Institut für Technikfolgenabschätzung und Systemanalyse (ITAS)

Sind Maschinen Teil der politisch-extremistischen Kommunikation?

Christian Büscher, Tim Röller, Octavia Madeira, Georg Plattner

Phänomenmonitoring

Zusammenfassung

Generative Künstliche Intelligenz (KI) verbreitet sich zunehmend, bewusst oder unbewusst, in unserem Alltag und wird somit allmählich zu einem Teil der gesellschaftlichen Wirklichkeit. Ein Charakteristikum dieser häufig als disruptiv beschriebenen neuen Technologie besteht wohl darin, auf der Grundlage technischer Operationen kommunikative Inhalte erzeugen zu können. Wenn Gesellschaft als Verkettung kommunikativer Prozesse aufgefasst wird, stellt sich die Frage, inwiefern generative KI dazu beitragen kann, die gesellschaftliche Wirklichkeit hervorzubringen, sie zu reproduzieren oder zu verändern. Diese Frage stellt sich auch und gerade im Zusammenhang mit kommunikativen Prozessen, die mit Phänomenen extremistischer Radikalisierung verbunden sind. Im vorliegenden Beitrag wird deshalb untersucht, inwieweit extremistische Akteur:innen als typische early adopters *neuer Technologien in kommunikativer Absicht bereits auf generative KI zurückgreifen, welche zukünftigen Entwicklungen sich dabei abzeichnen und welche Folgen für das Radikalisierungsgeschehen damit verbunden sein könnten. Zur Beantwortung dieser Fragestellungen wurden Expert:inneninterviews geführt, deren Auswertung mittels einer analytischen Heuristik erfolgt, um Erkenntnisse über die Entwicklung technischer Innovationen sowie technologiebedingter Bedrohungen durch böswillige Akteur:innen zu gewinnen.*

Stichworte

Künstliche Intelligenz (KI) | generative KI | KI-Modelle | Large Language Models (LLMs) | Sprachmodelle | KI-generierte Inhalte | Textgenerierung | Bildgenerierung | Deepfakes | Radikalisierung | authentische Kommunikation | Innovation | Technologie | malevolente Kreativität

Einleitung

Es besteht ein großes Interesse am Innovationspotenzial extremistischer oder terroristischer Akteur:innen (Personen, Gruppen, Organisationen) im Hinblick auf Technik – insbesondere im Zuge der Verbreitung von universell anwendbaren Technologien wie (generativer) Künstlicher Intelligenz (KI), die sich im Allgemeinen anschickt, Teil vielfältiger gesellschaftlicher Prozesse zu werden (Dickel, 2025; Kusche, 2023). Aktuell erscheinen immer mehr Studien, die der Frage nachgehen, inwieweit KI-Technologien extremistische oder terroristische Aktivitäten unterstützen oder ermöglichen können (Baele & Brace, 2024; Brundage et al., 2018; Klinkhammer, 2024; Nelu, 2024; Siegel & Doty, 2023).

Der Fokus auf neue Technologien spielt seit einiger Zeit eine große Rolle in der zivilen Sicherheitsforschung, wie in Fallstudien über die Nutzung von Technologien durch bekannte böswillige Akteur:innen (Argentino et al., 2021) oder quantitativen Studien über verurteilte Terrorist:innen (Gill et al., 2017), die sowohl altbekannte als auch neuartige Technologien für gewalttätige Angriffe eingesetzt haben. Gleichzeitig haben auch die für die zivile Sicherheit zuständigen Akteur:innen Zugang zu neuen Technologien, die sie für die Überwachung und Durchsetzung der Einhaltung von Rechtsnormen einsetzen, was die „Technologisierung der Sicherheit“ (Kaufmann, 2016) vorantreibt, zum Beispiel durch die Idee des „predictive policing“ (Büchner & Dosdall, 2021; Godé & Brion, 2024; Pelzer, 2018).

Wenn es um die begriffliche Einordnung von Techniknutzung und Extremismus/Terrorismus geht, dann wird vorzugsweise der Begriff des *dual use* verwendet (siehe kürzlich Grinbaum & Adomaitis, 2024), der ursprünglich das Potenzial einer Technologie sowohl für zivile als auch für militärische Zwecke bezeichnete (Forge, 2010; Mahfoud et al., 2018), heute aber oft für die Unterscheidung zwischen „benevolenten“ und „malevolenten“ Verwendungszwecken von Technologie steht (Oltmann, 2015). Der Begriff suggeriert, dass verantwortungsbewusste Entwickler und Vermarkter das schädliche Potenzial einer Technologie leicht erkennen können. Diese Annahme erscheint jedoch weniger plausibel, wenn man den vielseitigen, allgegenwärtigen und formbaren Charakter von Informations- und Kommunikationstechnologien bedenkt, zum Beispiel Social-Media-Plattformen, kryptografische Techniken, Künstliche Intelligenz, eine Fülle

anderer Softwareanwendungen oder das Internet im Allgemeinen (Montasari, 2024). Überdies fördert die breite Verfügbarkeit von Erfindungen (wie zum Beispiel Drohnen) Innovationen, die sowohl das Leben der Menschen verbessern als auch die innovative Nutzung durch extremistische und terroristische Akteur:innen ermöglichen können (Cronin, 2020).

Mit der zunehmenden Kommerzialisierung und gesellschaftlichen Verbreitung von KI-Anwendungen wird auch diese Technologie für die böswillige Nutzung durch extremistische Gruppen attraktiver. Im bisher durchgeführten Monitoring innerhalb des MOTRA-Projekts wurde deren Nutzung durch Extremist:innen als sehr relevante Möglichkeit identifiziert (Büscher et al., 2022; Madeira et al., 2023), die nun vertiefend analysiert werden soll. Im Sinne eines Monitorings ist dabei bedeutsam, welche gegenwärtigen Entwicklungen bereits sichtbar und welche zukünftigen Entwicklungen denkbar sind. Eine kontinuierliche Beobachtung kann politische Entscheidungen an der Schnittstelle zwischen Sicherheit und Technologie beraten. Um den potenziellen Missbrauch neuer Technologien nicht zu übersehen, müssen politische Entscheidungsträger:innen verstehen, wie böswillige Kreativität mit technischen Möglichkeiten interagiert.

Wir beginnen diesen Beitrag mit der theoretischen Prämisse, dass der Begriff des *dual use* nicht ausreicht, um das Potenzial des Technologiemissbrauchs und die potenziellen Schwierigkeiten bei der Unterscheidung zwischen wohlwollenden und böswilligen Anwendungsfällen zu erfassen (Büscher & Kusche, 2025). Wir wollen versuchen, die Suche nach malevolenten Potenzialen weiter zu differenzieren. Zur Annäherung werden wir Expert:innengespräche als empirische Grundlage mithilfe eines theoretischen Modells der Interaktion verschiedener Dimensionen der Technikentwicklung und -verbreitung auswerten (Rassler & Veilleux-Lepage, 2024). Rassler und Veilleux-Lepage bieten ein Konzept, das als Heuristik dienen kann, um Verfügbarkeit und Nützlichkeit als Suchkriterien weiter auszudifferenzieren in Richtung Weiterverbreitung und Weiterentwicklung. Damit gehen wir der Frage nach, ob und inwieweit KI immer weiter in soziale Kommunikationsketten eingeflochten wird und damit schlussendlich als Technologie „invisibilisiert“ wird.

Im zweiten Abschnitt dieses Beitrags werden wir zunächst grundlegende Konzepte von KI und malevolenter Kreativität einführen. Daran

anschließend erfolgt die Vorstellung der 4D-Heuristik als analytischer Rahmen, um das terroristische Innovationspotenzial im Bereich der KI systematisch zu erfassen und zu bewerten. Der dritte Abschnitt widmet sich den Ergebnissen der Analysen von Expert:inneninterviews mithilfe der 4D-Heuristik, die Einblicke in aktuelle Einschätzungen von Expert:innen und Erfahrungen zum Thema liefern. Im vierten Abschnitt werden diese Ergebnisse eingeordnet und kritisch diskutiert, wobei ein Abgleich mit der aktuellen wissenschaftlichen Literatur erfolgt, um bestehende Forschungsstände zu reflektieren und offene Fragen zu identifizieren. Abschließend bietet der fünfte Abschnitt einen Ausblick auf die gesamtgesellschaftlichen Herausforderungen, die sich aus der Integration von KI in Kommunikationsprozesse ergeben, und benennt weiterführende Problemstellungen sowie Handlungsfelder für Wissenschaft und Praxis.

Konzeptionelle Grundlagen

Künstliche Intelligenz (KI)

In dem Strategiepapier „Artificial Intelligence for Europe“ definiert die Europäische Union KI (tautologisch) als Systeme mit intelligentem Verhalten, die ihre Umgebung analysieren und mit einem gewissen Grad an Autonomie handeln, um bestimmte Ziele zu erreichen (Europäische Kommission, 2018). Andere Definitionen beziehen sich auf menschenähnliche Intelligenzfähigkeiten einer Maschine, wie Lernen, Urteilsvermögen und Problemlösung (Brundage et al., 2018; UNICRI & UNCCT, 2021). Ein Ziel der Entwicklung von Künstlicher Intelligenz ist vorwiegend die Automatisierung kognitiver menschlicher Tätigkeiten (Humm et al., 2021; Jäkel, 2025).

Die Technologie des maschinellen Lernens (ML) – eine Untergruppe oder Grundlagentechnologie (Humm et al., 2021) der Künstlichen Intelligenz – lehrt Maschinen, aus Daten und Erfahrungen zu lernen und Aufgaben zunehmend autonom zu erledigen. Im Gegensatz zu anderer Computersoftware benötigen solche Algorithmen nach dem anfänglichen Training keine ausdrücklichen Anweisungen von Menschen mehr. Stattdessen erkennen sie Muster und lernen implizite Regeln auf der Grundlage einer bestimmten Anzahl von Beispielen aus einer Datenbank (Brundage et al., 2018;

Schick, 2018). Die KI-Forschung unterscheidet auch zwischen „enger“ KI, die eine einzige Aufgabe erfüllt (diese Form der KI ist heute bereits weitverbreitet und basiert oft auf maschinellem Lernen), und „allgemeiner“ KI, die jede Aufgabe analog zur menschlichen Intelligenz erfüllen könnte (diese Form der KI ist derzeit im Bereich der Science-Fiction angesiedelt) (Jäkel, 2025; Schroeter, 2020, S. 8).

In den vergangenen Jahren haben kommerzielle KI-basierte Anwendungen enorme Fortschritte gemacht: Zunächst wurde Ende 2022 der Chatbot ChatGPT von OpenAI der Öffentlichkeit vorgestellt, der einen regelrechten Boom auslöste und binnen kurzer Zeit sehr hohe Nutzerzahlen erreichte (Hu, 2023). ChatGPT ist dazu in der Lage, auf der Basis von Inputs Fragen zu allen möglichen Themen eloquent zu beantworten, aber auch längere Texte nach bestimmten Vorgaben zu verfassen oder Aufgaben wie Programmieren zu übernehmen, wenngleich die Outputs aufgrund technischer Grenzen nicht immer zuverlässig sind. Die zugrunde liegende Technologie war das Sprachmodell GPT 3.5. Dabei handelt es sich um ein sogenanntes Transformermodell, eine Innovation im Bereich künstlicher neuronaler Netze, welche die enorme Steigerung der Leistungsfähigkeit auf dem Gebiet der maschinellen Sprachverarbeitung möglich macht (Albrecht, 2023).

Ein weiterer großer Schritt erfolgte im Bereich der bildgenerierenden KI. Seit Jahren gibt es Gimmicks für Nutzer:innen wie FaceApp, mit denen man sich selbst oder andere seit 2017 mithilfe von KI-basierter Bildtransformation hochgeladener Fotos älter oder jünger erscheinen lassen oder das Geschlecht manipulieren kann. Im Sommer 2022 wurde die generative KI-Anwendung *Midjourney* auf den Markt gebracht. Sie ermöglicht die Erstellung von Bildern in verschiedenen Stilen (von abstrakt bis fotorealistisch) auf Basis einer schriftlichen Vorgabe und revolutioniert die KI-Bildgenerierung (Bucher & Roy, 2023). Im Zuge des Höhenfluges um *Midjourney* wurden zahlreiche weitere Text-zu-Bild-Modelle veröffentlicht, von denen einige komplett kostenlos nutzbar sind.

Mit der zunehmenden Kommerzialisierung und Verbreitung von KI-Anwendungen wird die Technologie auch für extremistische Gruppen attraktiver, die sie für böswillige (malevolente) Zwecke nutzen wollen. Dies ist auch den Entwickler:innen nicht entgangen. Bei der Entwicklung seiner SORA-Anwendung setzt OpenAI auf den Einsatz von „Red Teams“, um das

Modell zu testen und so dessen Missbrauch beispielsweise für Deepfakes oder Desinformation zu verhindern (Lacy, 2024). Die Bemühungen von OpenAI, „adversarial testing“ durchzuführen, zeigen, dass sich die Entwickler:innen mittlerweile des böswilligen Potenzials ihrer Erfindungen sehr bewusst sind.

Malevolente Kreativität

Um zu verstehen, wie sich extremistische oder terroristische Akteur:innen neue Technologien zunutze machen, orientieren wir uns an der Vorstellung, dass es hierfür malevolenter Kreativität bedarf. Die nachfolgenden Ausführungen in diesem Abschnitt beschäftigen sich deshalb mit dem Zusammenhang von malevolenter Kreativität und dem Innovationshandeln von Extremist:innen oder Terrorist:innen in einer gesellschaftlichen Umgebung, die durch das Vorhandensein offener technologischer Innovationen geprägt wird. Diese Überlegungen gehen auf Argumente zurück, die wir bereits an anderer Stelle entwickelt und dort ausführlicher dargestellt haben (vgl. Gazos et al., 2025).

Gewöhnlich verwenden Extremist:innen Technologien, die kostengünstig, weitverbreitet, benutzerfreundlich und – insbesondere im Bereich von Propaganda, Kommunikation und Rekrutierung – allgemein zugänglich sind. Vor diesem Hintergrund hat Audrey Kurth Cronin eine Theorie entwickelt, die beschreibt, wie das gegenwärtige Paradigma offener technologischer Innovation die extremistische und terroristische Nutzung neuer Technologien begünstigt. Infolge der Digitalisierung sind günstige, leicht verfügbare und benutzerfreundliche Open-Source- und Standardtechnologien entstanden. Dies hat dazu geführt, dass sich ein leistungsfähiges technologisches Ökosystem für Extremist:innen und Terrorist:innen entwickelt hat, das deren Möglichkeiten der gezielten Anwendung von Gewalt quantitativ und qualitativ gesteigert hat (Cronin, 2020).

Die Verwendung von Technologie für einen anderen Zweck als den von ihren Erfinder:innen vorgesehenen oder die Verwendung dieser Technologie zur Schädigung anderer, die ihren Ursprung rein opportunistisch-benevolenten Zwecken verdankt, kann grundsätzlich als eine Form von Kreativität gelten. Cropley, Kaufman und Cropley (2008) nennen diese Art von Nutzung „malevolent creativity“. Sie beschreiben diese Art

von Kreativität als erforderlich für Gruppen oder Personen, um Ziele zu erreichen, die sie als erstrebenswert betrachten, aber die schwerwiegende negative Auswirkungen auf eine andere Gruppe haben. Diese werden von der ersten Gruppe bewusst akzeptiert, genauer gesagt sind sie der eigentliche Zweck (Cropley et al., 2008, S. 106). Dies trifft sowohl auf mit der Ausübung von körperlicher Gewalt einhergehende Handlungen zu als auch auf andere Handlungen, die eine schädliche Wirkung auf andere haben können, beispielsweise auf deren interne oder externe Kommunikation oder auch auf die Finanzierung extremistischer Aktivitäten. Innovation und Kreativität entstehen hier meist nicht spontan, sondern sind das Ergebnis einer vorsätzlichen Suche nach neuartigen Verbesserungen (Gill et al., 2013, S. 131).

Extremist:innen und Terrorist:innen wurde durch die digitale Revolution eine unvergleichliche Menge an Mitteln zur Verfügung gestellt, um ihre Ziele zu verfolgen. Diese haben sie innovativ genutzt und sich einen Vorteil gegenüber den Sicherheitsbehörden geschaffen. Es gibt viele verschiedene Arten extremistischer Aktivitäten, die so durch technische Innovationen erheblich erleichtert wurden. Diese reichen von Massenkommunikation für Propaganda und Rekrutierung bis zu Geldwäsche, Finanzierung, Operationen, Logistik und neuen Methoden der Zerstörung und des Terrors (Gazos et al., 2025). Insbesondere die Ausschöpfung der Möglichkeiten des Internets hat böswillige Akteur:innen in die Lage versetzt, Grenzen von Raum und Zeit zu überwinden, um sich mit Gleichgesinnten auszutauschen und zu vernetzen (Zeiger & Gyte, 2020). In all diesen Berührungspunkten hat die Technologie die Dynamik grundlegend verändert. In all diesen Bereichen gehörten böswillige Akteur:innen zu den „early adopters“.

Sowohl Rechtsextreme aus den USA (beispielsweise mit der Einrichtung des „Stormfront“-Forums Anfang der 1990er-Jahre) als auch Islamist:innen haben früh die Bedeutung des Internets für Propaganda und die Organisation von Anschlägen erkannt (Scrivens & Conway, 2019). Ein Beispiel für eine äußerst erfolgreiche und innovative Technologienutzung ist die Kampagne des sogenannten Islamischen Staates (IS) zur weltweiten Rekrutierung neuer Mitglieder (Piazza & Guler, 2021). Doch böswillige Akteur:innen nutzen auch andere neue Technologien in innovativer Weise: So verwendeten der IS und die Revolutionsgarden der Islamischen Republik Iran schon früh Drohnen als Waffe (Archambault & Veilleux-Lepage, 2024;

Franke, 2014). Bei dem Angriff auf Israel am 7. Oktober 2023 nutzten Kämpfer der Hamas Drohnen, um israelische Verteidigungsanlagen zu zerstören, bevor sie die Grenzen überquerten (Harding et al., 2023). Ein weiteres Beispiel ist die weltweit zunehmende Zahl von Fällen rechtsextremer Gewalt, in denen Waffen, die in Teilen mit einem 3D-Drucker hergestellt wurden, zum Einsatz kamen (Veilleux-Lepage, 2024).

Die 4D-Heuristik zu terroristischen Innovationspotenzialen

Es lässt sich immer mehr Literatur darüber auffinden, wie Terrorist:innen neue Technologien innovativ und kreativ nutzen, aber es gibt immer noch wenig darüber, wie Extremist:innen entscheiden, welche Technologie sie nutzen und welche sie ignorieren. Yannick Veilleux-Lepage und Don Rassler haben einen Rahmen vorgeschlagen, der vier Schlüsselelemente neuer Technologien hervorhebt, die deren Übernahme durch terroristische Gruppen erleichtern. Dieser Rahmen wurde entwickelt, um Folgendes zu bieten: „a visual and conceptual perspective to help researchers and practitioners to understand how terror innovation and technology-related terror threats evolve as part of an interactive system“ (Rassler & Veilleux-Lepage, 2024, S. 24). Dieses Konzept dient uns im Weiteren als analytische Heuristik, die erkennen lässt, wie es dazu kommt, dass bestimmte Technologien für malevolente Intentionen genutzt werden, andere hingegen nicht.

Die vier von Veilleux-Lepage und Rassler identifizierten Schlüsselfaktoren sind „dual use“, „democratized“, „diffusion“ und „directionality“. Die ersten beiden Faktoren beziehen sich auf die Zugänglichkeit einer Technologie und gehen auf die bereits zitierten Arbeiten von Cronin zurück. Dual use wird von den Autoren wie folgt beschrieben: „products, software, and technologies designed for civilian purposes but can also be adapted for military use“ (Rassler & Veilleux-Lepage, 2024, S. 24). Der zweite Begriff „demokratisiert“ bezieht sich auf die Verfügbarkeit von Systemen, die den Austausch von Know-how über die Nutzung dieser Güter mit doppeltem Verwendungszweck ermöglichen. Diese beiden ersten Faktoren spielen eine große Rolle beim Abbau von Barrieren, die insbesondere für nichtstaatliche Akteur:innen bestehen, wenn die innovative Nutzung neuartiger Technologien anvisiert wurde. Bei den beiden anderen Schlüsselfaktoren – Verbreitung und Ausrichtung – geht es um die Verstärkung der Auswirkungen dieser Technologien. Die Diffusion bezieht sich auf die

Verbreitung der (terroristischen) Nutzung einer bestimmten Technologie, während die Direktionalität die verschiedenen Wege betrachtet, die die Technologienutzung durch Lernen aus den zugänglichen Informationen nehmen kann (Rassler & Veilleux-Lepage, 2024).

Die Autoren befassen sich zwar in ihrer Arbeit speziell mit der terroristischen Nutzung und Verbreitung einer bestimmten Technologie, doch der vorgeschlagene Analyserahmen kann nach unserer Auffassung auch wertvolle Einblicke in die Aneignung neuer Technologien durch extremistische Akteur:innen im Allgemeinen (einschließlich gewaltfreier Aktivitäten) sowie möglicher Entwicklungspfade der Technologieverwendung im Besonderen bieten.

Zunächst möchten wir die genannte Annahme empirisch überprüfen, indem wir die angenommenen Schlüsselfaktoren, die miteinander interagieren, als Grundlage für die Analyse von Expert:inneninterviews zur aktuellen und zukünftigen Nutzung von generativer KI durch extremistische Gruppen verwenden. Anschließend werden wir die erzielten Ergebnisse validieren, indem wir sie im Kontext wissenschaftlicher Forschung zu ähnlichen Themen diskutieren.

Ergebnisse

Wir stützen unsere Analysen auf unterschiedliche Daten, die wir seit Beginn des Projekts erhoben haben. Schwerpunktmäßig interpretieren wir Aussagen aus leitfadengestützten Interviews mit insgesamt zwölf Expert:innen, die im vierten Quartal des Jahres 2023 via Microsoft Teams geführt wurden. Die Zielsetzung war, Einschätzungen darüber zu erhalten, welche aktuellen Entwicklungen im Hinblick auf Künstliche Intelligenz im Kontext extremistischer Radikalisierung bereits zu beobachten sind und mit welchen weiteren Entwicklungen dahin gehend zu rechnen ist.

In den Aussagen der befragten Expert:innen werden Anwendungen generativer KI für extremistische Zwecke thematisiert, die sich auf Systeme zur Erzeugung von Texten (Large Language Models), Bildern (Bildgeneratoren) und Videos (Deepfakes) konzentrieren. Die Erzeugung von gesprochener

Sprache wird dagegen nur am Rande und im Zusammenhang mit Deepfakes angesprochen, die Möglichkeit der Erzeugung von Musik, Geräuschen oder anderen Tönen wird gar nicht explizit genannt.

Die berichteten Einschätzungen stützen sich (sofern expliziert) auf die Beobachtung rechtsextremer Akteur:innen und werden nachfolgend anhand der von uns für die Beschreibung des Innovationsgeschehens zugrunde gelegten, dem Modell von Rassler und Veilleux-Lepage (2024) entnommenen Schlüsselkategorien „dual use“, „democratized“, „diffusion“ und „directionality“ dargestellt und analysiert.

Dual-use:
Effizienz, Masse, Variation und Glaubwürdigkeit generierter Inhalte

Die Interviewten sprechen eine Vielzahl denkbarer extremistischer Anwendungsfälle für generative KI an, wobei die Herstellung von Inhalten zur Verbreitung extremistischer Ideologien im Vordergrund steht. Textgenerierende Sprachmodelle eignen sich für die Zusammenfassung von Texten, für die Erstellung von Vorschlägen für Kampagnen, für die Entwicklung von Programmcodes (beispielsweise für Malware), für die Erstellung authentisch wirkender Internetseiten zur Verbreitung bestimmter Narrative, für die Anfertigung juristischer Schriftsätze zur Einschüchterung politischer Gegner oder zur Überflutung von Behörden sowie für die Entwicklung von Social Bots, die automatisiert interagieren und extremistisches Gedankengut verbreiten können. Mit Large Language Models können insbesondere zu jedem beliebigen Thema Texte maschinell erzeugt werden, die explizit (in Form von Hassrede oder Volksverhetzung), aber auch implizit (als Desinformation oder Verschwörungserzählung) in verschiedenen Kontexten Radikalisierung und Polarisierung befördern können.

Dadurch ist es Extremist:innen möglich, in sehr kurzer Zeit sehr viel mehr und variierende Inhalte zu produzieren, die niederschwellig in die Radikalisierung führen können. Insbesondere verschwörungsideologische Narrationen lassen sich auf diese Art und Weise leicht variieren und dabei auch komplexe Erzählweisen umsetzen. Diesen Sachverhalt veranschaulicht einer der befragten Experten mit dem Bild eines Füllhorns:

> *„Wenn ich jetzt plötzlich eine KI habe, kann ich natürlich auch sagen, erzähl mir eine Verschwörung über das Rheinland, die mit Ausländern[1] zu tun hat. Die KI schreibt mir dann etwas herunter und dann bin ich plötzlich in der Lage, aus einem neuen Füllhorn von Geschichten zu gewinnen."*

Bildgeneratoren ermöglichen Extremist:innen die Erstellung falscher Profilbilder und damit die anonyme, aber dennoch authentisch erscheinende Bewegung im digitalen Raum, etwa für Sockenpuppen-Accounts, Trolling-Accounts oder andere Accounts, welche die Wahrung der realen Identität nahelegen oder gar erfordern. Dies gilt ebenso für die Bewegung innerhalb der rechten Community, wenn es darum geht, etwa einen Kanal zu erstellen, einer Gruppe beizutreten oder einen Account zu pflegen. Ein weiteres Anwendungsfeld für KI-basierte Bildgenerierung ist die Herstellung propagandistischer Bilder oder Memes mit der Intention, emotionalisierende, hetzerische oder polarisierende Wirkungen zu erzielen und diese in unterschiedlichen Social-Media-Aktivitäten einzusetzen. Nach den Beobachtungen eines Experten verbinden Rechtsextreme damit eine weitere Möglichkeit, um:

> *„... Menschen, die man politisieren könnte über zum Beispiel Migration und Co, um sie dann auf den eigenen Plattformen oder in geschlossenen Räumen stärker zu radikalisieren. Das ist etwas, was auch bei jeder neuen Technologie immer wieder getestet wird. Erreicht man damit größere Massen?"*

Durch den Einsatz von Bildgeneratoren lässt sich mit reduziertem Aufwand und gesteigerter Effektivität extremistische Bildpropaganda produzieren, mit deren Hilfe sich Verschwörungsnarrative und Weltbilder nahezu beliebig mit passenden synthetischen Belegen versehen lassen, die zunehmend schwerer von authentischem Material unterscheidbar sind und dadurch Radikalisierungsprozesse begünstigen können.

---

1 *Die wörtlichen Zitate aus den Interviews werden so wiedergegeben, wie sie von den Sprecher:innen formuliert wurden. Deshalb nehmen wir in diesen Passagen keine Anpassungen vor, wenn es zu Grammatikfehlern oder Satzfehlern kommt oder wenn nicht die gegenderte Form verwendet wurde.*

Dies gilt ebenso für das Missbrauchspotenzial von Deepfakes, das als enorm eingeschätzt wird, weil die Rezipient:innen medialer Inhalte an die Augenscheinvalidität von Bildern gewöhnt sind. Die Technologie macht es möglich, Videos zu generieren, die als Beweise für Tatsachenbehauptungen stehen und dadurch die Gültigkeit nahezu jeder beliebigen (extremistischen) Weltanschauung plausibilisieren können:

> *„Wenn man das dann noch auf die einzelnen Weltbilder projiziert, Ausländer sind die Schlimmen, auf einmal gibts dann gefakte Videos, an Schulen, Rechtsextremisten, man kann ja wirklich alles machen, Islamisten, Frauen sind die Schlimmen, gibt ja auch Misogynie und die ganze Incel-Szene und so. Auch da. Du kannst ja mittlerweile alles so darstellen, ein Weltbild produzieren, was das alles stützt und supportet."*

Durch authentisch wirkende Deepfakes lassen sich Menschen einfacher täuschen und es ist mit einer Potenzierung von Desinformation zu rechnen, weil die Erzeugung täuschender Inhalte stark vereinfacht wird, während die Erkennung solcher Täuschungen technisch nur schlecht möglich ist. Paradoxerweise ist es gleichzeitig technisch auch schwierig nachweisbar, dass etwas echt ist.

Im Zuge gesteigerter audiovisueller Desinformation kann sich der Kreis der für extremistische Radikalisierung zugänglichen Personen erweitern und die Rekrutierung neuer Mitglieder kann erleichtert werden. Deepfakes können aber auch der Verbreitung von Narrativen innerhalb einer Community und damit der Binnenradikalisierung in diesen Gruppen dienen, indem vorgefertigte Meinungen durch generiertes Videomaterial bestätigt werden. In diesem Kontext kann eine weitere Radikalisierung nicht nur durch authentisch wirkende Täuschungen über Sachverhalte, sondern auch durch erkennbar generierte Videos begünstigt werden, indem beispielsweise in satirischer Machart politischen Gegner:innen Worte in den Mund gelegt werden, die eine versteckte Agenda offenbaren sollen.

Democratized:
Die Verfügbarkeit von Angeboten und das Wissen um ihre Beherrschung

Die nahezu schlagartig erfolgte allgemeine Verfügbarkeit von generativer KI durch die Veröffentlichung von ChatGPT hat die rechtsextreme Szene zunächst genauso überrascht wie den Rest der Bevölkerung. Schnell wurde in diesen Kreisen aber dann damit experimentiert. Die Nutzung für eigene Zwecke wurde getestet und im Internet darüber kommuniziert. Zunächst wurden dabei Grenzen deutlich, weil sich die gewünschten Resultate nicht einstellten:

> *„Deswegen war ChatGPT sehr schnell in der Szene verrufen als ein wokes Instrument. Ein Instrument, was geschaffen wurde von der Silicon-Valley-LGBTQI-Blase, was nicht genutzt werden könnte für die Szene. Gleichzeitig war das dann wiederum der Appell, man müsste da was Eigenes konstruieren oder man müsste diese Möglichkeiten umgehen …“*

Infolgedessen wurde versucht, die Beschränkungen der Erzeugung extremistischer Outputs aufgrund der Filter von ChatGPT über kreative Promptteingaben zu umgehen. Es wurde jedoch schnell deutlich, dass immer wieder Anpassungen durch die Entwickler:innen vorgenommen wurden, um diese Umgehung der Filter zu verhindern. Dennoch hat diese Phase des Ausprobierens dazu geführt, dass der Nutzen und die Vorteile von Large Language Models erkannt wurden. Aus den Einschränkungen, die sich gezeigt haben, wurde die Schlussfolgerung gezogen, dass die Szene eigene Sprachmodelle ohne Filter bräuchte.

Frei verfügbare, lokal einsetzbare Open-Source-Sprachmodelle könnten rechtsextremistische Akteur:innen auf der Grundlage rechtsextremer Ideologie – von Hassrede bis rechter Literatur – optimieren, sodass diese über keine Beschränkungen des Outputs verfügen würden. Zugleich würden sich die daraus resultierenden ideologisch angepassten Sprachmodelle leicht über das Internet für andere verfügbar machen lassen, indem sie zum Download bereitgestellt werden. Nach Ansicht eines Experten müssen Akteur:innen, die gewillt sind, Sprachmodelle in dieser Art und Weise für extremistische Zwecke bereitzustellen, nicht über tiefer gehende Informatikkenntnisse verfügen. Die dafür relevanten Informationen lassen sich demnach mit relativ geringem Aufwand ausfindig machen.

Weitere Plausibilität gewinnt ein solches Szenario durch die Annahme, dass für die Zukunft damit zu rechnen ist, dass Menschen, die in rechtsextremen Szenen aktiv sind, vermehrt Kompetenzen im Umgang mit Large Language Models mitbringen werden. Wenngleich sich Individuen hinsichtlich ihrer technischen Affinität, unter anderem auch in Abhängigkeit von Alter und Bildungsgrad, unterscheiden, wachsen die Mitglieder kommender Generationen mit generativer KI wie ChatGPT auf und erwerben diesbezüglich Wissen. Wenn Menschen aus diesen Alterskohorten sich zu einem anderen Zeitpunkt extremistischen Milieus anschließen, werden sie dieses Wissen auch dort einsetzen.

Eine derartige Ausweitung technologischer Kompetenzen wird die Demokratisierung generativer Sprachmodelle und damit auch deren mögliche Nutzung für extremistische Zwecke weiter befördern. Diese wird außerdem durch die freie Verfügbarkeit von Quellcodes und eine zunehmende Vereinfachung technischer und materieller Voraussetzungen sowie die Offenheit des Internets bedingt. Letzteres ermöglicht nicht nur den Austausch und die Verbreitung von Wissen und Programmcodes, sondern stellt auch eine wichtige Kontextbedingung für die Anwendungsmöglichkeiten generativer Sprachmodelle dar. Dies kann am Beispiel der Nutzung von Social Bots verdeutlicht werden. Dafür werden nicht nur ein gewisses Maß an Anwender:innenwissen und das Vorhandensein eines geeigneten Sprachmodells vorausgesetzt, sondern auch Plattformen wie Telegram, die durch eine relativ offene API[2] und eine relativ schwache Moderation gekennzeichnet sind.

Ähnliche Zusammenhänge wie in den Beobachtungen zu Large Language Models lassen sich in Bezug auf die Demokratisierung von Bildgeneratoren und Deepfakes ausmachen. Das zunehmende Vermögen von generativer KI zur Erstellung nahezu fotorealistischer Bilder unterliegt einer dynamischen Entwicklung, die in der rechten Szene zur Kenntnis genommen wurde. Kommerzielle Software wie Midjourney, Adobe Firefly oder DALL-E wird von Rechtsextremen insbesondere unter Effizienzgesichtspunkten erprobt, um falsche Profilbilder oder emotional aufgeladene Bilder für

2 *Das Akronym API steht für den englischen Begriff „application programming interface" (deutsch: Anwendungsprogrammierschnittstelle). Solche Schnittstellen ermöglichen es, die Funktionen einer Plattform wie zum Beispiel das Kommentieren von Inhalten über Programmcode automatisiert ausführen zu lassen. Dies ist eine der Grundvoraussetzungen für den Betrieb von Social Bots auf einer Social-Media-Plattform. Die verschiedenen Social-Media-Plattformen unterscheiden sich darin, wie offen sie ihre APIs gestalten (Clever et al., 2022).*

Propagandazwecke herzustellen. Dadurch ergibt sich eine Verringerung des Aufwands, weil man nicht mehr auf die Produktion von eigenem Material oder die Suche und den Erwerb passender Stock-Fotos angewiesen ist. Auch mit der Verwendung von Bildgeneratoren zur Erzeugung von Memes, die rechtsextreme Ideologie vermitteln sollen, wird bereits experimentiert.

Aufgrund der bisher nicht gegebenen lokalen Verfügbarkeit der leistungsfähigeren Modelle zur Bilderzeugung sowie des Vorhandenseins von Restriktionen durch Filter in den online verfügbaren Modellen konzentrieren sich entsprechende Versuche jedoch bisher auf subtile, nicht explizit extremistische Bildmotive, etwa Darstellungen, die das *Tradwife*-Ideal[3] abbilden. Dabei erweist sich das Finden geeigneter Prompts als entscheidendes Problem bei der Erstellung passender Bilder.

Zur Demokratisierung von Deepfakes lassen sich im Interviewmaterial drei divergierende Positionen ausmachen. Eine Expertin betont mit Blick auf den Stand der technischen Entwicklung die einfache Verfügbarkeit entsprechender Software, deren Open-Source-Charakter und das geringe Maß an erforderlichem Wissen zu deren malevolenter Verwendung. Im Gegensatz dazu verweist ein anderer Experte anhand von in Selbstversuchen gewonnenen Anschauungen auf praktische Herausforderungen bei der Realisierung der für den Technologiegebrauch unterstellten Möglichkeiten: Das Erstellen eines Deepfakes erfordert ein nicht unerhebliches Maß an Einarbeitung und eine Auseinandersetzung mit der Materie. In der dritten Position, basierend auf der Beobachtung des Verhaltens extremistischer Akteur:innen, verknüpfen sich die beiden zuvor genannten Perspektiven. Demnach setzt eine zur Erzeugung von Deepfakes ausreichende Beherrschung der Technik derzeit noch größere Anstrengungen zur Aneignung von erforderlichem Wissen voraus. Zugleich wird für die Zukunft im Zuge einer weiteren Vereinfachung und allgemeinen Verbreitung der Technologie ein Absinken des Umfangs und der Qualität des vorausgesetzten Know-hows angenommen.

---

3 *Der Begriff Tradwife setzt sich zusammen aus „traditional“ und „wife“, bedeutet also übersetzt „traditionelle Ehefrau“. Er bezieht sich auf Online-Communities, in denen Influencerinnen für eine Rückbesinnung auf eine extrem konservative, traditionelle Geschlechterrollenverteilung und den damit verbundenen Lebensstil werben. Die mediale Vermittlung dieser Auffassungen geht mit einer eigenen Ästhetik einher, die häufig an die 1950er-Jahre angelehnt ist. Problematisch an der Tradwife-Subkultur sind die damit verbundenen exkludierenden Tendenzen, wobei das Spektrum der vertretenen Meinungen von rechtskonservativ bis hin zu rechtsextrem reicht (siehe Proctor, 2022; Sykes & Hopner, 2024).*

Diffusion:
Generative KI als Experimentierfeld und Bedrohung im Werden

Zum Zeitpunkt der Befragungen Ende 2023 lässt sich auf Basis der Einschätzungen der Expert:innen der Stand der Nutzung und Verbreitung von generativer KI in der rechtsextremen Szene als Phase des Ausprobierens charakterisieren, die auch dadurch gekennzeichnet ist, dass die neue Technologie in diesem Feld bisher nicht in der Breite angekommen zu sein scheint. Über die Nutzung von Large Language Models wird vor diesem Hintergrund einerseits von Aktivitäten rechtsextremer Akteur:innen berichtet, sich mit ChatGPT die Arbeit zu vereinfachen. Andererseits werden aber auch Versuche beobachtet, die darauf abzielen, explizit extremistische Inhalte zu erzeugen, indem mithilfe von Methoden wie (hidden) Pre-Prompts oder Fine-Tuning Filter in den Modellen umgangen werden:

> *„Es wurde zum Beispiel ChadGPT geschaffen, was der Begriff ist, wie man sich als maskuliner Mann präsentieren sollte, das Gegenstück zur Woke-Kultur und das war ein Large Language Model, was entertainment-mäßig zum Beispiel in bestimmten der identitären Bewegung nahen Gruppen eingesetzt wurde, bei dem man dann fragen konnte, was denn chad und (unv.) ist und was denn woke und damit verrucht und blöd ist. Damit hatte man sich mehr untereinander ausgetauscht und Fragen dazu gestellt.“*

In solchen Fällen beschränkt sich die Nutzung eines Large Language Models noch darauf, sich wissentlich damit zu unterhalten und nicht eine vermeintlich authentische Konversation mit einer anderen Person vorzutäuschen. Im Gegensatz dazu wird aber auch bereits von der Verwendung von Open-Source-Sprachmodellen für Social Bots berichtet. Demnach gibt es in extremistischen Kontexten verifizierte Fälle, in denen Jugendliche auf diesem Weg gezielt auf Missstände und politische Themen angesprochen wurden, um diese zu radikalisieren. Überdies wird im Kontext von Radikalisierung vermutet, dass Large Language Models in Verbindung mit dem Einsatz größerer Ressourcen an Geld und professionellem Personal in Desinformationskampagnen aus Russland genutzt werden, um authentisch wirkende Blogs oder Nachrichtenwebseiten zu erstellen.

Die Erstellung von Profilbildern mit KI wird in der rechten Szene bereits praktiziert, wobei sich dieser Befund in den Interviews auf die Beobachtung einer geringen Zahl von Fällen stützt. Auch zur Herstellung von Propaganda verwenden verschiedene rechte Akteur:innen bereits Bildgeneratoren, genannt werden das Magazin *Compact* oder die AfD. Ein Experte rechnet damit, dass KI-generierte Bilder künftig häufiger auf Wahlplakaten zu sehen sein werden. Des Weiteren wird von Posts vereinzelter Internet-Accounts berichtet, die darauf schließen lassen, dass Bildgeneratoren zur Erstellung von Memes verwendet wurden. Dabei wird insbesondere auf einen, namentlich nicht näher bezeichneten, sehr aktiven rechten Account[4] auf Instagram hingewiesen. Die Ergebnisse dieser Versuche, rechte Memes mithilfe von KI zu erzeugen, scheinen jedoch (noch) nicht sonderlich erfolgreich zu sein und (noch) kein breiteres Publikum anzusprechen. Dies führt ein Experte auch darauf zurück, dass die dafür geeigneten Tools zum Befragungszeitpunkt bisher nicht verfügbar waren. Für die Zukunft (im Verlauf des nächsten Jahres) wird diesbezüglich mit weiteren Innovationen gerechnet.

Die Verbreitung von Deepfakes unter rechtsextremen Akteur:innen wurde ebenfalls schon wahrgenommen (explizit genannt werden die Identitäre Bewegung und die AfD). Das Ausmaß erweist sich bislang jedoch als geringer, als es einer der Experten zuvor erwartet hatte. Dennoch zeigt sich bei der Betrachtung des Zeitverlaufs von November 2022 bis Oktober 2023 eine deutliche Zunahme der Häufigkeit von Deepfakes, wobei sich nach Ansicht eines Experten deren Urheberschaft auf eher wenige Akteur:innen beschränkt, die sich in den Umgang mit der technischen Materie eingearbeitet haben und dazu in der Lage sind, damit Videos zu erstellen. Auffällig ist, dass die Qualität dieser Deepfakes unprofessionell wirkt und insgesamt eher schlecht ist, wenn etwa visuelle Ungereimtheiten klar erkennbar sind. Ein Experte merkt jedoch an, dass die Synthese von gesprochener Sprache bereits zuverlässig funktioniert und die Stimmen dabei schon authentisch wirken. Die inhaltliche Machart der Deepfakes, über die in den Interviews gesprochen wurde, ist humoristischer Art und dient der Desinformation, indem (politischen) Personen des öffentlichen Lebens Aussagen in den Mund gelegt werden, die bestimmte

4 *Dabei dürfte es sich um den Account „Wilhelm Kachel" handeln, der neben der Präsenz auf Instagram auch auf weiteren großen Social-Media-Plattformen aktiv ist (siehe hierzu Ayyadi, 2024; Deutschlandfunk, 2024).*

Narrative bedienen sollen. Trotz der angesprochenen offensichtlichen Unzulänglichkeiten verbreiten sich diese Deepfakes stellenweise rasant und nach Einschätzung eines Experten scheint sich die Verbreitung solcher Videos und deren Wirkung auf Rezipienten unabhängig von deren zugeschriebener Authentizität zu entfalten. Zum Teil erfolgt die Weiterverbreitung der Videos, weil sie für echt gehalten werden, zum Teil aber auch, obwohl sie als Deepfake erkannt werden, aber die vermittelte Botschaft Zustimmung findet:

> *„Das Spannende ist, wenn man von außen betrachtet nur das Video sähe, würde man denken, dass niemand das glaubt. Aber wenn ein solches Video innerhalb der Community verbreitet wird, trägt das zur Radikalisierung bei, weil immer jemand sagt: ‚Genau das ist es ja.'"*

Deepfakes scheinen somit derzeit in erster Linie ein Mittel zu sein, welches dazu dienen kann, Radikalisierungsprozesse innerhalb der Szene zu begleiten und zu verstärken, während sich Außenstehende auf der Grundlage der hier beobachteten Praktiken rechtsextremistischer Akteur:innen bei der Herstellung und Dissemination generierter Videos darüber eher nicht ohne Weiteres radikalisieren lassen dürften.

Directionality: Zwischen Verstärkung und Transformation bekannter Strategien

In den Beobachtungen der Expert:innen wird bereits ein konkreter Pfad der Veränderung der Nutzung generativer KI durch Rechtsextreme erkennbar. Dieser führt von der normkonformen Verwendung von ChatGPT als Werkzeug zur Bearbeitung und Erzeugung von Texten über das Ausloten von Möglichkeiten zur Erzeugung extremistischer Inhalte durch die Überwindung von Barrieren hin zur Schaffung schrankenloser, anhand rechtsextremer Ideologie optimierter Sprachmodelle. Für die Zukunft wird erwartet, dass es für die rechtsextreme Szene deutlich einfacher werden wird, solche eigenen Modelle anzubieten, die etwa auf einem Finetuning verfügbarer Modelle basieren. Diese Einschätzung beruht mitunter auf Ankündigungen von Akteur:innen wie Elon Musk, Large Language Models

veröffentlichen zu wollen, die eine absolut uneingeschränkte Redefreiheit garantieren sollen.[5]

Mit der Verwirklichung des Ziels, eigene rechtsextremistische Sprachmodelle zu entwickeln und zu verbreiten, geht die Fähigkeit einher, in kurzer Zeit große Mengen variierender, zweckangepasster Inhalte produzieren zu können. Einer der befragten Experten sieht hier eine „rechte Datensphäre“ im Entstehen. Daraus lassen sich verschiedene, an bereits bestehende Strategien rechtsextremer Akteur:innen anschließende, weitere neue Richtungen extremistischer Nutzungsweisen denken.

Ein Experte beschreibt das Bild einer „Radikalisierungspipeline“, indem zunächst mithilfe von Large Language Models große Massen an niederschwelligen, nicht unmittelbar als extremistisch erkennbaren Inhalten erzeugt werden, deren Dissemination stochastische Effekte mit sich bringt, die zu einer Verstärkung des Rabbit-Hole-Phänomens[6] führen könnten. Durch generative Sprachmodelle lässt sich auch die Wirksamkeit orchestrierter Aktionen steigern, indem der Aufwand zur Variation von Textbausteinen verringert wird. Solche Kampagnen waren in der Vergangenheit oftmals leicht durchschaubar, weil die darin involvierten Accounts Textbausteine verwendeten, die keine oder nur eine sehr geringe Varianz aufwiesen.

Frühere Befürchtungen in Bezug auf problematische Entwicklungen durch das Phänomen extremistischer Social Bots, die sich bisher nicht bestätigt hatten, könnten sich nun doch bewahrheiten. In einem extremen Szenario ist es sogar vorstellbar, dass damit gezielt versucht werden könnte, missliebige Teilnehmer:innen an Diskussionen anhand bestimmter Merkmale gezielt zu erkennen, um mit sehr wenig Aufwand mit bösartigen Reaktionen gegen diese vorzugehen, insbesondere deren Social-Media-Accounts mit Inhalten zu überfluten.

---

5 *Die Ankündigung von Elon Musk, eine derartige KI-Anwendung entwickeln zu wollen, wurde mit der Veröffentlichung des Chatbots Grok durch X.AI im November 2023 verwirklicht (Picciotto, 2023). Nachfolgend kam es im weiteren Betrieb immer wieder zu Kontroversen über unangemessenes Verhalten des KI-Systems, wenn dieses beispielsweise antisemitische Äußerungen verbreitete (Lindner, 2025).*

6 *Die Rabbit-Hole-Metapher geht auf den Roman* Alice im Wunderland *von Lewis Carroll zurück und bezeichnet im Kontext von Radikalisierung häufig die Vermutung, dass Nutzer:innen durch die Architektur von Empfehlungsalgorithmen nach und nach in ideologisch zunehmend extremere Informationsumgebungen geführt werden (siehe zum Beispiel Brown et al., 2022; Lewis et al., 2024).*

Eine weitere mögliche neue Richtung der Nutzung generativer Sprachmodelle, die auch mit bereits verfügbaren Angeboten realisierbar sein dürfte, basiert auf deren ausgeprägter Fähigkeit zur Verarbeitung juristischer Texte. Hier wird die Möglichkeit gesehen, dass sich Akteur:innen wie Reichsbürger dies zunutze machen, um die Strategie des sogenannten „Paper Terrorism" auf ein höheres Niveau zu heben, indem sie mit minimalem Aufwand massenhaft authentisch erscheinende pseudojuristische Schriftstücke erzeugen, um damit politische Gegner:innen einzuschüchtern oder Behörden zu lähmen.[7]

In den berichteten Beobachtungen der Nutzung von KI-Tools zur Bildgenerierung lassen sich ebenfalls Entwicklungen ausmachen, welche an bereits existierende Praktiken zur Lösung bestimmter Problemstellungen anknüpfen. Dies zeigt sich etwa im Hinblick auf die Verwendung von Bildmaterial zur Wahrung von Anonymität im digitalen Raum, wobei KI-basierte Bildgeneratoren neue Möglichkeiten eröffnen, die zu einer Veränderung und Optimierung der vorgängigen Praxis führen. Während in der Vergangenheit szenebezogene Dinge, Symbole, Memes oder Ähnliches zur Gestaltung von Accounts verwendet wurden, werden mittlerweile wieder vermehrt Profilbilder genutzt, von denen angenommen wird, dass sie KI-generiert sind. Dies hat den Vorteil, einen Account nahbarer wirken zu lassen. Dabei stellt die Nutzung von Anwendungen Künstlicher Intelligenz keineswegs eine abrupte Entwicklung dar, sondern es besteht eine gewisse Kontinuität. Ein Experte verweist auf die Verwendung der Website „This-person-does-not-exist-dot-com" mit der sich zufällige Porträts von Personen generieren lassen. Die entsprechenden Ergebnisse sind zwar noch relativ gut als generiert zu erkennen, lassen sich jedoch mit wenig Aufwand so überarbeiten, dass das weniger offensichtlich ist.

Auch die KI-basierte Generierung von Bildpropaganda setzt bei einer bestehenden Praxis an, die vereinfacht und optimiert wird. Die Möglichkeit, beliebiges Bildmaterial erzeugen zu können, macht das aufwendige

7 *Mit dem Begriff des Paper Terrorism wird eine Strategie der Sovereign-Citizen-Bewegung in den USA bezeichnet, deren verschwörungstheoretische Ideologie dem Staat jede Legitimität abspricht. Im Kern zielt diese darauf ab, auf Basis fingierter, teils bizarrer Rechtsbegriffe staatliche Behörden und Politiker:innen, aber auch politische Gegner oder sonstige Mitbürger:innen in zermürbende juristische Auseinandersetzungen zu verwickeln, um ihre Weltanschauung durchzusetzen (siehe Loeser, 2015; Southern Poverty Law Center, 2025). Die Anwendung dieser Strategie wird auch im deutschen Reichsbürgermilieu beobachtet (Fuchs et al., 2025).*

Suchen von passenden Bildern in Datenbanken obsolet. Gleichzeitig wird die Qualität der Ausgaben von Bildgeneratoren im Zuge der technischen Entwicklung immer besser. Dadurch wird es schwerer, bei extremistischen Inhalten zwischen Wirklichkeit und Fälschung zu unterscheiden. Infolgedessen lassen sich mit generierten Bildern extremistische Anschauungen zu beliebigen Sachverhalten mit authentisch erscheinenden Tatsachenbehauptungen belegen.

Die Erstellung rechtsextremer Memes könnte gleichermaßen in Zukunft durch die Verwendung von Bildgeneratoren vereinfacht werden, was sich nach Einschätzung der Expert:innen zum Interviewzeitpunkt noch als schwierig erweist. Dennoch könnten die auf diesem Gebiet erfolgenden Versuche zu Fortschritten führen, die mit einem stilistischen Wandel von Memes einhergehen. Eine weitere Richtung für Innovation wird vor diesem Hintergrund darin gesehen, dass die Durchführung koordinierter rechter Aktionsformen im Netz auch in Bezug auf visuelle Mittel erleichtert werden könnte, insbesondere das koordinierte Agieren auf mehreren Plattformen.

Analog zu der von den Expert:innen erwarteten Steigerung der Qualität von Bildgeneratoren, wird eine allgemeine Tendenz in diese Richtung auch für Deepfakes angenommen. Die Weiterentwicklung generativer KI führt dazu, dass sich damit zunehmend authentischer wirkende Videos erstellen lassen, die schließlich nicht mehr von der Wirklichkeit unterscheidbar sein werden. Gleichzeitig gehen die Expert:innen davon aus, dass sich die Zugangshürden zu entsprechenden Tools weiter absenken werden: Software zur Erstellung von Deepfakes wird zunehmend allgemein verfügbar werden und einfacher in der Handhabung sein.

Vor dem Hintergrund der Erwartung einer solchen allgemeinen Entwicklung leitet ein Experte ein Szenario für eine mögliche neue Richtung der Nutzung von Deepfakes durch Rechtsextremist:innen ab. Auf chinesischen Handelsplattformen sind bereits heute KI-generierte Verkaufsvideos zu sehen, in denen die Verkäuferinnen schon zum jetzigen Zeitpunkt nicht mehr als generiert zu erkennen sind. Mit Zugriff auf die Erstellung derart echt wirkender Deepfakes könnten rechtsextremistische Akteur:innen glaubwürdig wirkende und reichweitenstarke rechte Influencer:innen simulieren und zudem den Eindruck einer enormen Zahl an Unterstützer:innen erwecken. Dies würde auf bereits praktizierten Strategien aufbauen,

die auf Textbasis mit Sockenpuppen-Accounts operieren, aber dazu führen, dass rechte Akteur:innen auch auf Video-Plattformen wie TikTok noch stärker werden würden, wenn dort sowohl Inhalte als auch die Reaktionen darauf generiert werden könnten.

Diskussion und Einordnung

Bereits bevor mit der Veröffentlichung von ChatGPT die Möglichkeiten von Sprachmodellen einem breiten Publikum bekannt und zugänglich gemacht wurden, war das Missbrauchspotenzial für extremistische Propaganda Gegenstand von Forschung, die deutlich gemacht hat, wie sich mit generativer KI in kürzester Zeit überzeugende extremistische Outputs verschiedener ideologischer Ausrichtung und Machart erstellen lassen (McGuffie & Newhouse, 2020; Newhouse et al., 2019). Während hinsichtlich der Glaubwürdigkeit und Zuverlässigkeit solcher Outputs noch gewisse Zweifel geltend gemacht wurden (Baele, 2023), zeigen jüngere Forschungsergebnisse, dass sich mittlerweile selbst Expert:innen durch KI-generierte extremistische Texte täuschen lassen (Baele et al., 2024). Diese Erkenntnisse stützen die Einschätzungen aus den Interviews zum Potenzial von generativer KI in Bezug auf Effizienz, Skalierbarkeit, Variation und Glaubwürdigkeit generierter extremistischer Kommunikationsinhalte.

Für den Befund, dass dieses Potenzial auch von den extremistischen Akteur:innen selbst zunehmend erkannt wird und dadurch Aktivitäten des Ausprobierens motiviert werden, lassen sich weitere Belege in rechtsextremistischen Online-Diskursen finden, wenngleich auch ideologische Vorbehalte gegenüber generativer KI auszumachen sind (Allchorn, 2024). Dennoch sind Prozesse der Übersetzung von malevolenter Kreativität in malevolente Innovation bereits im Gange (Hunter et al., 2022), indem Extremist:innen gezielt nach Wegen und Methoden suchen, um generative KI für sich zu nutzen, dabei Einschränkungen kommerziell verfügbarer KI-Modelle umgehen oder auf Open-Source-Basis eigene Modelle entwickeln und das so gewonnene malevolente Wissen im Internet an einen breiteren Personenkreis weitergeben (Dean, 2025; Koblentz-Stenzler & Klempner, 2023; Siegel, 2023). Ein weiterer Grund zur Sorge ist vor diesem Hintergrund, dass die Anfälligkeit verfügbarer KI-Modelle gegenüber missbräuchlicher

Verwendung bereits ohne manipulative Eingriffe hoch zu sein scheint, wobei diese mit dem jeweiligen Zweck variiert und sich verschiedene Modelle dahin gehend deutlich unterscheiden (Weimann et al., 2024).

Die Verbreitung von verfügbarer generativer KI unter Akteur:innen der extremen Rechten und deren Nutzung wird in den Interviews als Situation beschrieben, die sich als „Bedrohung im Werden" charakterisieren lässt. Diese Einschätzung ist anschlussfähig an aktuellere empirische Untersuchungen, die Aufschluss über zeitlich nachgelagerte Entwicklungen geben und zeigen, wie und in welchem Ausmaß unterschiedliche Akteur:innen der extremen Rechten in Deutschland bereits KI-generierte Inhalte wie Bilder, Memes, Musik oder Videos erstellen und auf verschiedenen Social-Media-Plattformen verbreiten (CeMAS, 2025; Donner et al., 2024; Fielitz et al., 2024; Hiller & Maristany de las Casas, 2025). Auch in diesen Arbeiten wird deutlich, dass die Problematik KI-generierter Inhalte nicht allein an deren Glaubwürdigkeit gekoppelt ist, sondern auch und vielleicht vielmehr – gerade auf dem Gebiet extremistischer Radikalisierung – an deren Vermögen, die von Extremist:innen beabsichtigten Botschaften zum Ausdruck bringen zu können und dadurch ein entsprechendes Publikum effizienter ansprechen zu können (Kusche, 2024). In eine ähnliche Richtung weisen auch die Ergebnisse einer experimentellen Studie, wonach die Wirkung KI-generierter Bilder im Kontext politischer Kommunikation nicht so sehr davon abhängt, dass diese als echt wahrgenommen werden, sondern mehr davon, dass es eine Übereinstimmung zwischen den Bildinhalten und den politischen Einstellungen der Rezipient:innen gibt (Kruschinski et al., 2025).

Die in den Interviews angedeuteten Entwicklungsrichtungen extremistischer Nutzung generativer KI fokussieren kommunikative Intentionen, die auf gegenwärtigen rechtsalternativen Online-Strategien wie Mainstreaming, Desinformation, alternativen Medien oder Online-Hatespeech (Heil, 2023) aufbauen und diese in die Zukunft fortschreiben, wobei eine technisch ermöglichte Steigerung von deren Wirksamkeit antizipiert wird. Dass eine solche Zukunft gleichermaßen bereits greifbar, aber dennoch auch ein ganzes Stück weit noch entfernt ist, lässt sich am Beispiel der Vorstellung von KI-generierten Influencer:innen illustrieren. Hiller und Maristany de las Casas (2025) finden in ihrer Studie zu generativer KI in der deutschen extremen Rechten bereits vereinzelte Fälle, in denen virtuelle

rechte Influencerinnen simuliert und online als reale Personen dargestellt werden. Diese können jedoch aufgrund relativ offensichtlicher Merkmale leicht als KI-generiert identifiziert werden und erreichen nur in sehr geringem Ausmaß Interaktionen mit echten Nutzer:innen. Reaktionen auf Beiträge dieser KI-Influencerinnen markieren einerseits deren künstlichen Ursprung und machen sich teilweise darüber lustig. Andererseits sind aber auch Anschlusskommunikationen zu beobachten, welche deren synthetische Natur relativieren und die Bedeutsamkeit der gesendeten Botschaften betonen.

Im Sinne einer Methodenkritik beabsichtigen wir, vor dem Hintergrund unserer Ergebnisse auf zwei zusätzliche Punkte einzugehen. Wenn man den von Rassler und Veilleux-Lepage (2024) vorgeschlagenen theoretischen Rahmen als Prozessmodell versteht, der beschreibt, wie sich neue Technologien in extremistischen Kontexten entfalten, dann fällt eine Leerstelle zu Beginn auf. Diese lässt sich an der Emergenz einer möglicherweise disruptiv wirkenden Technologie (mit Dual-use-Charakter und Demokratisierungspotenzial) und der initialen Nutzung durch extremistische Akteur:innen festmachen. Der Schritt, wie es zu der Einsicht in die Verfügbarkeit und Nützlichkeit (für die eigenen Zwecke) kommt, bleibt in der Konzeption unscharf. Hier gibt es nur vage Verweise auf „kreative Prozesse“, den Zwang zur Innovation aufseiten der Nutzer (benelovent, malevolent) oder auf die Annahme, dass neuen offenen Technologien ein gesteigertes (malevolentes) Innovationspotenzial inhärent sei. Hier könnte das Konzept der malevolenten Kreativität ansetzen (Cropley et al., 2008).

Vor dem Hintergrund des Aspekts der Verfügbarkeit von Wissen über eine Technologie, die bei Rassler und Veilleux-Lepage (2024) in Bezug auf den Faktor Demokratisierung eine zentrale Rolle spielt, kann malevolente Kreativität womöglich als entscheidender Prozess der Wissensgenerierung verstanden werden, um das einer Technologie innewohnende Dual-use-Potenzial erschließen zu können. Technologien sind tendenziell verwendungsoffen und damit zwangsläufig auch offen für malevolente Kreativität (Gazos et al., 2024). Doch das Dual-use-Potenzial einer Technologie ist nicht zwingend offensichtlich, sondern ein Bewusstsein dafür – und im Weiteren ein kommunizierbares Wissen darum – entsteht bisweilen erst in der kreativen Auseinandersetzung mit Möglichkeiten malevolenter Verwendungsweisen. Demnach muss das Wissen um die

Dual-use-Eigenschaften von Technologien erst erzeugt werden. Unseres Erachtens eignet sich für weitere Forschungen dazu anstelle des Dual-use-Begriffs eher die Frage nach dem Maß an noch zu entdeckenden Affordanzen. Damit sind die angenommenen funktionalen Eigenschaften einer Technologie bezeichnet, die den Benutzer:innen mögliche Verwendungszwecke oder Handlungen aufzeigen. Affordanzen sind keine festen Merkmale, sondern entstehen aus der Beziehung zwischen der Technologie und dem Benutzer, abhängig von den Absichten, der Situation und dem kulturellen Wissen des Benutzers. Das bedeutet, dass verschiedene Benutzer:innen dieselbe Technologie auf unterschiedliche Weise wahrnehmen und nutzen können und dass einige Affordanzen je nach den Erwartungen oder dem Suchverhalten des Benutzers unbemerkt bleiben können (Büscher & Kusche, 2025, S. 4).

An diesem Punkt ist noch nichts darüber gesagt, wie die entsprechende Technologie verwendet werden muss, um in malevolenter Absicht wirken zu können, das heißt, wie sich die mithilfe malevolenter Kreativität gewonnenen Ideen in die Tat umsetzen lassen, wie es also zu malevolenter Innovation kommt. Hierfür sind Aktivitäten des Erkundens notwendig, die darauf abzielen, Möglichkeiten der Umsetzung der Ideen zu finden, auszuprobieren und zu verbessern (Hunter et al., 2022). Die Aneignung von Technologien durch extremistische Akteur:innen lässt sich genau in diesem Sinne als iterativer Lernprozess beschreiben, der in Wechselwirkung mit der allgemeinen technologischen Entwicklung steht und in dessen Verlauf die erfolgreiche Verwendung der Technologie in der beabsichtigten Art und Weise entsteht (Gartenstein-Ross et al., 2019). Letztlich offenbart sich auch hier das Problem der Notwendigkeit, Wissen zu erzeugen, das *a priori* nicht vorhanden ist. Erst mit dem Vorhandensein eines solchen Handlungswissens ist die Demokratisierung des malevolenten Gebrauchs der Technologie möglich, und erst dann wird eine Steigerung von deren Wirkung in Form der Weiterverbreitung und Weiterentwicklung wahrscheinlicher.

Ausblick

Im Zentrum der weiteren Entwicklung von generativer KI und der damit verbundenen Folgen steht mit Blick auf unsere Ergebnisse die Frage, ob sich die Unterscheidung zwischen artifizieller und menschlicher Kommunikation auflöst, oder genauer gesagt, ob sie weiterhin eine Rolle spielt (Dickel, 2025). Die dreifache Selektion von Information, Mitteilung und Verstehen (Luhmann, 1984, S. 191ff.) erlaubt die unterschiedliche Zurechnung, wer diese Selektion vornimmt, ein Mensch oder eine Maschine. Soziologische Studien sehen im Allgemeinen die Möglichkeit, dass sich Beiträge von Maschinen in den Kommunikationsfluss einflechten lassen, wenn diese interessant, informativ, zuverlässig und sogar unterhaltsam sein können.[8] Die weitere Überprüfung der Mitteilung auf ihren generativen Ursprung – der Selektion der Informationen – ist in der Dynamik von Kommunikationsflüssen in den meisten Fällen nicht zu leisten: „Normally there is not the time and the motivation to question whether it is a machine or a human being“ (Esposito, 2022, S. 108). Dies ist keine Deformation des Sozialen durch die Einmischung von Maschinen, sondern der soziale Regelfall, wie ihn die Organisationssoziologen March und Simon (1993, S. 186) anhand des Prinzips der Unsicherheitsabsorption beschrieben haben. Die Herleitung von Informationen kann in den allermeisten sozialen Situationen nicht oder nur in einem sehr begrenzten Maße kontrolliert werden. Dickel erörtert die Möglichkeit einer tatsächlich existierenden Künstlichen Intelligenz demzufolge primär unter dem Gesichtspunkt einer sozialen Fragestellung und weniger als ein technisches Problem. Die Anerkennung einer technischen Entität als kommunikatives Gegenüber hängt von der Anschlussfähigkeit ab, die sich in der Unterlassung der Zurechnung auf Maschinen manifestiert. Ob generative KI, zum Beispiel in Gestalt von Bots, Beitragende zur sozialen Interaktion werden kann, hängt davon ab, ob das Verstehen der Mitteilung auf Versuche der Identifikation von Markern der Maschinenhaftigkeit verzichtet (Dickel, 2025, S. 201). Auch unsere Interviewpartner:innen deuten diesen Entwicklungspfad an, wenn beispielsweise Deepfakes trotz erkennbarer Künstlichkeit in spezifischen Communitys einen Informationswert gewinnen und radikalisierende Anschlussfähigkeit präparieren. Oder wenn „Radikalisierungspipelines“ mit nicht auf den ersten

8 *Ganz im Sinne von Gregory Bateson (1995, S. 39 f.): „Jede Informationsaufnahme ist notwendig die Aufnahme einer Nachricht von einem Unterschied.“*

Blick als generiert erkennbaren extremen Inhalten geflutet werden, die mögliche Rezipient:innen nur mit erhöhtem kognitiven Prüfaufwand – und damit Zeitverbrauch – erkennen könnten. Eine spannende Forschungsfrage ergibt sich für uns deshalb im Weiteren: Welche Effekte wird eine „Indifferenz gegenüber der menschlichen oder maschinellen Agency" (Dickel, 2025, S. 203), das heißt der wachsenden Irrelevanz der Klärung des Ursprungs der Information, auf das Radikalisierungsgeschehen und möglicherweise auch auf Präventionsarbeit haben?

Literatur

Albrecht, S. (2023). *ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen.* Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag (TAB).

Allchorn, W. (2024). Global far-right extremist exploitation of artificial intelligence and alt-tech: the cases of the UK, US, Australia and New Zealand. *Counter Terrorist Trends and Analyses, 16*(3), 13–18.

Archambault, E. & Veilleux-Lepage, Y. (2024). The islamic state's drone innovation. In J. P. Rogers (Hrsg.), *De Gruyter Handbook of Drone Warfare* (S. 243–254). De Gruyter.

Argentino, M.-A., Maher, S. & Winter, C. (2021). *Violent extremist innovation: A cross-ideological analysis.* International Centre for the Study of Radicalisation. https://icsr.info/wp-content/uploads/2021/12/ICSR-Report-Violent-Extremist-Innovation-A-Cross%E2%80%91Ideological-Analysis.pdf

Ayyadi, K. (2024). Tannwald Media und Wilhelm Kachel: AfD setzt im Wahlkampf auf KI. *Belltower.News.* https://www.belltower.news/tannwald-media-afd-beauftragt-rechtsextremen-mit-ki-wahl-kampf-157509/

Baele, S. J. (2023). Artificial intelligence and extremism: the threat of language models for propaganda purposes. *Crest Security Review, 16*, 14–17.

Baele, S. J. & Brace, L. (2024). *AI Extremism: Technologies, tactics, actors.* VOX-Pol Network of Excellence.

Baele, S. J., Naserian, E. & Katz, G. (2024). Is AI-generated extremism credible? experimental evidence from an expert survey. *Terrorism and Political Violence, 37*(8), 1060–1076. https://doi.org/10.1080/09546553.2024.2380089

Bateson, G. (1995). *Geist und Natur – Eine notwendige Einheit* (4. Aufl.). Suhrkamp.

Brown, M. A., Bisbee, J., Lai, A., Bonneau, R., Nagler, J. & Tucker, J. A. (2022). *Echo chambers, rabbit holes, and algorithmic bias: how youtube recommends content to real users* (SSRN Scholarly Paper No. 4114905). Social Science Research Network. https://doi.org/10.2139/ssrn.4114905

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G. C., Steinhardt, J., Flynn, C., Ó hÉigeartaigh, S., Beard, S. J., Belfield, H., Farquhar, S., … Amodei, D. (2018). *The malicious use of artificial intelligence: forecasting, prevention, and mitigation.* https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

Bucher, B. & Roy, S. (2023). *Krasse Evolution der Foto-KI: Midjourney früher und heute.* chip.de. https://www.chip.de/news/Krasse-Evolution-der-Foto-KI-Midjourney-frueher-und-heute_184707368.html

Büchner, S. & Dosdall, H. (2021). Organisation und digitale Technologien. Predictive Policing im organisationalen Kontext. *Soziale Systeme, 26*(1–2), 217–239. https://doi.org/10.1515/sosys-2021-0008

Büscher, C. & Kusche, I. (2025). Monitoring new and emerging technologies in order to prevent extremism and terrorist violence. *Technological Forecasting and Social Change, 219,* 124274. https://doi.org/10.1016/j.techfore.2025.124274

Büscher, C., Kusche, I., Röller, T., Andres, F., Gazos, A., Hahn, J., Ladikas, M., Madeira, O., Plattner, G. & Scherz, C. (2022). Trends der zukünftigen Technologienutzung im Kontext von Extremismus und Terrorismus: Erste Erkenntnisse aus dem MOTRA-Technologiemonitoring.

In U. Kemmesies, P. Wetzels, B. Austin, C. Büscher, A. Dessecker, E. Grande & D. Rieger (Hrsg.), *MOTRA-Monitor 2021* (S. 248–281). BKA.

CeMAS (2025). *Autoritäre Strategien im Netz: Analyse und Monitoring digitaler Risiken rund um die Bundestagswahl 2025.* CeMAS – Center für Monitoring, Analyse und Strategie.

Clever, L., Klapproth, J. & Frischlich, L. (2022). Automatisierte (Gegen-)Rede? Social Bots als digitales Sprachrohr ihrer Nutzer*innen. In J. Ernst, M. Trompeta & H.-J. Roth (Hrsg.), *Gegenrede digital: Neue und alte Herausforderungen interkultureller Bildungsarbeit in Zeiten der Digitalisierung* (S. 11–26). Springer Fachmedien. https://doi.org/10.1007/978-3-658-36540-0_2

Cronin, A. K. (2020). *Power to the people: How open technological innovation is arming tomorrow's terrorists.* Oxford University Press.

Cropley, D. H., Kaufman, J. C. & Cropley, A. J. (2008). Malevolent creativity: A functional model of creativity in terrorism and crime. *Creativity Research Journal,* 20(2), 105–115. https://doi.org/10.1080/10400410802059424

Dean, L. (2025). *AI or aryan ideals? A thematic content analysis of white supremacist engagement with generative AI.* GNET. https://gnet-research.org/2025/01/13/ai-or-aryan-ideals-a-thematic-content-analysis-of-white-supremacist-engagement-with-generative-ai/

Deutschlandfunk (2024, 29. August). *Rechte Propaganda – Die völkischen KI-Memes von „Wilhelm Kachel".* Deutschlandfunk.de. Abgerufen am 25. Juni 2025, von https://www.deutschlandfunk.de/rechtsextreme-memes-kuenstliche-intelligenz-wilhelm-kachel-propaganda-100.html

Dickel, S. (2025). Im Imitationsspiel. Über die Kommunikation mit Maschinen und das Streben nach Artificial General Intelligence. *Zeitschrift für Soziologie,* 54(2), 190–206. https://doi.org/10.1515/zfsoz-2025-2011

Donner, C., Böswald, L.-M., Fielitz, M., Sick, H., Marcks, H., Gsenger, R. & Heft, A. (2024). Attention is all they need: Eine Analyse der Nutzung generativer KI in rechtsalternativen Netzwerken. *Machine Against the Rage,* 7(2). https://doi.org/DOI:%252010.58668/matr/07.2

Esposito, E. (2022). *Artificial communication: how algorithms produce social intelligence.* MIT Press. https://doi.org/10.7551/mitpress/14189.001.0001

Europäische Kommission (2018). *Mitteilung der Kommission an das Europäische Parlament, den Europäischen Rat, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen: Künstliche Intelligenz für Europa.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2018%3A237%3AFIN

Fielitz, M., Sick, H., Schmidt, M. & Donner, C. (2024). *Social-Media-Partei AfD? – Digitale Landtagswahlkämpfe im Vergleich* (No. 73; OBS-Arbeitspapier). Otto Brenner Stiftung. https://www.otto-brenner-stiftung.de/fileadmin/user_data/stiftung/02_Wissenschaftsportal/03_Publikationen/AP73_LTW_AfD_WEB.pdf

Forge, J. (2010). A note on the definition of "dual use". *Science and Engineering Ethics,* 16(1), 111–118. https://doi.org/10.1007/s11948-009-9159-9

Franke, U. E. (2014). The global diffusion of unmanned aerial vehicles (UAVs), or 'drones'. In M. Aaronson, R. Rauxloh, W. Aslam, T. Dyson, A. Barrinha, L. Alison & U. E. Franke (Hrsg.), *Precision Strike Warfare and International Intervention: Strategic, Ethico-Legal and Decisional Implications* (S. 52–73). Routledge.

Fuchs, C., Geisler, A., Polke-Majewski, K. & Steinhagen, M. (2025, 31. Juli). Reichsbürger: Reich ohne Steuern. *Die Zeit.* https://www.zeit.de/2025/32/reichsbuerger-steuerhinterziehung-finanzbehoerden-steuerfahndung-polizei

Gartenstein-Ross, D., Jones, D. & Shear, M. (2019). *Virtual plotters. Drones. Weaponized AI?: Violent non-state actors as deadly early adopters.* Valens Global. https://valensglobal.com/virtual-plotters-drones-weaponized-ai-violent-non-state-actors-as-deadly-early-adopters/

Gazos, A., Madeira, O., Plattner, G., Röller, T. & Büscher, C. (2024). New and emerging perspectives for technology assessment: Malevolent creativity and civil security. *TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis, 33*(2), 9–15. https://doi.org/10.14512/tatup.33.2.09

Gazos, A., Madeira, O., Plattner, G., Röller, T., & Büscher, C. (2025). Technologiemonitoring: Soziotechnische Innovationsdynamiken und systemische Resilienz. In U. Kemmesies, P. Wetzels, B. Austin, C. Büscher, A. Dessecker, S. Hutter & D. Rieger (Hrsg.), *MOTRA-Monitor 2023/24* (S. 264–285). MOTRA. https://doi.org/10.57671/ISBN.978-3-911329-01-9_2025_MOTRA

Gill, P., Corner, E., Conway, M., Thornton, A., Bloom, M. & Horgan, J. (2017). Terrorist use of the internet by the numbers. *Criminology & Public Policy, 16*(1), 99–117. https://doi.org/10.1111/1745-9133.12249

Gill, P., Horgan, J., Hunter, S. T. & D. Cushenbery, L. (2013). Malevolent creativity in terrorist organizations. *The Journal of Creative Behavior, 47*(2), 125–151. https://doi.org/10.1002/jocb.28

Godé, C. & Brion, S. (2024). The affordance-actualization process of predictive analytics: Towards a configurational framework of a predictive policing system. *Technological Forecasting and Social Change, 204*, 123452. https://doi.org/10.1016/j.techfore.2024.123452

Grinbaum, A. & Adomaitis, L. (2024). Dual use concerns of generative AI and large language models. *Journal of Responsible Innovation, 11*(1), 2304381. https://doi.org/10.1080/23299460.2024.2304381

Harding, E., Leiter, M. & Byman, D. (2023). *Hamas' october 7 attack: the tactics, targets, and strategy of terrorists.* Center for Strategic and International Studies (CSIS). https://www.csis.org/events/hamas-october-7-attack-tactics-targets-and-strategy-terrorists

Heil, F. (2023). Rechts-alternative Onlinestrategien und ihr Gefahrenpotenzial für demokratische Gesellschaften. In T.-G. Rüdiger & P. S. Bayerl (Hrsg.), *Handbuch Cyberkriminologie* (S. 1–27). Springer Fachmedien. https://doi.org/10.1007/978-3-658-35450-3_29-1

Hiller, A., & Maristany de las Casas, P. (2025). *Generative KI und die deutsche extreme Rechte: Narrative, Taktiken und digitale Strategien.* Institute for Strategic Dialogue (ISD). https://isdgermany.org/wp-content/uploads/2025/02/The-use-of-generative-AI-by-the-German-Far-Right.pdf

Hu, K. (2023, 2. Februar). ChatGPT sets record for fastest-growing user base—Analyst note. *reuters.com.* https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/

Humm, B. G., Lingner, S., Schmidt, J. C. & Wendland, K. (2021). KI-Systeme: Aktuelle Trends und Entwicklungen aus Perspektive der Technikfolgenabschätzung. *TATuP – Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis, 30*(3), 11–16. https://doi.org/10.14512/tatup.30.3.11

Hunter, S. T., Walters, K., Nguyen, T., Manning, C. & Miller, S. (2022). Malevolent creativity and malevolent innovation: A critical but tenuous linkage. *Creativity Research Journal, 34*(2), 123–144. https://doi.org/10.1080/10400419.2021.1987735

Jäkel, F. (2025). *Die intelligente Täuschung. Über die Fähigkeiten Künstlicher Intelligenz.* transcript Verlag. https://www.transcript-verlag.de/978-3-8376-7752-2/die-intelligente-taeuschung/

Kaufmann, S. (2016). Security through technology? Logic, ambivalence and paradoxes of technologised security. *European Journal for Security Research, 1*(1), 77–95. https://doi.org/10.1007/s41125-016-0005-1

Klinkhammer, D. (2024). Misuse of large language models: Exploiting weaknesses for target-specific outputs. *TATuP – Zeitschrift Für Technikfolgenabschätzung in Theorie Und Praxis, 33*(2), 29-34. https://doi.org/10.14512/tatup.33.2.29

Koblentz-Stenzler, L. & Klempner, U. (2023). *Navigating extremism in the era of artificial intelligence.* The International Institute for Counter-Terrorism (ICT). https://ict.org.il/navigating-extremism-in-the-era-of-artificial-intelligence/

Kruschinski, S., Jost, P., Fecher, H. & Scherer, T. (2025). *Künstliche Intelligenz in politischen Kampagnen* (No. 75; OBS-Arbeitspapier). Otto Brenner Stiftung. https://www.otto-brenner-stiftung.de/fileadmin/user_data/stiftung/02_Wissenschaftsportal/03_Publikationen/AP75_KI_Kampagnen.pdf

Kusche, I. (2023). Artificial intelligence and/as risk. In P. Klimczak & C. Petersen (Hrsg.), *AI – Limits and Prospects of Artificial Intelligence* (S. 143–162). transcript Verlag. https://doi.org/10.14361/9783839457320-007

Kusche, I. (2024). Politische Öffentlichkeit, Desinformation und das Problem von Deepfakes. In A. Bahr & G. Fröhlich (Hrsg.), *„Ain't Nothing Like the Real Thing?" Formen und Funktionen medialer Artefakt-Authentifizierung* (S. 149–168). transcript Verlag.

Lacy, L. (2024). *OpenAI can now turn words into ultra-realistic videos.* CNET. https://www.cnet.com/tech/openai-can-now-turn-words-into-ultra-realistic-videos/

Lewis, B., Farkas, J. & Maloney, M. (2024). Rabbit hole: creating the concept of algorithmic radicalization. In J. Farkas & M. Maloney (Hrsg.), *Digital Media Metaphors* (S. 90–102). Routledge. https://doi.org/10.4324/9781032674612

Lindner, R. (2025, 9. Juli). *Grok: Musk-KI empört mit antisemitischen Ausfällen und Lob für Hitler.* FAZ.NET. https://www.faz.net/aktuell/wirtschaft/unternehmen/grok-musk-ki-empoert-mit-antisemitischen-ausfaellen-und-lob-fuer-hitler-110583280.html

Loeser, C. (2015). From paper terrorists to cop killers: The sovereign citizen threat. *North Carolina Law Review, 93*(4), 1106.

Luhmann, N. (1984). *Soziale Systeme. Grundriß einer allgemeinen Theorie.* Suhrkamp.

Madeira, O., Plattner, G., Gazos, A., Röller, T. & Büscher, C. (2023). Technologiemonitoring: Das Potenzial von Metaverse und KI für extremistische Verwendungszwecke. In U. Kemmesies, P. Wetzels, B. Austin, C. Büscher, A. Dessecker, S. Hutter & D. Rieger (Hrsg.), *Motra-Monitor 2022* (S. 226–252). BKA.

Mahfoud, T., Aicardi, C., Datta, S. & Rose, N. (2018). The limits of dual use. *Issues in Science and Technology, 34*(4), 73–78.

March, J. G. & Simon, H. (1993). *Organizations* (2. Aufl.). Blackwell Publishers.

McGuffie, K. & Newhouse, A. (2020). *The Radicalization Risks of Gpt-3 and Advanced Neural Language Models.* Center on Terrorism, Extremism, and Counterterrorism. https://www.middlebury.edu/institute/sites/default/files/2020-09/The_Radicalization_Risks_of_GPT_3_and_Advanced_Neural_Language_Models_0.pdf

Montasari, R. (2024). The impact of technology on radicalisation to violent extremism and terrorism in the contemporary security landscape. In R. Montasari (Hrsg.), *Cyberspace, cyberterrorism and the international security in the fourth industrial revolution: threats, assessment and responses* (S. 109–133). Springer International Publishing. https://doi.org/10.1007/978-3-031-50454-9_7

Nelu, C. (2024). *Exploitation of Generative AI by Terrorist Groups.* International Centre for Counter-Terrorism – ICCT. https://icct.nl/publication/exploitation-generative-ai-terrorist-groups

Newhouse, A., Blazakis, J. & McGuffie, K. (2019). *The industrialization of terrorist propaganda. Neural language models and the threat of fake content generation* (Center on Terrorism, Extremism, and Counterterrorism Report). Center on Terrorism, Extremism, and Counterterrorism. https://www.middlebury.edu/institute/sites/default/files/2019-11/The%20Industrialization%20of%20Terrorist%20Propaganda%20-%20CTEC.pdf

Oltmann, S. (2015). Dual Use Research: Investigation across multiple science disciplines. *Science and Engineering Ethics, 21*(2), 327–341. https://doi.org/10.1007/s11948-014-9535-y

Pelzer, R. (2018). Policing of terrorism using data from social media. *European Journal for Security Research, 3*(2), 163–179. https://doi.org/10.1007/s41125-018-0029-9

Piazza, J. A. & Guler, A. (2021). The online caliphate: Internet usage and ISIS support in the arab world. *Terrorism and Political Violence, 33*(6), 1256–1275. https://doi.org/10.1080/09546553.2019.1606801

Picciotto, R. (2023). *Elon Musk debuts „Grok" AI bot to rival ChatGPT, others.* CNBC. https://www.cnbc.com/2023/11/05/elon-musk-debuts-grok-ai-bot-to-rival-chatgpt-others-.html

Proctor, D. (2022). The #Tradwife persona and the rise of radicalized domesticity. *Persona Studies, 8*(2), 7–26. https://doi.org/10.21153/psj2022vol8no2art1645

Rassler, D. & Veilleux-Lepage, Y. (2024). The paradox of progress: How 'disruptive,' 'dual-use,' 'democratized,' and 'diffused' technologies shape terrorist innovation. *TATuP – Journal for Technology Assessment in Theory and Practice, 33*(2), 22–28. https://doi.org/10.14512/tatup.33.2.22

Schick, U. (2018). *Was ist künstliche Intelligenz?*. SAP News Center. https://news.sap.com/germany/2018/03/was-ist-kuenstliche-intelligenz/

Schroeter, M. (2020). *Artificial Intelligence and Countering Violent Extremism: A Primer.* GNET. https://gnet-research.org/2020/09/28/artificial-intelligence-and-countering-violent-extremism-a-primer/

Scrivens, R. & Conway, M. (2019). The roles of „old" and „new" media tools and technologies in the facilitation of violent extremism and terrorism. In R. Leukfeldt & T. J. Holt (Hrsg.), *The Human Factor of Cybercrime* (S. 286–309). Routledge. https://doi.org/10.4324/9780429460593

Siegel, D. (2023). *'RedPilled AI': A new weapon for online radicalisation on 4chan.* GNET. https://gnet-research.org/2023/06/07/redpilled-ai-a-new-weapon-for-online-radicalisation-on-4chan/

Siegel, D. & Doty, M. B. (2023). *Weapons of mass disruption: Artificial intelligence and the production of extremist propaganda* (GNET Insights). GNET. https://gnet-research.org/2023/02/17/weapons-of-mass-disruption-artificial-intelligence-and-the-production-of-extremist-propaganda/

Southern Poverty Law Center (2025). *Sovereign citizens movement.* Southern Poverty Law Center. https://www.splcenter.org/resources/extremist-files/sovereign-citizens-movement/

Sykes, S. & Hopner, V. (2024). Tradwives: right-wing social media influencers. *Journal of Contemporary Ethnography, 53*(4), 453–487. https://doi.org/10.1177/08912416241246273

UNICRI & UNCCT (2021). *Algorithms and terrorism: The malicious use of artificial intelligence for terrorist purposes.* United Nations Interregional Crime and Justice Research Institute; United Nations Office of Counter-Terrorism. https://unicri.it/News/Algorithms-Terrorism-Malicious-Use-Artificial-Intelligence-Terrorist-Purposes

Veilleux-Lepage, Y. (2024). Printing terror: An empirical overview of the use of 3D-printed firearms by right-wing extremists. *CTC Sentinel, 17*(6), 31–45.

Weimann, G., Pack, A. T., Sulciner, R., Scheinin, J., Rapaport, G. & Diaz, D. (2024). Generating terror: The risks of generative AI exploitation. *CTC Centinel, 17*(1), 17–24.

Zeiger, S. & Gyte, J. (2020). Prevention of radicalization on social media and the internet. In A. P. Schmid (Hrsg.), *Handbook of Terrorism Prevention and Preparedness* (S. 358–395). ICCT Press. https://doi.org/10.19165/2020.6.01