

## **Gaza, künstliche Intelligenz und Todeslisten**

Eine der größten Herausforderungen in der Kriegsführung ist die Identifizierung militärischer Ziele. Die israelische Armee hat ein auf künstlicher Intelligenz basierendes System namens „Lavender“ entwickelt, das diesen Prozess automatisiert, indem es riesige Mengen an Überwachungsdaten durchforstet und anhand von Mustern in diesen Daten mögliche Kämpfer der Hamas oder des Islamischen Dschihad (PIJ) identifiziert. Dieser Ansatz verspricht eine schnellere und genauere Zielerfassung; jedoch haben Menschenrechtsorganisationen wie Human Rights Watch (HRW) und das Internationale Komitee vom Roten Kreuz (IKRK) vor Defiziten bei der Rechenschaftspflicht für Verstöße gegen das humanitäre Völkerrecht (IHL) gewarnt und argumentiert, dass bei diesen halb- oder sogar vollautomatisierten Systemen bei den menschlichen Offizieren eine gewisse „Routinisierung“ eintritt, die „die Notwendigkeit der Entscheidungsfindung“ verringert und die lebenswichtige Bedeutung der Entscheidung verschleiert. Zudem sind militärische Befehlshaber, die die Verantwortung für fehlerhafte Zielauswahl (Verstöße gegen das humanitäre Völkerrecht) tragen, möglicherweise nicht mehr in der Lage, die algorithmische „Black Box“, die sie berät, zu überwachen.

Im Folgenden werden wir diese Bedenken untersuchen und aufzeigen, wie die Verantwortung für Verstöße gegen das humanitäre Völkerrecht weiterhin einem Staat zuzuschreiben ist, der automatisierte oder halbautomatisierte Systeme in der Kriegsführung einsetzt. Dabei werden wir darlegen, dass das bestehende humanitäre Völkerrecht gut gerüstet ist, um mit den Herausforderungen umzugehen, die sich aus den neuen technologischen Möglichkeiten ergeben.

### **KI in der Kriegsführung – Vorteile und Risiken**

Die Vorteile von KI in der Kriegsführung sind im Wesentlichen dieselben wie in jedem anderen Bereich. Aufgrund ihrer Fähigkeit, enorme Datenmengen sehr schnell zu verarbeiten, Muster in Daten zu erkennen und diese Erkenntnisse auf neue Datensätze anzuwenden, verspricht KI eine deutliche Steigerung der Geschwindigkeit, Genauigkeit und Effizienz militärischer Entscheidungsfindung. Somit bietet KI nicht nur Vorteile für Militärangehörige, die relevante Angriffsziele identifizieren wollen, sondern auch für den Schutz der Zivilbevölkerung. Bei richtiger Programmierung und Nutzung sind KI-Systeme in der Lage, geschützte zivile Strukturen genauer und schneller zu kennzeichnen als menschliche Offiziere und präzisere Angriffe zu planen und durchzuführen, um zivile Opfer zu reduzieren. Eine Verringerung der menschlichen Beteiligung an der Entscheidungsfindung kann ebenfalls zum Schutz der Zivilbevölkerung beitragen, indem die Quelle unbeabsichtigter menschlicher Voreingenommenheit beseitigt wird.

Allerdings haben KI-Systeme einen Grad an Ausgereiftheit und Komplexität erreicht, der es Menschen oft unmöglich macht, die Gründe für ihre Bewertungen nachzuvollziehen, weshalb diese Systeme oft als „Black Boxes“ bezeichnet werden. Dies führt zu der Sorge, dass menschliche Bediener sich der Verantwortung entziehen könnten, indem sie behaupten, sie hätten keine sinnvolle Kontrolle über die Maschine ausüben können und könnten daher nicht für deren Entscheidungen zur Rechenschaft gezogen werden. Diese Verlagerung der rechtlichen Verantwortung vom Menschen auf die Software bezeichnen Organisationen wie Human Rights Watch als „Verantwortungslücke“. Um es ganz offen zu sagen: Wir glauben nicht, dass eine solche Verantwortungslücke existiert. Um Bedenken hinsichtlich KI-bedingter „Verantwortungslücken“ auszuräumen und aufzuzeigen, wie Verantwortung dennoch zugewiesen werden kann, veranschaulicht dieser Artikel zunächst anhand des israelischen Systems „Lavender“, das im aktuellen Gaza-Krieg eingesetzt wird, wie ein solches System funktioniert, bevor er sich einer eingehenden Analyse der Verantwortung für dessen Empfehlungen nach dem humanitären Völkerrecht zuwendet.

## **Automatisiertes Entscheidungshilfetooll zur Zielauswahl – „Lavender“**

Laut einem Bericht des +972 Magazine und von Local Call hat Israel dieses System genutzt, um Zehntausende potenzieller Ziele der Hamas und des Islamischen Dschihad in Gaza zur Eliminierung zu markieren.

Das System wurde mit Daten über bekannte Hamas-Mitglieder gefüttert und aufgefordert, gemeinsame Merkmale unter ihnen zu finden. Solche Merkmale könnten die Mitgliedschaft in bestimmten Chatgruppen oder der häufige Wechsel von Mobiltelefon und Adresse sein. Nachdem diese Muster gefunden waren, konnten dem System neue Daten über die allgemeine Bevölkerung zugeführt und es aufgefordert werden, jene gemeinsamen Merkmale aufzuspüren, die vermutlich auf eine Zugehörigkeit zur Hamas hindeuten. Im Wesentlichen unterscheidet sich dieser Ansatz nicht wesentlich von dem Verfahren, das zuvor von menschlichen Geheimdienstmitarbeitern durchgeführt wurde, doch die Automatisierung macht ihn wesentlich schneller.

Laut den Aussagen von sechs anonymen israelischen Geheimdienstmitarbeitern, die alle während des aktuellen Krieges in Gaza im Einsatz waren und Erfahrungen aus erster Hand mit diesem System hatten, stützte sich die Armee in den ersten Wochen des Krieges fast ausschließlich auf Lavender zur Zielidentifizierung. Während dieser Zeit stuft das System bis zu 37.000 Palästinenser als mutmaßliche Militante ein und markierte sie und ihre Häuser für mögliche Luftangriffe. Ein zweites KI-System namens „Where’s Daddy“ wurde speziell dafür entwickelt, sie in ihren Familienhäusern aufzuspüren und nicht während militärischer Aktivitäten, da es einfacher war, die Ziele zu lokalisieren, wenn sie sich in ihren Privathäusern befanden. Dem Bericht zufolge akzeptierte das System Kollateralschäden von 15 bis 20 Zivilisten für einen einzelnen rangniedrigen Kämpfer der Hamas oder des Islamischen Dschihad (PIJ) und über 100 zivile Opfer für einen hochrangigen Kommandeur.

Eine Quelle berichtet, dass die Armee den Offizieren pauschal die Genehmigung erteilte, die von „Lavender“ generierte Zielliste ohne weitere Prüfung zu übernehmen, obwohl sie wusste, dass das System eine Fehlerquote von etwa zehn Prozent aufweist und gelegentlich Personen markierte, die nur lose oder gar keine Verbindungen zu einer militanten Gruppe hatten. Mitarbeiter berichteten, dass sie oft nur als „Abnicker“ für die Entscheidungen der Maschine dienten, und fügten hinzu, dass sie persönlich etwa „20 Sekunden“ für jedes Ziel aufwendeten, bevor sie einen Bombenangriff genehmigten, wobei sie oft nur bestätigten, dass es sich bei dem Ziel um einen Mann handelte. Darüber hinaus erklärten die Quellen, dass manchmal eine erhebliche Zeitspanne zwischen dem Moment, in dem „Where’s Daddy“ einen Offizier darauf hinwies, dass ein Ziel sein Haus betreten hatte, und dem Bombenangriff selbst lag, was dazu führte, dass ganze Familien getötet wurden, ohne dass das beabsichtigte Ziel überhaupt getroffen wurde

## **Einschränkungen durch das humanitäre Völkerrecht**

Die beschriebenen Praktiken werfen viele Fragen hinsichtlich potenzieller Verstöße gegen das humanitäre Völkerrecht (IHL) auf, z. B. gegen den Grundsatz der Unterscheidung (Art. 48 AP I; siehe auch Art. 51(1) und (2) AP I; Art. 13(1) und (2) AP II; ICRC-Gewohnheitsregeln 1, 7) – die Verpflichtung, streng zwischen zivilen und militärischen Zielen zu unterscheiden. Angesichts der im System programmierten zulässigen Kollateralschäden erscheinen Verstöße gegen den Grundsatz der Verhältnismäßigkeit (Art. 51 (5b) AP I; ICRC-Gewohnheitsregel 14) oder das Versäumnis, Vorsichtsmaßnahmen zu treffen (Art. 57, 58 AP I; ICRC-Gewohnheitsregel 15), wahrscheinlich. Die Tatsache, dass Militante der Hamas und der PIJ (beides nichtstaatliche organisierte bewaffnete Gruppen) in ihren Häusern ins Visier genommen wurden, ist besonders problematisch, wenn sie gemäß Art. 43(I) AP I nicht als Kombattanten gelten (zur Erörterung der anwendbaren Normen des

humanitären Völkerrechts und der Konfliktklassifizierung siehe hier und hier). Der Kombattantenstatus für Kämpfer der Hamas und der PIJ ist umstritten, da der Konflikt von verschiedenen Parteien entweder als internationaler oder als nicht-internationaler bewaffneter Konflikt eingestuft wird. Der Kombattantenstatus gilt nicht in nicht-internationalen bewaffneten Konflikten. Israel (und viele andere Beobachter) betrachten den Konflikt zwischen Israel und der Hamas nicht als internationalen Konflikt, da die Hamas keinen Staat repräsentiert.

Zudem würde die Gewährung des Kombattantenstatus für Hamas-Kämpfer es der Hamas aus Sicht des Kriegsrechts ermöglichen, Soldaten der israelischen Streitkräfte (IDF) anzugreifen. Andererseits können Personen ohne Kombattantenstatus rechtlich nur angegriffen werden, wenn sie zu diesem Zeitpunkt aktiv an Feindseligkeiten beteiligt sind (Art. 51 Abs. 3 API), was nicht angenommen werden kann, wenn sie sich in ihren Häusern aufhalten, um zu schlafen.

Gemäß Art. 91 API (Art. 3 der Haager Konvention über die Gesetze und Gebräuche des Landkriegs von 1907 und den Gewohnheitsrechtsregeln 149, 150) ist eine Konfliktpartei, die gegen das humanitäre Völkerrecht verstößt, schadensersatzpflichtig. In diesem Fall kann ein Staat für alle Handlungen verantwortlich gemacht werden, die von Personen begangen werden, die Teil seiner Streitkräfte sind,

Personen oder Einrichtungen, die er zur Ausübung staatlicher Hoheitsrechte ermächtigt hat, Personen oder Gruppen, die auf Anweisung, unter Anleitung oder Kontrolle des Staates handeln, sowie Personen oder Gruppen, deren Verhalten der Staat anerkennt und als sein eigenes übernimmt.

Dementsprechend sind die Handlungen aller staatlichen Organe, die in ihrer offiziellen Funktion ausgeführt werden, seien sie militärischer oder ziviler Natur, dem Staat zuzurechnen.

Die staatliche Verantwortung besteht zusätzlich zu der Verpflichtung, Einzelpersonen wegen schwerer Verstöße gegen das humanitäre Völkerrecht strafrechtlich zu verfolgen (Gewohnheitsrechtliche Regel 151, Art. 51 der Ersten Genfer Konvention; Art. 52 der Zweiten Genfer Konvention; Art. 131 der Dritten Genfer Konvention; Art. 148 der Vierten Genfer Konvention). Zahlreiche militärische Handbücher bekräftigen die individuelle strafrechtliche Verantwortung für Kriegsverbrechen, und diese ist in der Gesetzgebung vieler Staaten verankert.

Ein Staat ist auch für das Unterlassen seiner Organe verantwortlich, wenn diese zur Handlung verpflichtet sind, wie beispielsweise im Fall von Kommandanten und anderen Vorgesetzten, die für die Verhinderung und Bestrafung von Kriegsverbrechen verantwortlich sind (siehe Gewohnheitsrechtliche Regel 153 und Art. 2 der ILC-Entwürfe zu Artikeln über die Staatenverantwortlichkeit).

### **Gewährleistung der Einhaltung des humanitären Völkerrechts beim Einsatz von KI in der Kriegsführung**

Diese Regeln des humanitären Völkerrechts müssen in der Kriegsführung eingehalten werden, unabhängig davon, wie Entscheidungen getroffen werden. Dies bedeutet insbesondere, dass Staaten sicherstellen müssen, dass die von ihnen eingesetzten Instrumente – oder sogar fortschrittliche Instrumente, an die sie gesamte Entscheidungen delegieren – ebenfalls diesen Regeln entsprechen.

Werden Entscheidungen an automatisierte Systeme delegiert, ergeben sich an drei entscheidenden Punkten bestimmte Verantwortlichkeiten: erstens in der Programmierphase; zweitens auf der Führungsebene, wo Entscheidungen über den strategischen Gesamteinsatz des fertigen Programms getroffen werden; und drittens in der Phase des täglichen Einsatzes vor Ort.

In der Programmierphase müssen die Grundsätze des humanitären Völkerrechts in den Code des KI-Systems selbst integriert werden. Dies bedeutet, dass Trainingsdaten sorgfältig ausgewählt werden müssen, um spätere Fehlalarme auszuschließen, dass wichtige Einstellungen und Sicherheitsvorkehrungen festgelegt werden, die die Regeln des humanitären Völkerrechts widerspiegeln, und dass das erforderliche Maß an menschlicher Aufsicht gewährleistet ist, um Fehler oder Fehlfunktionen zu erkennen. Die Trainingsphase ist der Zeitpunkt, an dem das System mit gekennzeichneten Daten gefüttert und aufgefordert wird, Muster zu finden, die eine Gruppe von einer anderen unterscheiden, wie beispielsweise Hamas-Aktivisten von nicht angehörigen Personen. Im Fall von Lavender wurde laut dem Bericht von +972 „der Begriff ‚Hamas-Kämpfer‘ sehr weit gefasst verwendet, und es wurden Personen, die als Mitarbeiter des Zivilschutzes tätig waren, in den Trainingsdatensatz aufgenommen.“ Dies könnte sich als entscheidender Schritt im Hinblick auf das humanitäre Völkerrecht erweisen, da Softwareentwickler dem Programm damit „beigebracht“ haben, nach gemeinsamen Merkmalen nicht nur militanter Hamas-Kämpfer, sondern auch von Zivilisten zu suchen.

Zweitens tragen die befehlshabenden Offiziere die Verantwortung für die angemessene Nutzung. Dies beinhaltet die Aufsicht über den gesamten Prozess und die Sicherstellung, dass die menschlichen Bediener des betreffenden KI-Systems das humanitäre Völkerrecht einhalten (IKRK, Gewohnheitsregeln 15–24). Im hier untersuchten Fall lässt sich ein potenzieller Verstoß gegen diesen Grundsatz darin erkennen, dass die befehlshabenden Offiziere die vom KI-System generierten Tötungslisten ohne weitere Überprüfung pauschal genehmigten und damit das Verfahren der menschlichen Aufsicht auf einen „Stempelabdruck“ reduzierten.

Drittens müssen menschliche Bediener in der Ausführungsphase ihrer Verpflichtung gemäß Art. 57 ZPI (Gebräuchliche Regel 16 des IKRK) nachkommen, alles Mögliche zu tun, um zu überprüfen, ob es sich bei den Zielen um militärische Objekte handelt (zur Zielauswahl der IDF vor dem 7. Oktober siehe hier) und ob die Entscheidung ein Gleichgewicht zwischen militärischer Notwendigkeit und humanitären Erwägungen widerspiegelt (Grundsatz der Verhältnismäßigkeit). Dies ist im vorliegenden Fall fragwürdig, da die Überprüfung jedes einzelnen Falls angeblich nur 20 Sekunden dauerte, in denen der menschliche Bediener oft lediglich bestätigte, dass es sich bei dem Ziel um einen Mann handelte.

Selbst wenn „Lavender“ und „Where’s Daddy“ also unter dem (schlecht definierten) Oberbegriff „künstliche Intelligenz“ zusammengefasst werden und von manchen als „autonom“ wahrgenommen werden könnten, werden ihre Entwicklung und ihr Betrieb dennoch von menschlichen Entscheidungen und menschlichem Verhalten bestimmt, was diese Personen für ihre Entscheidungen verantwortlich macht. Menschliche Beamte können sich ihrer rechtlichen Verantwortung nicht entziehen, indem sie sich hinter einem KI-System verstecken und mangelnde Kontrolle geltend machen, wenn sie die ihnen zur Verfügung stehende Kontrolle schlichtweg nicht ausüben.

Das Hauptproblem, das sich aus fortschrittlichen militärischen KI-Werkzeugen ergibt, ist somit eines des Ausmaßes: Da KI es ermöglicht, Tausende potenzieller Ziele fast gleichzeitig zu markieren, stellt sie die menschlichen Kapazitäten zur Überprüfung und Verifizierung auf die Probe. Es könnte daher für menschliche Entscheidungsträger verlockend sein, sich ohne ordnungsgemäße Überprüfung auf die Ergebnisse der KI zu verlassen und so ihre Entscheidungsbefugnis und Verantwortung an die Maschine zu delegieren. Es liegt in der Verantwortung von Staaten, befehlshabenden Offizieren und Einsatzkräften vor Ort, dieser Versuchung zu widerstehen und den verantwortungsvollen und rechtmäßigen Einsatz dieser neuen Technologien sicherzustellen. Sollten sie dies jedoch versäumen, bleiben die bestehenden Vorschriften des humanitären Völkerrechts ein wirksames Instrument, um die staatliche Rechenschaftspflicht für Verstöße gegen die Regeln bewaffneter Konflikte sicherzustellen.

## Fazit

Abschließend muss betont werden, dass dies keine endgültige Beurteilung der Rechtmäßigkeit eines bestimmten israelischen Angriffs darstellt. Solche Beurteilungen sind im Nebel des Krieges oft unzuverlässig, da es an offiziell bestätigten Insiderinformationen darüber mangelt, wie die IDF ihre Angriffe durchführt. Es lässt sich jedoch feststellen, dass der Einsatz von Lavender und anderen KI-basierten Zielauswahl-Tools Verstöße gegen das humanitäre Völkerrecht wahrscheinlicher machen könnte.

Entscheidend ist, dass der Einsatz von KI-Tools wie „Lavender“ keine Verantwortungslücken bei Verstößen gegen das humanitäre Völkerrecht schafft, da menschliche Entscheidungen weiterhin bestimmen, was diese Anwendungen tun können und was nicht und wie sie eingesetzt werden. Für diese Entscheidungen und den Entscheidungsprozess (z. B. das vollständige Verlassen auf die von der KI generierte Zielliste) können menschliche Verantwortliche nach dem humanitären Völkerrecht zur Rechenschaft gezogen werden. Gaza und die besetzten Gebiete sind seit langem ein Testfeld für neue Überwachungstechnologien und KI-Kriegsführung. Es ist wichtig, diese Systeme zu bewerten und sicherzustellen, dass Technologien, die von Natur aus (aufgrund menschlicher Programmierung) gegen das humanitäre Völkerrecht verstoßen, nicht weiterbestehen. Für eine solche Bewertung ist Transparenz hinsichtlich der Trainingsdaten und der nachrichtendienstlichen Prozesse unerlässlich.

Dennoch sind die bestehenden Regeln des humanitären Völkerrechts gut gerüstet, um die neuen Herausforderungen zu bewältigen, die sich aus dem Einsatz von KI in bewaffneten Konflikten ergeben. Solange Menschen weiterhin die Aufsicht und Kontrolle über KI-Systeme ausüben, können sie für ihre Handlungen zur Rechenschaft gezogen werden, wenn diese gegen das humanitäre Völkerrecht verstoßen; geben sie diese Kontrolle ab, kann diese Entscheidung an sich bereits einen Verstoß gegen das humanitäre Völkerrecht darstellen.

ZITATVORSCHLAG Wiese, Lisa; Langer, Charlotte: Gaza, Künstliche Intelligenz und Tötungslisten, VerfBlog, 16.05.2024, <https://verfassungsblog.de/gaza-artificial-intelligence-and-killlists/>, DOI: 10.59704/07a0756a3c08e64a.