reuter weist darauf hin, dass der Begriff der rechtswahrenden Gewalt semantisch offener sei und nicht nur die Erhaltung, sondern auch die Ermöglichung und Durchsetzung basaler Menschenrechte umfasse (Reuter 2014, 39).

Entscheidung für Waffengewalt und die Art der Kriegsführung werden sich also an diesem Ziel messen lassen müssen. Deswegen scheint mir eine bisher noch zu wenig beachtete Kategorie im derzeitigen Diskurs das *jus post bellum* zu sein, von dem her sich *ius in bello* und *ius ad bellum* aufschlüsseln. Gemeint sind damit moralische Normen, die Stabilität nach einem Krieg (wieder) ermöglichen sollen. In Worten des Philosophen Larry May: »Jus post bellum normally concerns how to move to a situation of stability after war« (May 2012, 6). Konkret beinhaltet dies unter anderem die Fähigkeit, Kriegsverbrecher für ihr Handeln zur Rechenschaft zu ziehen und eine Versöhnung der konfligierenden Parteien herzustellen (May 2012, 1).<sup>13</sup>

Vor diesem Hintergrund aber ergeben sich mit Blick auf Autoregulation in Waffensystemen in mindestens zwei Bezugspunkten Fragen: Zunächst betrifft dies die Fortsetzung eines Trends, der sich bereits für Drohnen und deren desaströse Wirkung auf die Zivilbevölkerung nachweisen lässt (International Human Rights and Conflict Resolution Clinic 2012). Vor allem die Beeinträchtigung des täglichen Lebens von Zivilisten (ebd., VII) dürfte sich fortsetzen, wenn diese nicht mehr von Menschenhand gesteuert werden. Hinzu käme nämlich die Furcht vor einem Fehler, der dann deutlich fataler ausfiele als bei einem Menschen, weil ein menschlicher Operateur denselben Fehler nicht in atemberaubender Geschwindigkeit wiederholt (Scharre 2016, 6). Ist jedoch ein solcher Fehler erst einmal aufgetreten und lassen sich Verantwortlichkeiten im Anschluss nicht sinnvoll rekonstruieren, dürfte dies der Ausbildung langfristiger Friedensstrukturen zuwiderlaufen. Entsprechend ist abzusehen, dass der Einsatz solcher Technik langfristigen Frieden eher verhindern als befördern könnte. Und zwar selbst dann, wenn der Einsatz von Autoregulation zur Reduktion von Schädigungen führte.

Darüber hinaus spielt die Komponente des technischen Risikos im Zusammenhang mit Autoregulation in Waffensystemen eine nicht zu unterschätzende Rolle. Dies trifft vor allem dann zu, wenn die Maschine fehlerhaft operiert oder aber ihr ordnungsgemäßes Funktionieren nicht im Sinne der Einsetzenden liegt, wie dies oben bereits am

(13) Auch wenn die Friedensdenkschrift diese Ideen nicht direkt unter dem Stichwort eines *ius post bellum* ausweist, sind diese Gedanken maßgeblich vorhanden und stecken den Rahmen der Denkschrift ab, da sie auf Frieden hin orientiert ist. Ich halte dennoch die explizite Bezugnahme auf diese Debatten sinnvoll, weil die Fachdebatte unter diesem Schlagwort diskutiert wird, aber auch weil so der friedensethische Rahmen vom Ende her, also aus Sicht eines nachhaltigen Friedens, pointiert gedacht werden kann.

Büroklammer-Gedankenexperiment deutlich wurde. Dieses Risiko aber wird im Kriegseinsatz vornehmlich auf die Gegenseite verlagert. Damit setzt sich zugleich ein Trend fort, der bereits für Drohnen zu beobachten ist und den der amerikanische Jurist Paul W. Kahn als »Paradox risikoloser Kriegsführung« (paradox of riskless warfare; Kahn 2002) bezeichnet hat: Während die Seite, die Autoregulation nutzt, sich keinerlei Risiko mehr aussetzt, nimmt das Risiko für die Gegenseite sogar noch zu – ein Thema, das beispielsweise Hans-Richard Reuter für die bewaffneten Drohnen bereits kritisch beleuchtet hat (Reuter 2014). Wird nun aber das Recht, im Krieg zu töten, über das Selbstverteidigungsrecht der menschlichen Soldat:innen rekonstruiert, wirft der Einsatz letaler autoregulativer Waffensysteme enorme Zweifel auf.<sup>14</sup>

Mit dem Begriff des Risikos kann wiederum aus friedensethischer Perspektive zugleich das Problem des Kombattantenstatus erschlossen werden, wie dies beispielsweise bei Larry May geschieht: Seiner Meinung nach ist eine Beteiligung an gegenwärtigen Kriegen problematisch, weil etwa in asymmetrischen Konflikten das Risiko zunimmt, Personen zu schädigen, die letztlich ungerechtfertigt anvisiert werden. Dies gilt umso mehr, wenn die Annahmen der Revisionisten innerhalb der Tradition des gerechten Krieges zu Grunde gelegt werden. In dieser Adaptierung der Tradition des gerechten Krieges verläuft die Grenze zwischen legitimem und illegitimem Ziel nicht mehr statisch zwischen Kombattant:in und Nichtkombattant:in, wie es etwa das Humanitäre Völkerrecht vorsieht, sondern wird innerhalb der involvierten Gruppen vorgenommen, sodass nicht jede:r Soldat:in ein legitimes Ziel und nicht jede Zivilperson ein illegitimes Ziel darstellt, da die Verstrickungen in Kriege oft vielschichtiger sind, als eine Uniform dies vorgibt (May 2015; McMahan 2009). Eine solche Unterscheidung jedoch macht Kriegsführung nicht nur moralisch risikoreicher, sondern auch komplexer, da die einzelne Tötungsentscheidung deutlich situationsabhängiger ausfallen dürfte als herkömmlich. Es ist jedoch höchst zweifelhaft, ob ein autoregulatives System diesen Bedingungen schon allein aus technischer Sicht gerecht werden kann.

#### ⇒ 5 Fazit

Es steht also zu erwarten, dass der Einsatz von Autoregulation in Waffensystemen aus ethischer Perspektive durchaus problematisch

(14) Dass diese Herleitung keineswegs unumstritten und selbstverständlich ist, hat David Rodin nachgewiesen (2002).

ist. Dies bedeutet zugleich nicht zwangsläufig, dass jede Verwendung von Autoregulation – wie etwa Funktionen, die nicht den Tod eines Menschen intendieren – in Gewaltkontexten grundsätzlich abzulehnen ist. Zu überlegen ist auch, ob es »einfache« Szenarien geben kann, in denen sich der Einsatz einer solchen Waffe sicher gestalten lässt, etwa im sehr tiefen Meer oder aber innerhalb eines klar abgegrenzten Luftraumes. Jedoch stellt sich auch hier die Frage nach dem Status der Kombattant:innen und der Situationserwägung. Zu bedenken wäre etwa eine Art »Autoregulationsskala«, wie sie De Greef für Hochrisikoszenarien vorgeschlagen hat, etwa bei der Suche nach Überlebenden in Folge eines Erdbebens (De Greef 2012). In solchen Situationen erweist sich die Zusammenarbeit von Mensch und Maschine dann als besonders effektiv, wenn das Automatisierungsniveau der Maschine dynamisch an den menschlichen Part angepasst wird. Ein solches Verfahren der Adaptive Automation, bzw. Adaptable Automation nimmt in Abhängigkeit von der Arbeitsbelastung der Operateur:innen eine Anpassung des Automatisierungsniveaus vor. 15 So kann bei Überforderung des menschlichen Parts problemlos das Automatisierungslevel angehoben oder bei weniger Arbeitsbelastung dieses wieder abgesenkt werden. In einem solchen Rahmen, etwa bei Totalausfall des menschlichen Parts, ist auch eine Umsetzung von Autoregulation in einem sehr beschränkten Maße durchaus vorstellbar. Grundsätzlich jedoch ist, sowohl aus technischen, als auch als friedensethischen Gründen, die Umsetzung einer »bedeutungsvollen menschlichen Kontrolle« zu befürworten: Nur der Mensch hat die Fähigkeit, Situationen umfassend zu bewerten, und nur durch ausreichenden Einfluss des Menschen auf die Maschine lässt sich verhindern, dass nicht mehr der Mensch die Maschine, sondern die Maschine den Menschen kontrolliert – eine Situation, die besonders im Hinblick auf letale Gewalt unbedingt zu vermeiden ist.

(15) Beide Vorschläge unterschieden sich in der Frage, wer die Anpassung vornimmt. Während Adaptive Automation eine automatische Anpassung auf Initiative des maschinellen Parts hin vornimmt, so wird diese Anpassung beim Konzept der Adaptable Automation vom Menschen vorgenommen. Beide Systeme haben jeweils Vor- und Nachteile. So ist etwa auf den Kontrollverlust, der mit einer automatischen Adaptierung einhergeht hinzuweisen, während sich andererseits kritisch rückfragen lässt, ob ein bereits überforderter Mensch noch in der Lage ist hinreichend über eine Abgabe an Kompetenzen an die Maschine zu entscheiden – zumal eine solche Situation die Arbeitsbelastung wieder erhöht (DeGreef 2012; Scerbo 2019).

## ⇒ Literaturverzeichnis

Altmann, Jürgen; Sauer, Frank (2017): Autonomous Weapons Systems and Strategic Stability, in: *Survival*, 5/59, 117–142.

Amoroso, Daniele; Tamburrini, Guglielmo (2021): Toward a Normative Model of Meaningful Human Control over Weapons Systems, in: *Ethics & International Affairs*, 2/35, 245–272.

Arkin, Ronald C. (2009): *Governing Lethal Behavior in Autonomous Robots*, Boca Raton: CRC Press.

Charbonnier, Ralph (2019): Wahrnehmen, entschieden, handeln – werden digitale Maschinen menschlich?, in: Görder, Björn; Zeyher-Quattlender, Julian (Hg.): *Daten als Rohstoff. Die Nutzung von Daten in Wirtschaft, Diakonie und Kirche aus ethischer Sicht*, Münster: LIT, 61–82.

Evangelische Kirche in Deutschland (EKD, 2007): *Aus Gottes Frieden leben – für gerechten Frieden sorgen. Eine Denkschrift des Rates der Evangelischen Kirche in Deutschland*. Gütersloh: Gütersloher Verlagshaus.

Fuchs, Thomas (2020): Menschliche und Künstliche Intelligenz. Eine Klarstellung, in: Ders.: *Verteidigung des Menschen. Grundfragen einer verkörperten Anthropologie*. Berlin: Suhrkamp, 21–70.

Galloit, Jai (2016): Lethal Autonomous Weapons Systems. Proliferation, Disengagement, Disempowerment, in: German Federal Foreign Office: *Lethal Autonomous Weapons Systems Technology, Definition, Ethics, Law & Security*, Frankfurt M.: Druck- und Verlagshaus Zarbock GmbH & Co. KG, 85–96.

Geiss, Robin; Lahmann, Henning (2017): Autonomous weapons systems. A paradigm shift for the law of armed conflict?, in: Ohlin, Jens David (Hg.): *Research handbook on remote warfare*. Cheltenham/Northampton: Edward Elgar, 371–404.

De Greef, Tjerk (2012): *ePartners for Dynamic Task Allocation and Coordination*, Delft.

Grünwald, Reinhard; Kehl, Christoph (2020): *Autonome Waffensysteme. Endbericht zum TA-Projekt*, Bad Honnef: Wienands Print + Medien GmbH.

Heikoop, Daniël D. (2020): Meaningful human control. Designing safety into automated driving systems, in: Research OUTREACH, 114. Verfügbar unter: <https://researchoutreach.org/articles/meaningful-human-control-designing-safety-into-automated-driving-systems>. Eingesehen am 22.11.21.

Heylighen, Francis; Joslynn, Cliff (<sup>3</sup>2001): Cybernetics and second-order cybernetics, in: R. Meyers (Hg.): Encyclopedia of Physical Science and Technology, New York: Academic Press.

Heyns, Christopher (2016): A Human Rights Perspective on Autonomous Weapons in Armed Conflict. The Rights to Life and Dignity, in: Federal Foreign Office of Germany: Lethal Autonomous Weapons Systems. Technology, Definition, Ethics, Law & Security Frankfurt M.: Druck- und Verlagshaus Zarbock GmbH & Co. KG, 148–159.

Hoppe, Thomas; Werkner, Ines-Jacqueline (2017): Der gerechte Frieden. Positionen in der katholischen und evangelischen Kirche, in: Werkner, I.-J.; Ebeling, K. (Hg.): Handbuch Friedensethik, Wiesbaden: Springer VS, 344–359.

ICRC, Views of the ICRC on autonomous weapon systems, paper submitted to the Convention on Certain Conventional Weapons Meeting of Experts on Lethal Autonomous Weapons Systems (LAWS), 11. April 2016. Verfügbar unter: <https://www.icrc.org/en/document/views-icrc-autonomous-weapon-system>. Eingesehen am 25.11.2021.

International Human Rights and Conflict Resolution Clinic at Stanford Law School and Global Justice Clinic at NYU School of Law (2012): Living under Drones. Death, Injury, and Trauma to Civilians from the US Drone Practices in Pakistan.

Jensen, Jessica (2015): Krieg um des Friedens willen. Zur Lehre vom gerechten Krieg, Baden-Baden: Nomos

Kahn, Paul (2002): The Paradox of Riskless Warfare, in: Philosophy and Public Policy Quarterly, 22, 2–7.

Kant, Immanuel (1977): Werke in zwölf Bänden. Band 7, Frankfurt M.: Suhrkamp.

Koch, Bernhard (2016): Maschinen, die uns von uns selbst entfremden. Philosophische und ethische Anmerkungen zur gegenwärtigen Debatte um autonome Waffensysteme, in: Militärseelsorge. Dokumentation 54, 99–119.

- Koch, Wolfgang (2019): FCAS – Herausforderungen für Sensordatenfusion und Ressourcenmanagement, in: mpc Sonderheft, 8–11.
- Leveringhaus, Alex (2016): Ethics and Autonomous Weapons, London: Palgrave Pivot.
- Loh, Janina (2019): Roboterethik. Eine Einführung, Berlin: suhrkamp taschenbuch wissenschaft.
- McMahan, Jeff (2009): Killing in War, Oxford: Oxford University Press.
- May, Larry (2012): After War Ends. A Philosophical Perspective, New York: Cambridge University Press.
- May, Larry (2015): Contingent Pacifism. Revisiting Just War Theory. New York: Cambridge University Press.
- Misselhorn, Catrin (2018): Grundfragen der Maschinenethik, Ditzingen: reclam.
- Noble, Safiya Umoja (2018): Algorithms of Oppression. How Search Engines Reinforce Racism. New York: New York University Press.
- O’Neil, Cathy (2017): Weapons of Math Destruction. Harlow, England: Penguin Books.
- Reuter, Hans-Richard (2014): Kampfdrohnen als Mittel rechtswahrender militärischer Gewalt? Aspekte einer ethischen Bewertung, in: epd-Dokumentation 49, 37–46.
- Rodin, David (2002): War and Self-Defense. Oxford: Oxford University Press.
- Roff, Heather M.; Moyes, Richard (2016): Meaningful Human Control. Artificial Intelligence and Autonomous Weapons. Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems, UN Convention on Certain Conventional Weapons.
- Rosert, Elvira; Sauer, Frank G. C. (2019): Prohibiting Autonomous Weapons. Put Human Dignity First, in: Global Policy 3/10, 370–375.
- Santoni de Sio, Filippo; van den Hoven, Jeroen (2018): Meaningful Human Control over Autonomous Systems. A Philosophical Account, in: Frontiers in Robotics and AI 5, 1–14.
- Scerbo, Mark. (2019): Theoretical Perspectives on Adaptive Automation, in: Mouloua, Mustapha; Hancock, Peter A.; Ferraro James (Hg.):

Human Performance in Automated and Autonomous Systems, 103–126.

Scharre, Paul (2016): Autonomous Weapons and Operational Risk. Verfügbar unter: <https://www.cnas.org/publications/reports/autonomous-weapons-and-operational-risk>. Eingesehen am 22.11.2021.

Sharkey, Noel (2016): Staying in the loop. Human supervisory control of weapons, in: Bhuta, N.; Beck, S. et al. (Hg.): Autonomous Weapons Systems. Law, Ethics, Policy, Cambridge: Cambridge University Press, 23–38.

Smith, Brian Cantwell (2019): The Promise of Artificial Intelligence. Reckoning and Judgement, Cambridge/London: The MIT Press.

Sparrow, Robert (2007): Killer Robots, in: Journal of Applied Philosophy 24, 62–77.

Suchman, Lucy; Weber, Jutta (2016): Human-machine autonomies, in: Bhuta, N.; Beck, S. et al. (Hg.): Autonomous Weapons Systems. Law, Ethics, Policy, Cambridge: Cambridge University Press, 75–102.

Online-Quellen:

<https://www.iai.co.il/p/harpy>. Eingesehen am 19.11.2021.

<https://www.ipraw.org>. Eingesehen am 19.11.2021.

<https://meetings.unoda.org/meeting/ccw-gge-2021>. Eingesehen am 19.11.2021.

---

---

**Zitationsvorschlag:**

Kunkel, Nicole (2021): Autoregulative Waffensysteme. Automatisierung als friedensethische Herausforderung – ein Werkstattbericht. (Ethik und Gesellschaft 2/2021: Friedensethik und Geopolitik). Download unter: <https://dx.doi.org/10.18156/eug-2-2021-art-6> (Zugriff am [Datum]).

---

---



**ethikundgesellschaft**  
**ökumenische zeitschrift für sozialetik**

**2/2021: Friedensethik und Geopolitik**

Peter Rudolf: Ein neuer ›kalter Krieg‹? Friedensethisch relevante geopolitische Trends

Wolfgang Huber: Streit um den gerechten Frieden – Aktuelle Herausforderungen der Friedensethik

Bernhard Koch: Die kirchliche Friedensdebatte – Beobachtungen aus philosophischer Sicht

Julian Zeyher-Quattlender: Wieviel Gewaltfreiheit verträgt der Gerechte Frieden? Zur gegenwärtigen Debatte um Aufbrüche jenseits der Rechtsethik innerhalb der evangelischen Friedensethik in Deutschland

Max Weber: To Hack Back or Not? Eine friedensethische Analyse von Cyberoperationen vor dem Hintergrund des Leitbilds des Gerechten Friedens

Nicole Kunkel: Autoregulative Waffensysteme. Automatisierung als friedensethische Herausforderung – ein Werkstattbericht