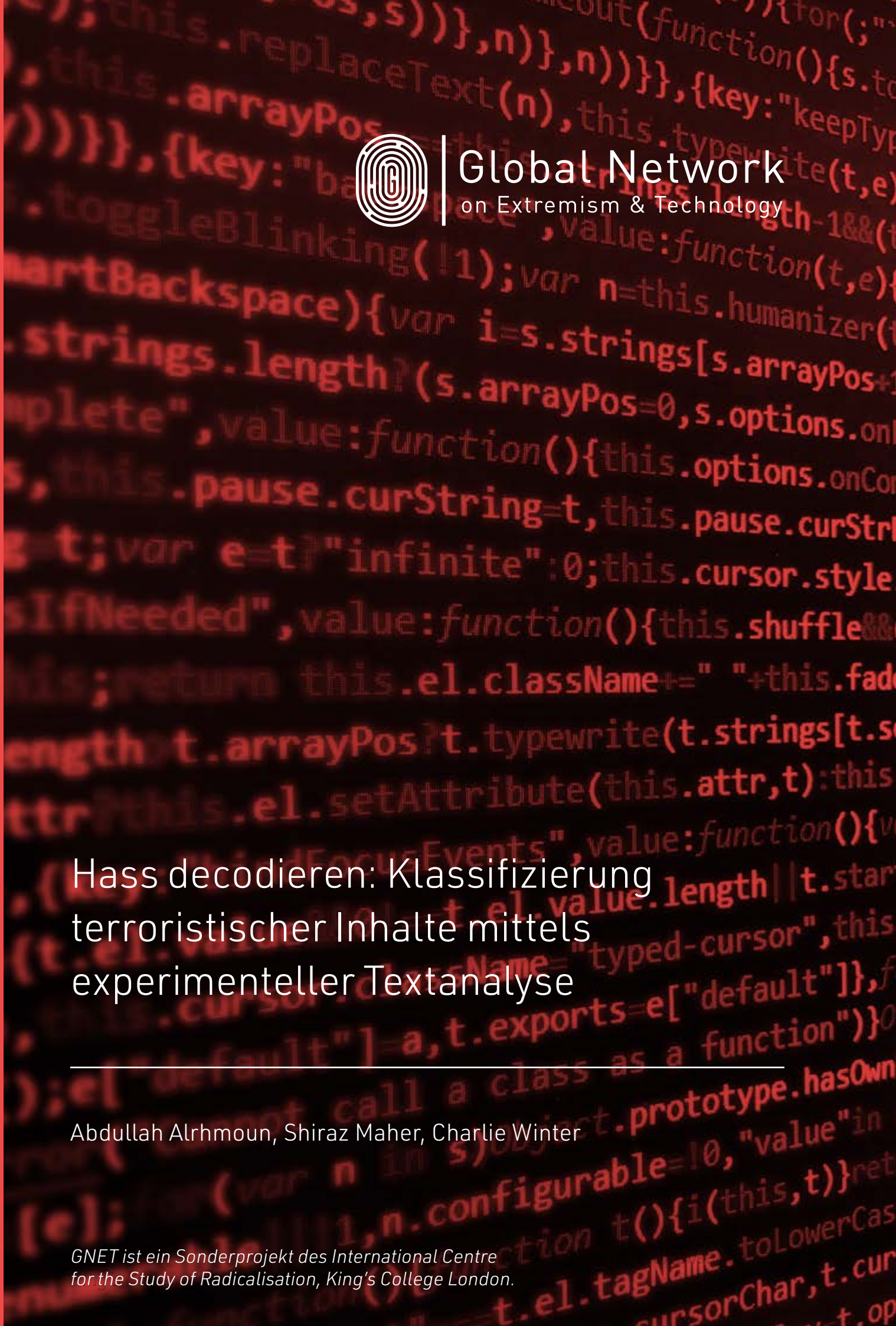


Global Network
on Extremism & Technology

Hass decodieren: Klassifizierung terroristischer Inhalte mittels experimenteller Textanalyse

Abdullah Alrhmoun, Shiraz Maher, Charlie Winter

GNET ist ein Sonderprojekt des International Centre for the Study of Radicalisation, King's College London.

Die Autoren dieses Berichts sind Abdullah Alrhmoun, Doktorand an der Central European University in Wien (Österreich), Dr. Shiraz Maher, Director des International Centre for the Study of Radicalisation (ICSR) am King's College London, und Dr. Charlie Winter, Senior Research Fellow am ICSR.

Das Global Network on Extremism and Technology (GNET) ist eine akademische Forschungsinitiative mit Unterstützung des Global Internet Forum to Counter Terrorism (GIFCT), eine unabhängige, aber von der Wirtschaft finanzierte Initiative mit dem Ziel, die Nutzung von Technologie für terroristische Zwecke besser zu verstehen und einzudämmen. GNET wird einberufen und geleitet vom International Centre for the Study of Radicalisation (ICSR), einem akademischen Forschungszentrum innerhalb des Department of War Studies am King's College London. Die in diesem Dokument enthaltenen Ansichten und Schlussfolgerungen sind den Autoren zuzuschreiben und sollten nicht als die ausdrücklichen oder stillschweigenden Ansichten und Schlussfolgerungen von GIFCT, GNET oder ICSR verstanden werden.

Diese Arbeit wurde durch einen Forschungspreis von Facebook im Rahmen seines Forschungsprojekts „Content Policy Research on Social Media Platforms" unterstützt. Die in diesem Dokument enthaltenen Ansichten und Schlussfolgerungen sind den Autoren zuzuschreiben und sollten nicht als die ausdrücklichen oder stillschweigenden Ansätze von Facebook verstanden werden.

KONTAKTANGABEN

Im Falle von Fragen oder zur Anforderung weiterer Exemplare wenden Sie sich bitte an:

ICSR
King's College London
Strand
London WC2R 2LS
United Kingdom

T. **+44 20 7848 2098**
E. **mail@gnet-research.org**

Twitter: **@GNET_research**

Wie alle anderen GNET-Publikationen kann auch dieser Bericht kostenlos von der GNET-Website unter www.gnet-research.org heruntergeladen werden.

© GNET

Inhalt

1 Einleitung

Gegenstand dieser Arbeit ist die automatische Textanalyse – d. h. der Prozess, mit dem unstrukturierter Text extrahiert, organisiert und in eine aussagefähige Darstellungsform umgewandelt wird –, um Tools zu entwickeln, mit denen sich Propaganda des Islamischen Staates (IS) im großen Maßstab analysieren lässt.[1] Obwohl in diesem Fall ein statisches Archiv an IS-Material verwendet wurde, sind diese Methoden grundsätzlich auf Inhalte beliebiger gewalttätiger, extremistischer Bewegungen in Echtzeit anwendbar. Die Studie ist folglich als Ergänzung der Forschungsarbeit angelegt, die sich mit den technologiegestützten Strategien von Social-Media-, Video-Hosting- und Filesharing-Plattformen im Kampf gegen Verbreiter gewalttätiger, extremistischer Inhalte befasst.[2] Generell verfolgen diese Plattformen das Ziel, Materialien von Terror- und Hassorganisationen zu entfernen, außer in sehr spezifischen Umständen (wie beispielsweise die Nutzung durch Journalisten oder Wissenschaftler).[3] In den letzten Jahren sind die kollektiven Bemühungen solcher Plattformen sehr effektiv geworden und nahezu alle terroristischen Inhalte werden schon vor ihrer Meldung wieder entfernt.[4]

Allerdings sind nicht alle terroristischen Inhalte gleich;[5] einige Materialien müssen dringender gelöscht werden als andere.[6] Die Automatisierung kann hier gewisse Probleme mit sich bringen, insbesondere wenn schädliche Inhalte mittels Technologie identifiziert und dann menschlichen Gutachtern zur endgültigen Entscheidung vorgelegt werden. Die Frage, was vorrangig zu überprüfen ist und wie und wann es entfernt wird, stellt eine ernsthafte Herausforderung dar.[7] Anders ausgedrückt: Ist es möglich, eine Technologie zu entwickeln, die das Material effektiv und genau bewertet? Es gilt, zwischen Materialien zu unterscheiden, die eine sofortige Prüfung erfordern oder die in eine Warteschlange gestellt werden können,[8] wie beispielsweise ein Foto, das ein eher harmloses Bild der terroristischen Sozialisation vermittelt, und ein Video, das grafische Gewalt zeigt. Das relative Risiko ist hier sehr unterschiedlich.

1 Justin Grimmer und Gary King, „General Purpose Computer-Assisted Clustering and Conceptualization“, Proceedings of the National Academy of Sciences, 2011. Abgerufen: https://j.mp/2nRjqbO; Gary King and Justin Grimmer, „Method and Apparatus for Selecting Clusterings to Classify A Predetermined Data Set“, USA 8,438,162 (7. Mai), 2013. Abgerufen: https://j.mp/2ovzAuR.

2 Ein Beispiel hierfür finden Sie unter: James Vincent, „UK creates machine learning algorithm for small video sites to detect ISIS propaganda“, The Verge, 13. Februar 2018. Abgerufen: https://www.theverge.com/2018/2/13/17007136/uk-government-machine-learning-algorithm-isis-propaganda.

3 Guy Rosen, „How are we doing at enforcing our community standards?“ Facebook, 15. November 2018. Abgerufen: https://newsroom.fb.com/news/2018/11/enforcing-our-community-standards-2/.

4 Monica Bickert, „Hard questions: What are we doing to stay ahead of terrorists?“ Facebook, 8. November 2018. Abgerufen: https://newsroom.fb.com/news/2018/11/staying-ahead-of-terrorists/.

5 Siehe zum Beispiel: Charlie Winter, „Understanding jihadi stratcom: The case of the Islamic State“, Perspectives on Terrorism Band 13:1 (2019): 54–62; Charlie Winter und Dounia Mahlouly, „A tale of two caliphates: Comparing the Islamic State's internal and external messaging priorities“, VOX-Pol, Juli 2019; Stephane Baele und Charlie Winter, „From music to books, from pictures to numbers: The forgotten—yet crucial—components of IS' propaganda“, in Stephane Baele, Travis Coane und Katharine Boyd (Hg.), The Propaganda of the Islamic State, Oxford: Oxford University Press (2019).

6 Bickert, „Hard questions“.

7 Alex Schulz und Guy Rosen, „Understanding the Facebook: Community standards enforcement report“, Facebook, Mai 2020. Abgerufen: https://fbnewsroomus.files.wordpress.com/2018/05/understanding_the_community_standards_enforcement_report.pdf. S.17.

8 Bickert, „Hard questions“.

Dazu kommt, dass Automatisierung bisher gewöhnlich auf den *Kontext* ausgerichtet war, in dem Beiträge in den sozialen Medien existieren, statt auf den *Inhalt* selbst. Infolgedessen werden Tools, die für Moderatoren bei Technologieunternehmen am praktikabelsten sind, bei solchen Untersuchungen nicht angemessen genutzt.

Dieses Projekt untersucht, wie Technologieunternehmen zeitnah und zuverlässig über solche Unterschiede entscheiden können. Am Beispielfall IS wird die Einstufung schädlicher Inhalte nuanciert. Unsere Grundprämisse besagt, dass die Absicht schädlicher Inhalte durch eine Untersuchung und Prüfung der Beweggründe für ihre Produktion kodifiziert werden kann.[9] Wenn sich klar zwischen taktischem, handlungsorientiertem Inhalt einerseits und strategischem, markenförderndem Inhalt andererseits unterscheiden lässt[10], dann wird es in der Tat möglich sein, die Überprüfung und Entfernung von Materialien nach ihrem Risikopotenzial zu priorisieren.

Zu diesem Zweck kombinieren wir in diesem Beitrag fachliches Know-how und Datenwissenschaft, indem wir unser Repositorium offizieller IS-Inhalte mithilfe experimenteller Textverarbeitung untersuchen und kategorisieren. Unser Hauptziel ist es, automatisierte Methoden zu entwickeln, die sich auf ähnliche (auch sehr viel größere) Materialsammlungen anwenden lassen, um ihre Aufschlüsselung sowie gegebenenfalls die Priorisierung ihrer Moderation und/oder Weiterleitung zu beschleunigen. Dies wiederum wird dazu beitragen, die vorhandenen Richtlinien für die Moderation von Inhalten zu verbessern.

Angesichts der enormen Menge an Inhalten, die Minute für Minute produziert werden, ist die Dringlichkeit eines solchen Ansatzes unbestreitbar. Allein auf YouTube werden pro Minute mehr als 300 Stunden an Videomaterial hochgeladen, und Nutzer sehen täglich über eine Milliarde Stunden Video an.[11] Im Durchschnitt werden tagtäglich 500 Millionen Tweets veröffentlicht, was einer Gesamtproduktion von rund 200 Milliarden pro Jahr entspricht.[12] Im ersten Quartal 2020 registrierte Facebook mehr als 2,6 Milliarden monatlich aktive Nutzer, und Instagram, das sich im Besitz von Facebook befindet, hostet täglich mehr als 500 Millionen „Instagram Stories“.[13] Die überwältigende Mehrheit der Benutzer besucht diese Plattformen natürlich für völlig harmlose und legitime Zwecke. Es geht auf keinen Fall darum, diese Inhalte in irgendeiner Weise zensieren oder überwachen zu wollen. Allerdings treten inmitten dieser Flut an Material auch böswillige, gewalttätige Extremisten auf, und deshalb sind effektive automatisierte Methoden erforderlich, um die von ihnen produzierten Inhalte zu identifizieren, zu analysieren und aufzuschlüsseln.

9 Wie Berger anmerkt, haben extremistische Bewegungen stark unterschiedliche Selbstdarstellungen, aber sehr ähnliche Strukturen, in die diese Darstellungen eingebettet sind. Es könnten also zukünftige Iterationen des Analysesystems entwickelt werden, um die Inhaltsrichtlinien und -praktiken von Facebook in Bezug auf andere Formen des Extremismus zu unterstützen. Siehe: JM Berger, Extremism, Cambridge: The MIT Press (2018): 51–112.

10 Für eine einleitende Darstellung dieser Unterscheidung siehe: Winter, „Understanding jihadi stratcom“.

11 Daten gesammelt am 11. August 2020 bei „YouTube for Press“. Abgerufen: https://www.youtube.com/intl/en-GB/about/press/.

12 David Sayce, „The number of tweets per day in 2020“, David Sayce, Mai 2020. Abgerufen: https://www.dsayce.com/social-media/tweets-day/.

13 Jessica Clement, „The number of monthly active Facebook users worldwide as of the first quarter of 2020“, Statista, 10. August 2020. Abgerufen: https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/; Maryam Mohsin, „10 Instagram Stats Every Marketer Should Know in 2020“, Oberlo, 6. Februar 2020. Abgerufen: https://www.oberlo.co.uk/blog/instagram-stats-every-marketer-should-know.

Aus diesem Grund haben wir uns auf IS-Material konzentriert, an dem das Problem gewalttätiger extremistischer Inhalte sehr deutlich wird. Generell wird Technologie von der breiteren dschihadistischen Bewegung effektiver zu Propagandazwecken eingesetzt als von anderen Bewegungen.[14] In den 1990er Jahren lieferten statische Websites wie Azzam.com englischsprachigen Zielgruppen Nachrichten über dschihadistische Kampagnen in Tschetschenien, Bosnien und Afghanistan. Nach der Irak-Invasion 2003 entwickelten sich passwortgeschützte Chat-Foren wie Ansar al-Mujahideen („Unterstützer der Mudschaheddin"), Faloja (ein Verweis auf die irakische Stadt Falludscha, die zu einer Brutstätte aufständischer Aktivitäten wurde) und Shamukh („erhaben" oder „jemand, zu dem man aufschauen kann") zu den wichtigsten Verbreitungsformen gewalttätiger extremistischer Inhalte, darunter Videos und Mitteilungen von Gruppen wie al-Qaida, al-Shabaab und Boko Haram.[15]

Diese Foren waren relativ statische und insulare Umgebungen, d. h. gewalttätige extremistische Inhalte existierten in eher unzugänglichen Bereichen des Internets und waren nur durch gezielte Suchen zu finden. Zum Zeitpunkt des Arabischen Frühlings im Jahr 2011 waren soziale Medien das wichtigste Mittel für Gewaltakteure, um sowohl Inhalte zu verbreiten als auch neue Anhänger zu gewinnen. Ihren dramatischsten Ausdruck fand diese Entwicklung mit dem Aufstieg des IS und der al-Qaida-nahen Al-Nusra-Front zwischen 2011 und 2016.[16] Das Problem war nicht nur ihre Präsenz auf diesen Plattformen, sondern auch die Tatsache, dass extremistische Inhalte nun für jeden leicht zugänglich waren, oft sogar rein zufällig. Für Technologieunternehmen, Strafverfolgungsbehörden und politische Entscheidungsträger im Bereich der Terrorismusbekämpfung wurde also die Handhabung solcher Inhalte zu einer besonders dringenden Frage.[17]

Vor diesem Hintergrund untersuchen wir in unserem Beitrag die Möglichkeiten, wie eine Automatisierung helfen könnte, diese Inhalte zu identifizieren und aufzuschlüsseln, sodass Technologieunternehmen sie inmitten anderer Materialien, veröffentlicht von legitimen Nutzern ihrer Plattformen, leichter erkennen können.

Zu Beginn sei darauf hingewiesen, dass die hier entwickelten Instrumente potenziell auch für andere, nicht dschihadistische Arten extremistischer Inhalte von Bedeutung sein können, die einzelne Unternehmen auf der Grundlage ihrer eigenen Bedürfnisse überwachen möchten. Schädliche Online-Inhalte können viele Formen annehmen, darunter Beleidigung, Mobbing, Belästigung,

14 Zur Technologie-Seite siehe: Brian Fishman, „Crossroads: Counter-terrorism and the Internet", Texas National Security Review, Februar 2019. Abgerufen: https://tnsr.org/2019/02/crossroads-counter-terrorism-and-the-internet/. Für eine behördliche Sichtweise siehe: „How social media is used to encourage travel to Syria and Iraq: Briefing note for schools", britisches Bildungsministerium, Juli 2015. Abgerufen: https://www.gov.uk/government/publications/the-use-of-social-media-for-online-radicalisation. Für eine internationale Regierungsperspektive siehe: „The use of the Internet for terrorist purposes", United Nations Office on Drugs and Crime, September 2012. Abgerufen: https://www.unodc.org/documents/frontpage/Use_of_Internet_for_Terrorist_Purposes.pdf.

15 Evan Kohlmann, „A beacon for extremists", CTC Sentinel, Februar 2010, Band 3, Ausgabe 2. Abgerufen: https://ctc.usma.edu/a-beacon-for-extremists-the-ansar-al-mujahideen-web-forum/; Manuel R. Torres-Soriano, „The Hidden Face of Jihadist Internet Forum Management: The Case of Ansar Al Mujahideen", Terrorism and Political Violence Band 28, Ausgabe 4 (2016).

16 Gunnar J. Weimann, „Competition and Innovation in a Hostile Environment: How Jabhat Al-Nusra and Islamic State Moved to Twitter in 2013–2014", Studies in Conflict & Terrorism, Band 42 (2019): 1–2, 25-42.

17 Zur Technologie-Seite siehe: Fishman, „Crossroads". Für eine Regierungsperspektive siehe: „How social media is used". Für eine internationale Regierungsperspektive siehe: „The use of the Internet for terrorist purposes".

Drohungen, Frauenfeindlichkeit, Hassreden und terroristische oder gewalttätige Propaganda, um nur einige zu nennen. Gelegentlich schlagen sich dann diese Online-Aktivitäten in realen Schäden nieder, wobei wohl keine so dramatische Formen annehmen wie die dschihadistische Gewalt.[18] Auf jeden Fall erfordern auch sie die Art von nuancierter, kontextualisierter Moderation, von der wir im Folgenden sprechen.

18 Ende 2018, nur vier Monate vor der Befreiung der letzten IS-kontrollierten Gebiete in Syrien durch mit der Koalition verbündete Streitkräfte, wurde die Nutzung sozialer Medien durch IS als direkte „Bedrohung für die Stabilität im Nahen Osten und in Afrika" angesehen. Antonia Ward, „ISIS's use of social media still poses a threat to stability in the Middle East and Africa", RAND Corporation, 11. Dezember 2018. Abgerufen: https://www.rand.org/blog/2018/12/isiss-use-of-social-media-still-poses-a-threat-to-stability.html.

2 Literaturübersicht

Es gibt bereits einen großen Fundus an Literatur, die sich mit gewalttätigen extremistischen Inhalten im Internet befasst, insbesondere mit IS. Nach seinen Anfängen in Syrien und dem Irak profilierte sich der Islamische Staat schnell als Pionier der extremistischen strategischen Kommunikation und Online-Propaganda. Wissenschaftliche Analysen dieser Materialien verfolgen generell drei Ansätze: (i) quantitative Analyse der Anhängerschaft in den sozialen Medien, (ii) qualitative Analyse individueller Propaganda-Texte oder -Genres und (iii) datenbasierte Analyse des aggregierten Medienoutputs.

Im ersten Ansatz wird untersucht, wie IS-Anhänger online miteinander interagieren. Speziell seit 2014 hat der IS-Aktivismus auf Mainstream-Plattformen wie Twitter, Facebook und YouTube große Aufmerksamkeit erregt. Die Untersuchung von Carter, Maher und Neumann war einer der ersten Versuche, Online-Beeinflussungsnetzwerke unter englischsprachigen Dschihadisten zu betrachten und diese vielfältigen Communitys zu kartieren. Es folgten ähnlich ausgerichtete Studien angesehener Kollegen wie Klausen, Berger und Morgan.[19] Nachfolgende Untersuchungen zum gleichen Thema von Conway und anderen sowie von Alexander zeigen, dass die dschihadistische Präsenz auf Mainstream-Plattformen seit 2015 zurückgegangen ist und neuere Dienste, die maximale Vertraulichkeit bieten, als bevorzugte Kommunikationsmittel an ihre Stelle getreten sind.[20] Trotz dieser Verschiebung haben Winterbotham und andere auch gezeigt, dass Mainstream-Plattformen wie Twitter und Facebook weiterhin eine Bedeutung für die Bewegung haben.[21] Darüber hinaus gibt es eine Reihe von Studien, die sich mit der idiomatischen Dynamik innerhalb extremistischer Gemeinschaften in den sozialen Medien befassen; das Ziel ist, anhand linguistischer Modelle radikalisierte Diskurse

19 Joseph A. Carter, Shiraz Maher und Peter R. Neumann, „#Greenbirds: Measuring Importance and Influence in Syrian Foreign Fighter Networks“, International Centre for the Study of Radicalisation, April 2014. Abgerufen: https://icsr.info/wp-content/uploads/2014/04/ICSR-Report-Greenbirds-Measuring-Importance-and-Infleunce-in-Syrian-Foreign-Fighter-Networks.pdf; Jytte Klausen, „Tweeting the jihad: Social Media Networks of Western Foreign Fighters in Syria and Iraq“, Studies in Conflict & Terrorism, Band 38:1: 1–22; J. M. Berger und Jonathon Morgan, „The ISIS Twitter census: Defining and describing the population of ISIS supporters on Twitter“, The Brookings Project on U.S. Relations with the Islamic World, Nr. 20, März 2015. Abgerufen: https://www.brookings.edu/wp-content/uploads/2016/06/isis_twitter_census_berger_morgan.pdf.

20 Maura Conway et al., „Disrupting Daesh: Measuring Takedown of Online Terrorist Material and Its Impacts“, Studies in Conflict & Terrorism, Band 42, 2019. Abgerufen: http://doras.dcu.ie/21961/1/Disrupting_DAESH_FINAL_WEB_VERSION.pdf; Audrey Alexander, „Digital Decay: Tracing Change Over Time Among English-Language Islamic State Sympathizers on Twitter“, Program on Extremism, George Washington University, Oktober 2017. Abgerufen: https://extremism.gwu.edu/sites/g/files/zaxdzs2191/f/DigitalDecayFinal_0.pdf; Miron Lakomy, „Mapping the online presence and activities of the Islamic State's unofficial propaganda cell: Ahlut-Tawhid Publications“, Security Journal, nur online.

21 Ugur Kursuncu et al., „Modeling Islamist Extremist Communications on Social Media Using Contextual Dimensions: Religion, Ideology and Hate“, Protokoll des ACM on Human-Computer Interaction, 3:1, August 2019; Moustafa Ayad, „‚The Baghdadi Net': How a Network of ISIL-Supporting Accounts Spread across Twitter“, Institute for Strategic Dialogue, November 2019. Abgerufen: https://www.voxpol.eu/download/report/E28098The-Baghdadi-NetE28099-How-A-Network-of-ISIL-Supporting-Accounts-Spread-Across-Twitter.pdf; Leevia Dillon et al., „A comparison of ISIS foreign fighters and supporters social media posts: an exploratory mixed-method content analysis“, Behavioural Sciences of Terrorism and Political Aggression, nur online; Airbus Defence and Space, „Mapping Extremist Communities: A Social Network Analysis Approach“, NATO Strategic Communications Centre of Excellence, Januar 2020. Abgerufen: https://www.voxpol.eu/download/report/web_stratcom_coe_mapping_extremist_strategies_31.03.2020_v2.pdf.

zu ernennen und teils auch Verhaltensweisen zu prognostizieren.[22] Diese Studien sind im Moment noch experimenteller Natur und nicht auf bestimmte Gruppen ausgerichtet.

Der zweite Ansatz beinhaltet die qualitative Untersuchung einzelner propagandistischer Produkte und Genres. In den letzten Jahren haben sich unzählige Studien mit den fremdsprachigen Magazinen des IS beschäftigt, einige davon auch mit der offiziellen arabischen IS-Publikation *al-Naba*.[23] Forscher wie Winkler und andere sowie Adelman haben sich schwerpunktmäßig auf die Hunderte von Infografiken konzentriert, die der IS seit 2015 veröffentlicht hat, während andere wie Nanninga, Dauber und Robinson zur Videoproduktion forschen.[24] El Damanhoury und Milton gehören zu den wenigen, die das umfangreiche Archiv von IS-Standbildern untersucht haben, über das noch viel mehr gesagt werden kann und sollte.[25] Trotz der thematischen Vielfalt kommen diese genrebasierten Studien generell zu ähnlichen Schlussfolgerungen hinsichtlich der Dominanz westlicher visueller Motive in der IS-Propaganda.

Der dritte Ansatz findet sich in den Archivarbeiten von Forschern wie Zelin, Milton und Winter rund um die offizielle IS-Medienproduktion.[26] Ihre Ergebnisse decken sich im Großen und Ganzen: Sie alle stellen einen Netto-Rückgang der IS-produzierter Propaganda fest, in etwa parallel mit dem Schrumpfen der IS-kontrollierten Gebiete seit 2015, was auch Nanninga anmerkt. Es muss jedoch betont werden, dass dieser Rückgang nicht notwendigerweise

22 Tom De Smedt et al., „Automatic Detection of Online Jihadist Hate Speech", Computation and Language Band 7:1–31, 2018; Adam Bermingham et al., „Combining Social Network Analysis and Sentiment Analysis to Explore the Potential for Online Radicalisation", 2009 International Conference on Advances in Social Network Analysis and Mining, 2009: 231–6; Edna Reid et al., „Collecting and Analyzing the Presence of Terrorists on the Web: A Case Study of Jihad Websites", International Conference on Intelligence and Security Informatics, 2005: 402–11; Enghin Omer, „Using machine learning to identify jihadist messages on Twitter", Uppsala Universitet, 2015. Abgerufen: http://www.diva-portal.org/smash/get/diva2:846343/FULLTEXT01.pdf.

23 Haroro J. Ingram, „An analysis of Islamic State's Dabiq Magazine", Australian Journal of Political Science, Band 51:3, 2016: 458–577; Julian Droogan and Shane Peattie, „Mapping the thematic landscape of Dabiq magazine", Australian Journal of Political Science, Band 71:6, 2017: 591–620; Haroro J. Ingram, „An analysis of Inspire and Dabiq: Lessons from AQAP and Islamic State's propaganda war", Studies in Conflict & Terrorism, Band 40:5, 2017: 357–75; Nuria Lorenzo-Dus et al., „Representing the West and ‚non-believers' in the online jihadist magazines Dabiq and Inspire", Critical Studies on Terrorism, Band 11:3, 2018; Carol K. Winkler et al., „The medium is terrorism: Transformation of the about to die trope in Dabiq", Terrorism and Political Violence, Band 31:2, 2019; Peter Wignell et al., „Under the shade of AK47s: a multimodal approach to violent extremist recruitment strategies for foreign fighters", Critical Studies on Terrorism, Band 10:3, 2017: 429–52; Logan Macnair and Richard Frank, „Changes and stabilities in the language of Islamic state magazines: A sentiment analysis", Dynamics of Asymmetric Conflict, Band 11:2, 2018: 109–20; Orla Lehane et al., „Brides, black widows and baby-makers; or not: an analysis of the portrayal of women in English-language jihadi magazine image content", Critical Studies on Terrorism, Band 11:3, 2018; Dounia Mahlouly und Charlie Winter, „A Tale of Two Caliphates: Comparing the Islamic State's Internal and External Messaging Priorities", VOX-Pol, 2018. Abgerufen: https://www.voxpol.eu/download/vox-pol_publication/A-Tale-of-Two-Caliphates-Mahlouly-and-Winter.pdf; Miron Lakomy, „Towards the ‚olive trees of Rome‚: Exploitation of propaganda devices in the Islamic State's flagship magazine ‚Rumiyah'", Small Wars & Insurgencies Band 31:3, 2020: 540–68; Michael Zekulin, „From Inspire to Rumiyah: does instructional content in online jihadist magazines lead to attacks?", Behavioural Sciences of Terrorism and Political Aggression, nur online.

24 Pieter Nanninga, „Meanings of savagery", in Lewis, J. (Hg.), The Cambridge Companion to Religion and Terrorism, Cambridge: Cambridge University Press, 2017: 172–90; Cori E. Dauber und Mark Robinson, „ISIS and the Hollywood visual style", Jihadology, 6. Juli 2015. Abgerufen: https://jihadology.net/2015/07/06/guest-post-isis-and-the-hollywood-visual-style/; Cori E. Dauber et al., „Call of Duty: Jihad – How the Video Game Motif has Migrated Downstream from Islamic State Propaganda Videos", Perspectives on Terrorism Band 13:3, 2019; Pieter Nanninga, „Branding a Caliphate in Decline: The Islamic State's Video Output (2015–2018)", International Centre for Counter-terrorism – Den Haag, April 2019. Abgerufen: https://icct.nl/publication/branding-a-caliphate-in-decline-the-islamic-states-video-output-2015-2018/.

25 Kareem El Damanhoury et al., „Examining the military-media nexus in ISIS's provincial photography campaign", Dynamics of Asymmetric Conflict, Band 11:2, 2018: 89–108; Daniel Milton, „Fatal attraction: Explaining variation in the attractiveness of Islamic State propaganda", Conflict Management and Peace Science Band 37:4, 2018; Carol Winkler et al., „Intersections of ISIS media leader loss and media campaign strategy: A visual framing analysis", Media, War & Conflict, 2019, nur online.

26 Aaron Y. Zelin, „Picture Or It Didn't Happen: A Snapshot of the Islamic State's Media Output", Perspectives on Terrorism Band 9:4, 2015; Daniel Milton, „Communication Breakdown: Unraveling the Islamic State's Media Efforts", Combating Terrorism Center at West Point, 2016. Abgerufen: https://ctc.usma.edu/communication-breakdown-unraveling-the-islamic-states-media-efforts/; Daniel Milton, „Down, but Not Out: An Updated Assessment of the Islamic State's Visual Propaganda", Combating Terrorism Center at West Point, 2018. Abgerufen: https://ctc.usma.edu/down-but-not-out-an-updated-examination-of-the-islamic-states-visual-propaganda/; siehe auch: Charlie Winter, „Apocalypse, later: A longitudinal study of the Islamic State brand", Critical Studies in Media Communication, Band 35:1, 2018: 103–21.

durch die Gebietsverluste verursacht wurde, auch wenn beides miteinander zu korrelieren scheint. Obwohl sich diesbezüglich eine Reihe intuitiver Schlussfolgerungen ziehen lassen, besteht keine endgültige Übereinstimmung darüber, worauf die Verlangsamung zurückzuführen ist, und es überrascht kaum, dass sich der IS nie mit dieser Frage befasst hat.[27]

Die vorliegende Studie setzt an den Arbeiten nach dem ersten und dritten Ansatz an. Hinsichtlich des ersten Ansatzes liefert sie eine neue Form der experimentellen Textverarbeitung, bei der nicht Social-Media-Posts im Mittelpunkt stehen, sondern der Inhalt selbst. In Bezug auf den dritten Ansatz ist der Beitrag dieser Studie angesichts des archivarischen und multimedialen Charakters der Daten recht offensichtlich. Wir hoffen, der Debatte darüber, wie und warum sich die Outreach-Aktivitäten des IS in den letzten Jahren entwickelt haben, mit unserer Arbeit weitere Nuancen verleihen zu können.

27 Es sei darauf hingewiesen, dass kein vollständiger Konsens herrscht; laut einem Outlier-Bericht von Fisher ist kein solcher Rückgang der Produktivität festzustellen. Ali Fisher, „ISIS: Sunset on the ‚decline narrative'", Online Jihad, 2018. Abgerufen: https://onlinejihad.net/2018/06/01/isis-sunset-on-the-decline-narrative/.

3 Methodik

Der wichtigste Aspekt unserer Methodik ist die Art und Weise, in der das automatisierte Textanalyse-Tool entwickelt und eingesetzt wurde. Es bildet die Grundlage unseres Modells für das automatische Einlesen umfangreicher Inhalte, um dadurch potenziell schädliches Material zu identifizieren.

Die von uns verwendete Datensammlung entstammt einer statischen, von einem unbekannten IS-Anhänger (bzw. mehreren Anhängern) verwalteten Website.[28] Diese im Surface Web leicht zugängliche Website ist seit einigen Jahren unter Dschihadisten im Umlauf, wobei die Links in öffentlichen Diskussionsforen, auf gängigen Social-Media-Plattformen sowie in weniger öffentlich ausgerichteten Diensten wie Telegram erscheinen. Die komplette Website mit insgesamt 6.290 individuellen Einträgen – von Fotoreportagen und Videos über Erklärungen der Führung bis hin zu Radiomeldungen und Magazinen – wurde im Februar 2020 heruntergeladen, archiviert und gespeichert.

Entwicklung unseres thematischen Rahmens

Der von uns entwickelte Algorithmus (der unten erläutert wird) basiert auf einem analytischen Rahmen zur Aufschlüsselung von IS-Inhalten nach Themen, der von einem der Autoren, Dr. Winter, für ein früheres Projekt entwickelt wurde.[29] Themen sind in zwei Kategorien, (i) Krieg und (ii) Zivilleben, mit 22 thematischen Subkategorien unterteilt,[30] die nachfolgend aufgeführt sind. Neun beziehen sich auf Krieg und 13 auf das Zivilleben.

(i) Kriegsthemen

1. *Operationen*: Offensive Militäroperationen. Diese variieren je nach Kontext, erklärter Zielsetzung und Taktik(en). Die drei häufigsten Iterationen betreffen Bodenangriffe, IED-Operationen (Improvised Explosive Device) und Selbstmordattentate. Andere Taktiken reichen von Drohnenangriffen bis hin zu nächtlichen Überfällen aus dem Hinterhalt.

2. *Zusammenfassung*: Aggregierte Nachrichten aus den Territorien des Kalifats in Form von täglichen, wöchentlichen oder monatlichen Zusammenfassungen, oft begleitet von Statistiken.

3. *Indirekte Kriegsführung*: Angriffe auf feindliche Stellungen mit Raketen, Flugkörpern und Mörsern. In der Regel wird der Abschuss der Geschosse dargestellt, nur selten die Nachwirkungen.

---

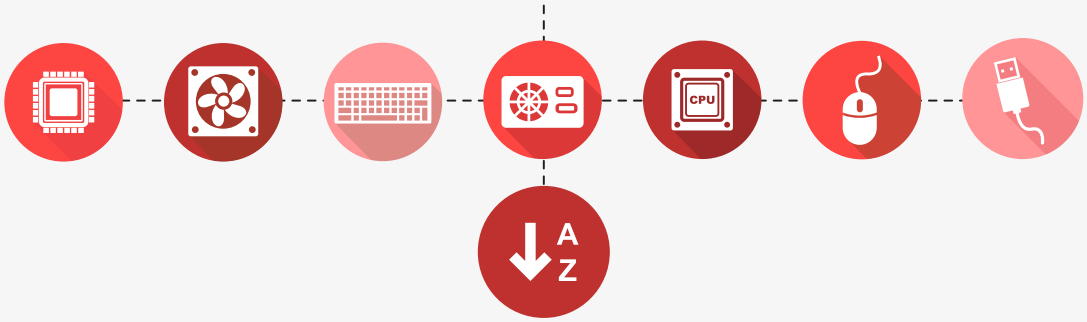
28 Wir haben uns dafür entschieden, das fragliche Archiv nicht zu nennen, weil es online weiterhin leicht zugänglich ist und wir unnötige Werbung vermeiden möchten. Leser werden gebeten, sich im Falle von Fragen direkt an die Autoren zu wenden.

29 Charlie Winter, „The Terrorist Image: A Mixed Methods Analysis of Islamic State Photo-Propaganda“, unveröffentlichte Doktorarbeit, King's College London, Juli 2020.

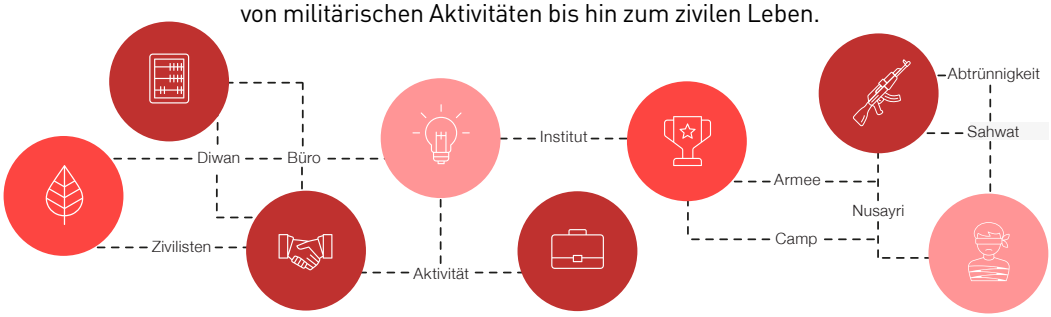
30 Siehe auch: Milton, „Communication Breakdown“; Milton, „Down, but Not Out“; Zelin, „Picture Or It Didn't Happen“; Winter, „Apocalypse, later“.

Das Archiv wurde heruntergeladen und gesichert.

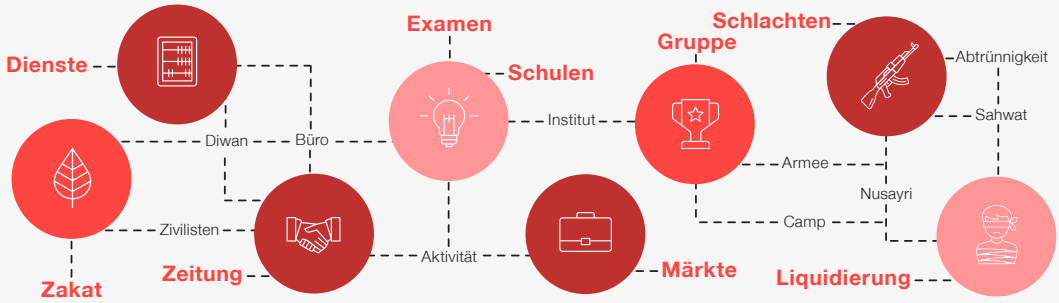
Die darin enthaltenen 6.290 Einträge wurden nach Medientyp und Veröffentlichungsdatum sortiert. Es wurden die 803 Wörter identifiziert, die am häufigsten in den Titeln vorkommen.



Diese Wörter wurden dann ein bis drei Themen zugeordnet, von militärischen Aktivitäten bis hin zum zivilen Leben.



Wörter, die sich ausschließlich auf ein Thema bezogen, wurden zu Super-Tags erklärt.



Dieses System von Tags und Super-Tags wurde dann von dem Algorithmus auf den gesamten Datenbestand angewendet.

4. *Märtyrertum*: Diese Materialien verherrlichen Männer und Jungen, die im Kampf im Namen des IS sterben. Fast alle diese gepriesenen Personen werden in den Fotos lebend dargestellt, entweder in idyllischen ländlichen oder städtischen Umgebungen oder in den Bombenfahrzeugen, in denen sie sterben werden. Sie reichen von Selbstmordattentätern über Anführer der unteren oder mittleren Ränge bis zu Propaganda-Beauftragten.

5. *Garnisondienst*: Material, das das Leben an der IS-Front verherrlicht. Im Allgemeinen vermittelt es einen Überblick über den Alltag in den IS-Stützpunkten. Im Vordergrund stehen Dinge wie Gebetszeit, Essenszubereitung, Waffenreinigung und körperliche Aktivitäten.

6. *Hinrichtungen*: Material, das die Hinrichtung von „„Spionen“ oder Kriegsgefangenen dokumentiert, die bei Entführungen oder Angriffen gefangen genommen wurden. Dazu gehören auch Mitglieder rivalisierender extremistischer Gruppen. Propaganda zu Hinrichtungen in einem offenkundig kriegerischen Kontext sollte nicht mit Propaganda verwechselt werden, die sich auf Hinrichtungen in einem zivilen Kontext bezieht.

7. *Defensive Operationen*: Defensive militärische Aktivitäten. Die meisten dieser Materialien drehen sich um vereitelte feindliche Offensiven und „„erfolgreiche“ Gegenangriffe. Andere Beiträge informieren über Flugabwehr, militärische Bereitschaftsmaßnahmen, den Bau von Befestigungen und die Wartung von Waffensystemen.

8. *Folgen*: Die Folgen eines Anschlags. In dieser Kategorie gibt es vier klare Gruppen: Kriegsbeute, feindliche Gefangene, feindliche Gefallene sowie beschädigte Bodenfahrzeuge und Drohnen.

9. *Ausbildung*: Diese Materialien zeigen Trainingslager mit Aktivitäten wie Waffenübungen, Fitnesstraining und Nahkampf.

(ii) Themen des zivilen Lebens

10. *Recht und Ordnung*: Administration von Recht und Ordnung in IS-Territorien. Es gibt drei Hauptgruppen: Bilder von der Arbeit der Religionspolizei (*Hisba*), Bilder von Strafverfahren und Bilder von Polizeieinsätzen.

11. *Opfertum*: Diese Materialien dokumentieren die Folgen von Angriffen auf IS-Territorien; sie zeigen in der Regel tote oder verletzte Kinder und zerstörte öffentliche Infrastruktur. Ihr Zweck ist, die IS-Herrschaft zu rechtfertigen, Unterstützung zu gewinnen und zu gewalttätigen Vergeltungsmaßnahmen anzustiften.

12. *Outreach*: Dieses Material folgt zum Großteil der Arbeit von Medienbeauftragten, vor allem deren Verbreitung von Inhalten und Schaffung medialer Infrastruktur. Andere Materialien zeigen Aktivitäten wie Versöhnungstreffen zwischen rivalisierenden Stämmen und Zusammenkünfte mit Zivilisten und Würdenträgern.

13. *Führungen*: Diese Materialien dokumentieren den „Alltag“ in IS-Hochburgen und zeigen zumeist geplante Besuche in spezifischen Dörfern, Städten oder Stadtvierteln. Inhalte dieser Kategorie decken diverse Aspekte ab, von der Religionsausübung und Vogelwelt bis hin zu Handel und Freizeitaktivitäten.

14. *Religiöses Leben*: Diese Materialien konzentrieren sich hauptsächlich auf religiöse Aktivitäten und zeigen Zivilisten in verschiedensten „islamischen“ Kontexten, von Eid- und Ramadan-Feierlichkeiten bis hin zu Freitagsgebeten und Koran-Rezitationswettbewerben.

15. *Geschäftsleben*: Materialien, die geschäftliche Aspekte des Lebens im Kalifat zeigen. Die meisten stellen Handel und Gewerbe dar, andere wiederum Besichtigungen von Märkten, Geschäften und Ausstellungsräumen.

16. *Kommunale Dienste*: Die Bereitstellung kommunaler Dienste in IS-Gebieten. Diese sehr vielfältigen Materialien zeigen öffentliche Stellen bei diversen Arbeiten, von der Reparatur von Strommasten bis zur Wartung von Abwasserkanälen.

17. *Soziale Wohlfahrt*: Materialien dieser Kategorie veranschaulichen die Berechnung, Vorbereitung und Verteilung von Sozialhilfe (Geld und Nahrungsmittel) unter der Zivilbevölkerung in IS-kontrollierten Gebieten. Die meisten sind als „Ein Tag im Leben“-Porträts verfasst und bieten einen allgemeinen Überblick über die Aktivitäten der Fürsorgestelle.

18. *Industrielles Leben*: Materialien zu industriellen Aktivitäten im Kalifat, von Rohrfabriken und Werkstätten für Klimaanlagen bis zu Einrichtungen zur Käseproduktion und Sonnenblumenkern-Trocknung.

19. *Landwirtschaftliches Leben*: Diese Materialien decken landwirtschaftliche Tätigkeiten ab, von der Aussaat bis zur Obsternte und -vermarktung.

20. *Bildung*: Dieses Material betrifft generell die Aktivitäten der offiziellen Stellen für religiöse Bildung. Andere zeigen Schulbesuche in verschiedenster Form, von Jugendlichen bei Zwischen- und Abschlussprüfungen bis hin zu Kleinkindern, die in den Pausen spielen.

21. *Gesundheitswesen*: Diese Materialien stellen medizinische Aktivitäten unter dem IS vor. Sie reichen von Führungen durch Krankenhäuser und zahnärztliche Kliniken über Hausbesuche von Ärzten bis zu Initiativen zur Impfung von Kindern.

22. *Landschaften und Natur*: In dieser Kategorie erscheinen hauptsächlich Fotografien von Stätten natürlicher oder monumentaler Schönheit.

Entwicklung des Algorithmus

Unser Algorithmus sucht nach linguistischen Markern in den Titeln der 6.290 individuellen Einträge unserer Datenbank. Viele millenaristische und reaktionäre Bewegungen verwenden ihre eigenen lexikalischen und syntaktischen Marker, die von der Zugehörigkeit zu dieser exklusiven Gruppe zeugen. Eines der bekanntesten Beispiele für solche Marker ist die Verwendung von dreifachen Klammern um die (((Namen))) prominenter jüdischer Personen durch Mitglieder einiger Alt-Right- und Neonazi-Communitys. Der Zweck hierbei ist, Juden bzw. Menschen mit jüdischem Hintergrund klar zu identifizieren.[31] Menschen im Umkreis

31 Matthew Yglesias, „The (((echo))) explained“, Vox, 6. Juni 2016. Abgerufen: https://www.vox.com/2016/6/6/11860796/echo-explained-parentheses-twitter.

solcher Alt-Right- und Neonazi-Bewegungen wüssten dann, dass diese Person unter dem Blickwinkel antisemitischer Verschwörungen oder Intrigen zu betrachten ist.

Ein weiteres Beispiel für die sprachlichen Eigenheiten bestimmter Online-Subkulturen findet sich in der Incel-Szene (Incel = „Involuntary Celibate"), in der sich sexuell inaktive Männer mit einem Hass gegen Frauen und den Feminismus zusammenschließen.[32] Mit dem Wort „Chad" beispielsweise verweist man in diesen Gruppen abfällig auf attraktive, beliebte Männer, die vermutlich viel Sex mit Frauen haben. Parallel dazu ist „Stacy" die Bezeichnung für attraktive Frauen, die nur an „Chads" interessiert sind.[33]

Auf ähnliche Weise geht der IS mit den Titeln für seine Inhalte vor. Linguistische Marker weisen auf spezifische Themen hin. Wenn man also die Sprache kennt, kann man allein vom Titel auf die thematische Natur des Inhalts schließen. Dieser Ansatz ist natürlich nicht unfehlbar, aber ausreichend für unser Ziel, Methoden zur Identifizierung, Klassifizierung und Aufschlüsselung extremistischer Inhalte sowie deren anschließende, priorisierte Prüfung durch Menschen zu finden. Deshalb haben wir unsere Textverarbeitungstools auf dieser Grundlage ausgerichtet und erweitert.

Vor der Entwicklung des Algorithmus mussten wir jedoch zunächst die Gültigkeit unseres Rahmenwerks belegen. Dazu haben wir ein Codebuch erstellt und auf Intercoder-Reliabilität getestet.[34] Eine Zufallsstichprobe von 286 Einträgen (d. h. 5 % des gesamten Materialbestands) wurde von drei Wissenschaftlern des ICSR unabhängig voneinander codiert. Alle drei waren arabischsprachig und mit IS-Inhalten vertraut. Bis auf 41 Fälle (14 % der Stichprobe) wurden alle Einträge gleich codiert. Die restlichen Einträge wurden von den Codierern individuell untereinander besprochen und abgestimmt. Die meisten dieser Abweichungen ergaben sich aus unterschiedlichen Auslegungen der genauen syntaktischen oder kontextbezogenen Verwendung eines spezifischen arabischen Begriffs.

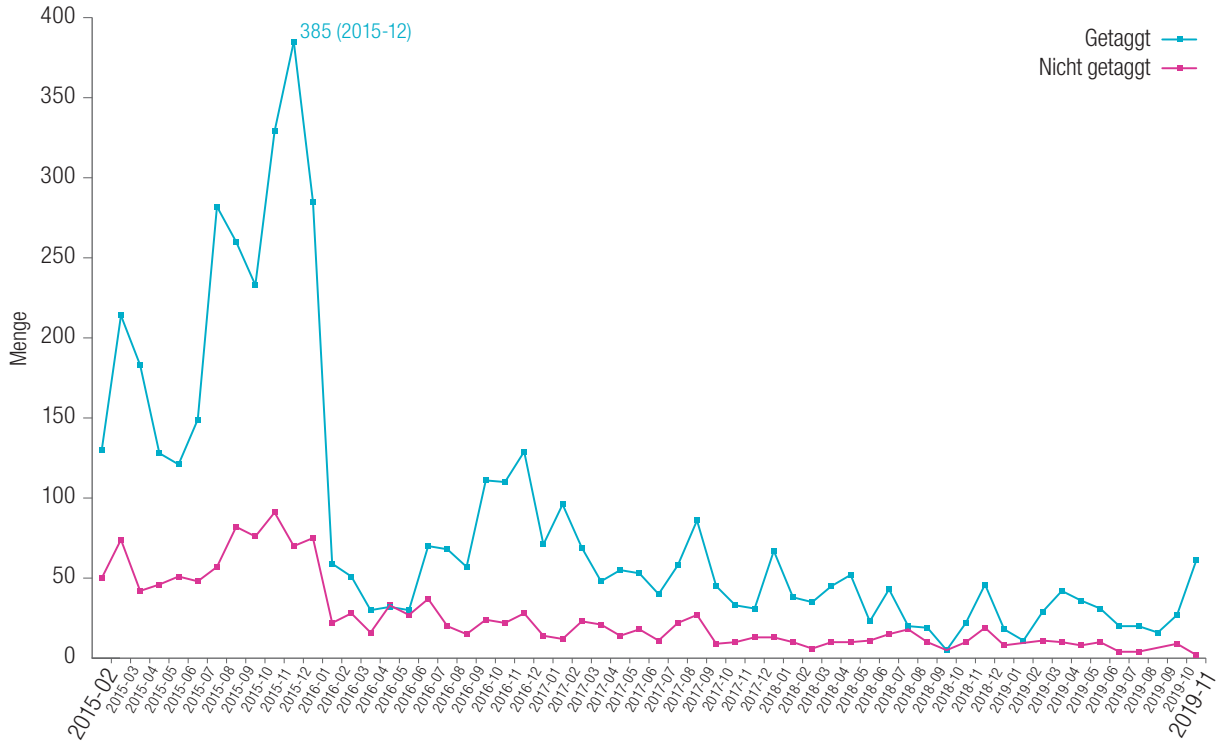
Nach Kalibrierung des Rahmenwerks mithilfe dieses Intercoder-Verfahrens programmierten wir dann unseren Algorithmus, den gesamten Bestand an archiviertem Material zu kategorisieren, indem er bestimmte Wörter mit bestimmten Themen verbindet. Dieser Prozess ist in der Infografik auf Seite 12 dargestellt. Im Hinblick auf unseren Datensatz wurden zunächst die 803 am häufigsten auftretenden Wörter identifiziert, um eine vollständige Erfassung des Korpus zu gewährleisten – d. h., dass alle 6.290 Einträge in diesem Teil des Prozesses erfasst wurden. Die Zusammenstellung der Liste war ein Prozess des Ausprobierens, bis wir die effizienteste Methode zur Erfassung des gesamten Materialbestands gefunden hatten. Diese 803 Wörter (die keine Ortsnamen oder Eigennamen enthielten) wurden dann jeweils ein bis drei Themen aus der Liste der 22 Themen unseres Rahmenwerks zugeordnet.

32 Rebecca Jennings, „Incels categorize women by personal style and attractiveness", Vox, 28. April 2018. Abgerufen: https://www.vox.com/2018/4/28/17290256/incel-chad-stacy-becky.

33 „A parent's guide to the secret language of internet extremists", CBS News, 16. März 2020. Abgerufen: https://www.cbsnews.com/news/incels-radicalization-glossary-parents-cbsn-orginals-extremists-next-door/.

34 Moin Syed und Sarah Nelson, „Guidelines for Establishing Reliability When Coding Narrative Data", Emerging Adulthood Band 3:6, 2015; Paul J. Lavrakas, Encyclopedia of Survey Research Methods, Thousand Oaks, CA: Sage Publications, 2008.

Abb. 1: Codierte Einträge ggü. uncodierten Einträgen



Wörter, die im Zusammenhang mit mehr als einem Thema erscheinen, wurden entsprechend getaggt. Beispielsweise wurde das Wort *mahud* („Institution“) mit den Themen „Bildung“, „Ausbildung und „Outreach“ verknüpft, weil es auf IS-Schulen, Militärlagern und religiöse Seminare verweisen kann. Ähnlich erhielt *riddah* („Abtrünnigkeit“) die thematischen Tags „Operationen“, „Hinrichtungen“ und „Recht und Ordnung“, da es im Zusammenhang mit allen dreien verwendet wird. Durch dieses Taggen der 803 Wörter konnten wir alle 6.290 Einträge des Korpus unter ein bis drei Themen erfassen.

Die Tatsache, dass ein Eintrag potenziell zu ein bis drei unterschiedlichen Themen passt, ist an sich noch keine nützliche Erkenntnis. Deshalb wurden über diesen anfänglichen Prozess hinaus eine Reihe von „Super-Tags“ definiert. Jeder „Super-Tag“ besteht aus einem einzigen Wort, das ausschließlich im Zusammenhang mit einem bestimmten Thema verwendet wird. So wurde das Wort *aswaq* („Märkte“) als Super-Tag für das Thema „Geschäftsleben“ identifiziert, weil es nur im Zusammenhang mit geschäftlichen Aktivitäten erscheint, sonst nicht. Wenn nun die Wörter „Institution“ und „Märkte“ im gleichen Titel erscheinen, wäre dieser aufgrund des Super-Tags „Märkte“ automatisch der Kategorie „Geschäftsleben“ zuzuordnen, ungeachtet etwaiger anderer vorhandener Begriffe. Insgesamt wurden 232 Super-Tags definiert.

Durch Anwendung dieses Super-Tag-Systems zusätzlich zur anfänglichen Textanalyse auf Grundlage der 803 linguistischen Identifikatoren konnten 4.848 Einträge (79 % des archivierten Materials) erfolgreich einem einzigen Thema zugeordnet werden. Anschließend haben wir die vom Algorithmus generierten Ergebnisse

mit denen unserer menschlichen Wissenschaftler verglichen. Der Algorithmus stimmte in 91 % der Fälle mit dem komplett abgestimmten Coding unseres menschlichen Teams überein. Dies ist zwar immer noch nicht perfekt, doch nach unserer Einschätzung war diese Fehlerquote gering genug für die vorliegende Untersuchung, die vorerst Erkundungscharakter hat.

Dem Algorithmus war es nicht möglich, das restliche Archiv mit 1.432 Einträgen zu codieren. Hierbei handelte es sich vorwiegend um Audio-Statements und Videos, deren Titel in der Regel nicht deskriptiv sind, sondern eher islamische Schriften oder andere, noch obskurere Quellen zitieren. Sie enthielten also keine der 803 linguistischen Identifikatoren oder 232 Super-Tags und wurden daher keinem Thema zugeordnet.

Der Algorithmus hatte zudem Probleme mit Titeln, die Wörter zu mehr als einem Thema enthielten, aber keine Super-Tags. So enthielt beispielsweise eine Fotoreportage mit dem Titel „Die Produktion von Booten für die Fischerei“ den Marker „Fischerei“, der mit dem Thema „Landwirtschaftliches Leben“ verbunden ist, sowie den Marker „Produktion“, der zum Thema „Industrielles Leben“ gehört. Aufgrund der Anwesenheit dieser beiden Marker ohne Super-Tag konnte der Algorithmus diesen Titel keinem einzelnen Thema zuordnen.

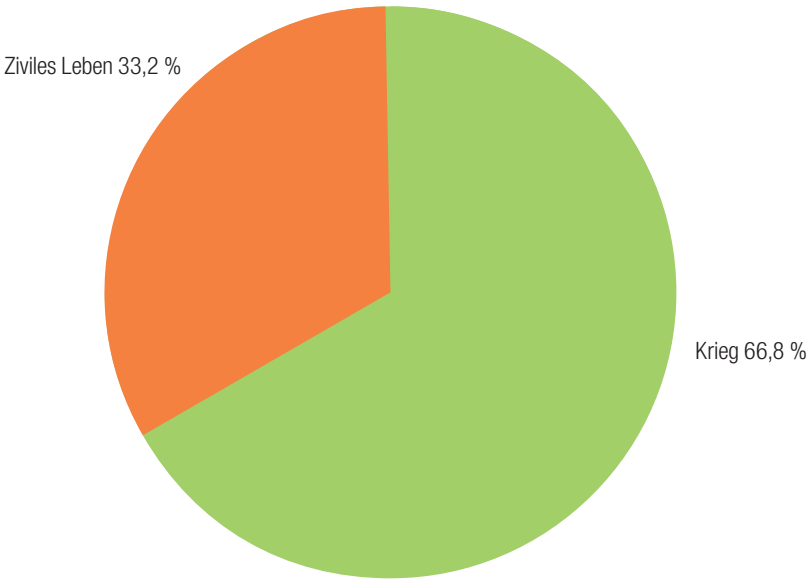
Abb. 1 zeigt die vom Algorithmus erfolgreich codierten Einträge sowie die nicht codierten Einträge. Die Korrelation der beiden Linien deutet auf die Gültigkeit des Algorithmus im gesamten Verlauf des Betrachtungszeitraums hin. Unsere Studie zeigt, dass nahezu ein Fünftel des Korpus nicht kategorisiert werden konnte, was eine erhebliche Einschränkung für unseren Ansatz darstellt. Eine Lösung würde ein vollkommen neues methodologisches Instrumentarium sowie eine Untersuchung anderer Faktoren über die linguistische Analyse hinaus voraussetzen, was den Rahmen dieser Studie sprengen würde. Wir werden in Zukunft zu dieser Frage zurückkehren.

4 Erkenntnisse

Priorisierung der Daten

Nach der algorithmischen Analyse der 6.290 Einträge in unserem Archiv stellten wir fest, dass der Algorithmus auch bei der Identifizierung und Priorisierung von Material helfen konnte. Wie Abb. 2 zeigt, bestand das Archiv zu 66,8 % aus kriegsbezogenen Themen und nur zu 33,2 % aus Themen des Zivillebens. Auf praktischer Ebene ist dies definitiv von Bedeutung für Technologieunternehmen, die spezifische Inhalte analysieren möchten, wenn auch nur mit einer ganz einfachen Unterscheidung: Krieg oder Zivilleben.

Abb. 2: Übergreifende Themen der getaggten Inhalte



Noch interessanter wird es, wenn wir uns ansehen, welche Arten von kriegsbezogenen Materialien am häufigsten auftreten. Wie Abb. 3 verdeutlicht, war das Thema „Operationen“ mit 30,3 % des Datenbestands am stärksten vertreten, gefolgt von „Zusammenfassung“ mit 10,4 %, „Indirekte Kriegsführung“ mit 7,2 % und „Hinrichtungen“ mit 3,2 %. Hinrichtungen bilden in dieser Kategorie also nicht die größte Gruppe gewalttätiger Inhalte. Dennoch könnten Unternehmen beschließen, diesem Material eine größere Dringlichkeit zuzuweisen, weil es fast immer schockierend und schädlich ist. Auch wenn die anderen Kategorien stärker verbreitet sind und sich ebenfalls auf kriegerische Inhalte beziehen, sind sie wahrscheinlich weniger oft wirklich schockierend und zeigen z. B. militärische Ausrüstung, Waffen und Ähnliches.

Wir möchten hier keinesfalls den Eindruck erwecken, als würden Technologieunternehmen nur unwirksame und stumpfe Instrumente einsetzen, um gewalttätige extremistische Inhalte auf ihren Plattformen zu identifizieren. Dies ist ganz offensichtlich nicht der Fall – auf der

Suche nach schädlichen Inhalten kommt eine große Vielzahl komplexer Instrumente zum Einsatz. Der Zweck dieses algorithmischen Tools besteht darin, diese existierenden Verfahren durch die Identifizierung linguistischer Marker zu ergänzen, um so die Klassifizierung und Priorisierung weiter zu verbessern.

Unsere Ergebnisse, dargestellt in Abb. 4, lassen zudem die häufigsten Themen rund um das Zivilleben erkennen: „Recht und

Abb. 3: Themen in kriegsbezogener Propaganda

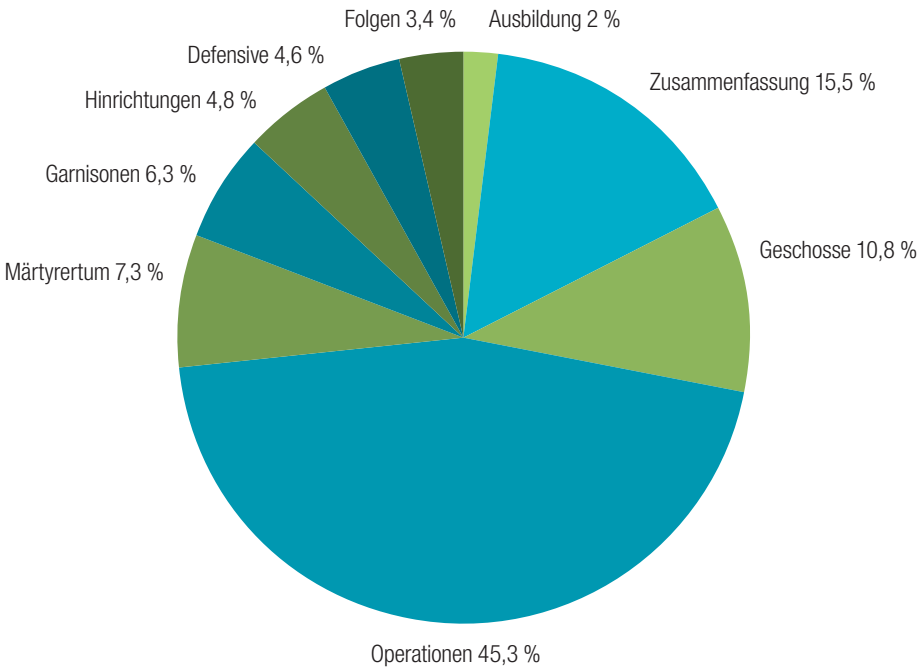
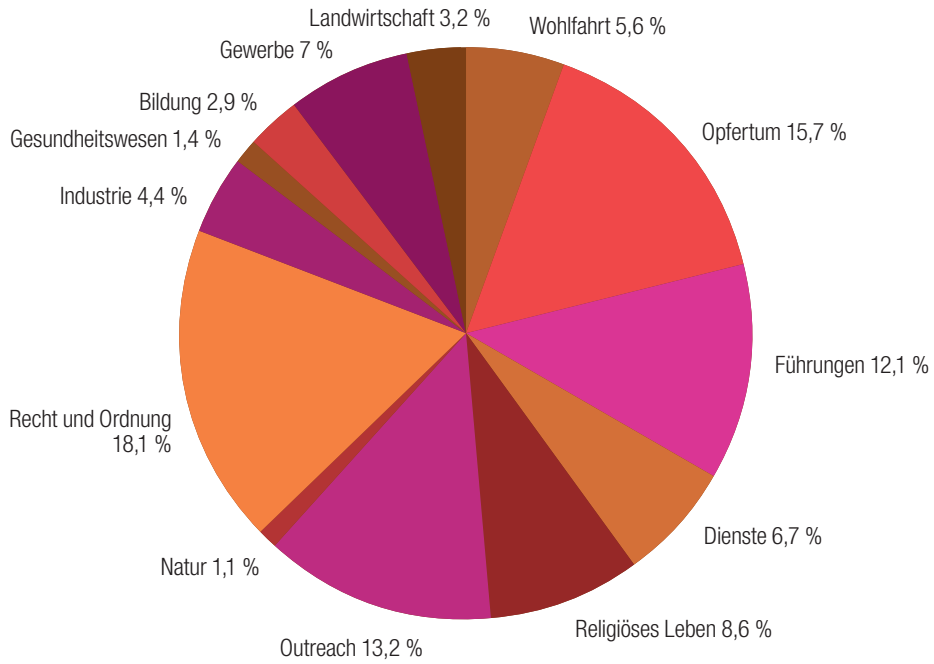


Abb. 4: Themen in Propaganda zum Zivilleben



Ordnung“ mit 6 %, „Opfertum“ mit 5,2 %, „Outreach“ mit 4,4 % und „Führungen“ mit 4 %. Da IS-Inhalte rund um Recht und Ordnung regelmäßig sowohl eher harmlose Szenen von patrouillierenden Polizeikräften zeigen als auch gewalttätige Szenen, in denen beschuldigte Kriminelle hingerichtet und/oder verstümmelt werden, sollte dieses Material für eine dringende menschliche Prüfung priorisiert werden, unmittelbar nach dem kriegsbezogenen Thema „Hinrichtungen“. Bei anderen Materialien zum Zivilleben ist eine sofortige Prüfung weniger wichtig, weil sie nur selten, wenn überhaupt, Gewalt beinhalten. Eine Ausnahme bildet hierbei Propaganda rund um Opfertum, bei der es oft um Opfer unter der Zivilbevölkerung geht.

Der Algorithmus macht es zudem möglich, Verschiebungen bei thematischen Schwerpunkten oder der Priorisierung im Laufe der Zeit zu erkennen und zu visualisieren, wie in Abb. 5 dargestellt. Ersichtlich wird ein allmählicher Rückgang an Material zum Zivilleben, das 2015 für rund 49 % der Einträge aufkam, 2019 dagegen nur für 7 %. Dies spiegelt die veränderte Situation wieder: Nach den Gebietsverlusten des IS rückte die Darstellung des angenehmen Lebens in der vermeintlichen Utopie in den Hintergrund; sie wurde ersetzt durch anderes Material von eher kriegerischer Natur nach den traditionellen dschihadistischen Tropen feindlicher „Kreuzfahrer“ im „Krieg gegen den Islam“. Das Ziel der Propaganda verlagerte sich daher von der Idee, dass Anhänger nach Syrien und in den Irak auswandern, um ihr „Kalifat“ zu unterstützen, zu dem Aufruf, zuhause zu bleiben und dort Terroranschläge zu verüben.

Abb. 5: Thematische Prioritäten der getaggten Inhalte (nach Monat)

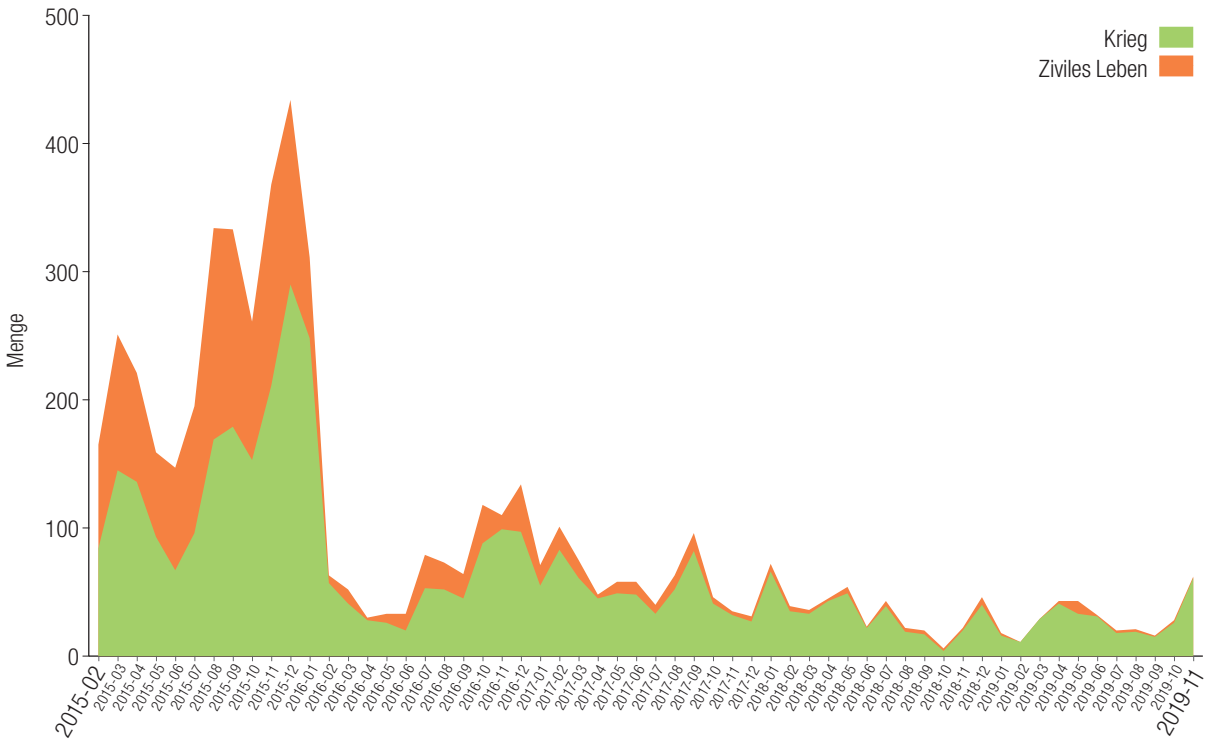


Abb. 6: Auf Krieg und auf Zivilleben bezogene Inhalte, 2015

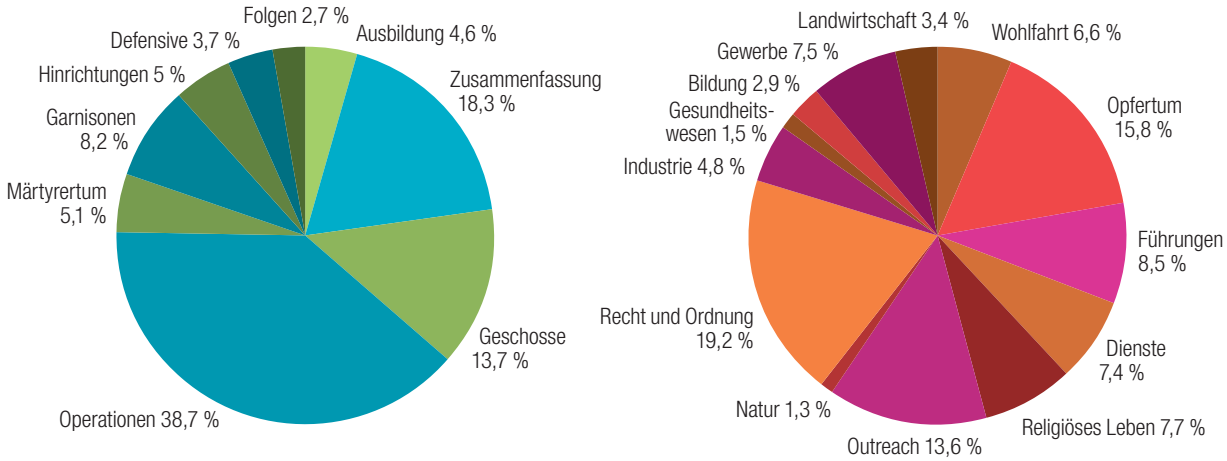
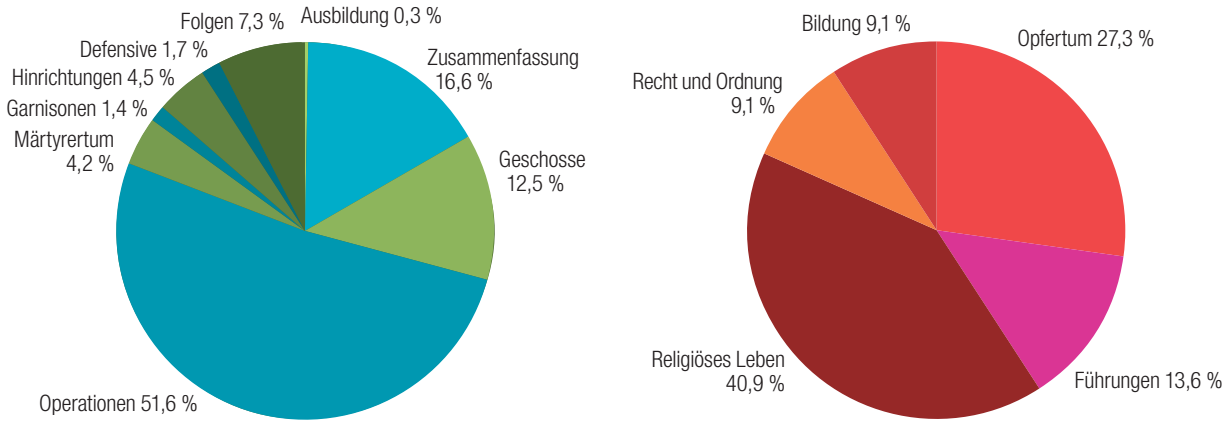


Abb. 7: Auf Krieg und auf Zivilleben bezogene Inhalte, 2019



Während dieser Zeit kam es auch zu einer allmählichen Vereinfachung des IS-Diskurses insgesamt. Parallel zum Rückgang der Produktivität von 2016 bis 2019 reduzierte sich auch die thematische Bandbreite. Dies wird aus den Daten ersichtlich. Wie Abb. 6 und 7 zeigen, setzte sich der Output im Jahr 2015 aus allen 22 der oben beschriebenen Kategorien zusammen. Im Jahr 2019 dagegen deckte der Output nur noch 14 Kategorien ab – neun davon kriegsbezogen.

Diese gleichzeitige Verschiebung und Vereinfachung der IS-Inhalte beschränkt sich keineswegs auf das untersuchte Archiv. Vielmehr ist diese Dynamik bezeichnend dafür, wie sich die offiziellen Outreach-Aktivitäten des IS den neuen, situationsbedingten Gegebenheiten vor Ort angepasst haben. Einfach ausgedrückt: Angesichts schrumpfender Gebiete und eines sich wandelnden Charakters seines Kriegs begann der IS ein neues Paradigma des Rebellenkriegs. Statt um wirtschaftliche und territoriale Gewinne zu kämpfen, kehrte man zu Aktivitäten zurück, die in erster Linie

auf eine fortschreitende Konsolidierung im Untergrund ausgerichtet waren. Dazu veränderte sich auch die propagandistische Botschaft. Es ging eher darum, Entschlossenheit und nachhaltige Präsenz zu signalisieren, als neue Kämpfer anzuwerben oder weltweite Empörung zu provozieren. Vor diesem Hintergrund fokussierten die veröffentlichten Materialien weniger auf das zivile Leben im Proto-Staat des IS als auf dessen kriegerische Aktivitäten, wie die Abb. 5, 6 und 7 zeigen.

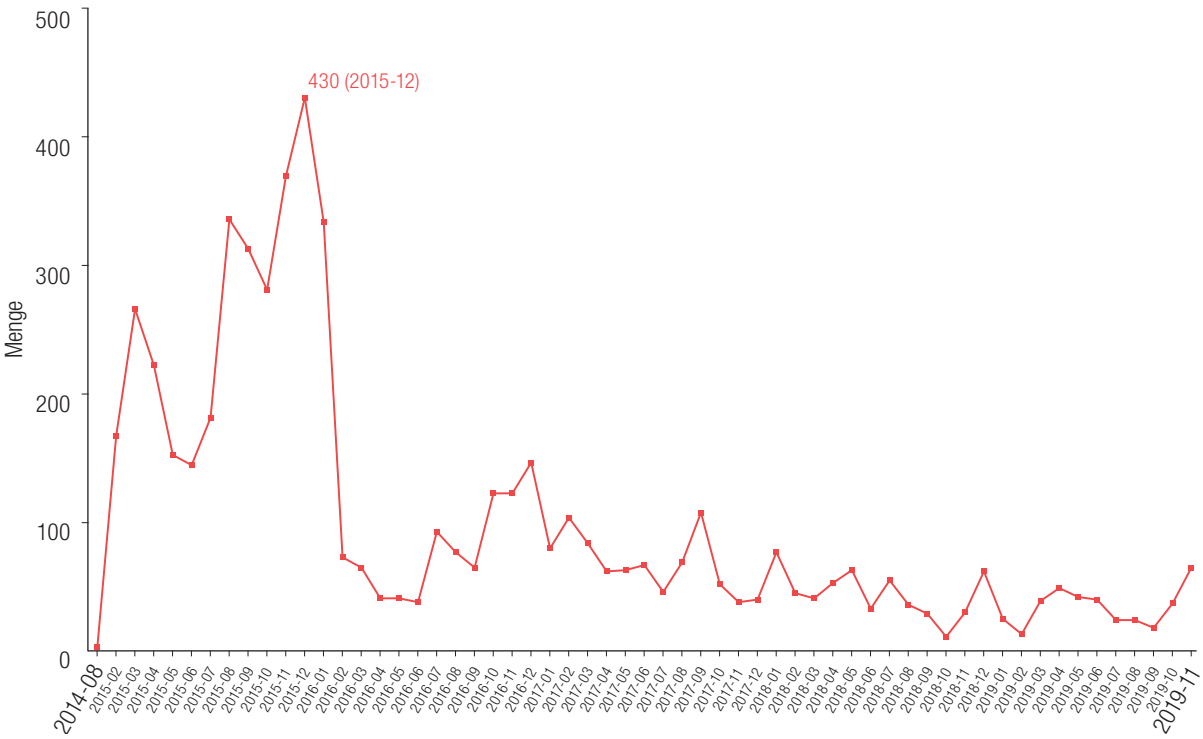
Mithilfe dieser Visualisierung, die unser Algorithmus ermöglicht, können Technologieunternehmen folglich Änderungen der thematischen Schwerpunkte und Prioritäten im Zeitverlauf verfolgen, Verschiebungen besser antizipieren und sich darauf einstellen.

Nutzen der Identifizierung temporaler Merkmale

Abb. 8 stellt die zeitliche Verteilung aller 6.290 Einträge im Archiv dar, die nach Veröffentlichungsdatum sortiert und angeordnet wurden. Dies half uns, den Umfang der für diesen Datensatz veröffentlichten Inhalte im Zeitverlauf zu verstehen und zu visualisieren. Diese Darstellung dient zur Veranschaulichung. Bei der Verfolgung von Live-Inhalten würde man eine solche Visualisierung fortlaufend in Echtzeit aktualisieren.

Wie die Kurve zeigt, stammt der Inhalt zum Großteil von 2015. Danach geht die Menge stark zurück und bleibt 2016 bis 2019 in etwa gleich. Dies sagt wahrscheinlich mehr über den partiellen Charakter des Archivs aus als über etwaige andere Aspekte.

Abb. 8: Gesamtmenge an Inhalt (nach Monat)



Wie wir wissen, ging die Produktion von IS-Propaganda in der Zeit zwischen 2017 und 2019 deutlich zurück. Wir können also mit einiger Sicherheit davon ausgehen, dass das Archiv für diese Jahre ein vollständigeres Bild liefert. Für die Jahre 2015 und 2016 erfasst das Archiv zwar erhebliche Mengen an Propaganda, aber dies ist nur ein kleiner Anteil der tatsächlichen Produktion aus dieser Zeit. Darüber hinaus enthält es nur einen einzigen Artikel aus dem Jahr 2014, in dem viele Tausende IS-Beiträgen veröffentlicht wurden. Die wahrscheinlichste Erklärung für diese Diskrepanz ist, dass die Person, die das Archiv angelegt hat, erst nach der Löschung der meisten offiziellen IS-Inhalte aus dem Surface Web damit begann. Daraus würde folgen, dass sie nur einen kleinen Teil des offiziellen Materials erfassen konnte, das in den betreffenden Jahren auftauchte.

Auf jeden Fall entspricht der in Abb. 8 erkennbare Rückgang von Jahr zu Jahr jedoch im Großen und Ganzen der Entwicklung der Medienproduktionskapazitäten des IS nach 2015. Wie zahlreiche Wissenschaftler bereits gezeigt haben, stellte das Jahr 2015 in Bezug auf die IS-Propagandaproduktion eine Blütezeit dar.[35]

Dieser Rückgang unterstreicht den starken Zusammenhang für den IS zwischen der Größe der kontrollierten Gebiete und den Kapazitäten zur Produktion von Inhalten. Als der IS in verschiedensten Formen über Millionen von Menschen herrschte, genossen seine Propagandisten nicht nur mehr Bewegungsfreiheit sowie einen besseren Zugang zu Finanzen und Personal, sondern hatten auch zahlreiche Themen zur Verfügung, um immer neues Material zu erstellen.[36] Dazu kommt, dass der IS zu seinen territorial besten Zeiten gleichzeitig mit zumeist konventionellen Mitteln an mehr als einem Dutzend Fronten kämpfte.[37] Dies lässt sich besser zu propagandistischen Zwecken ausschlachten als verdeckte Operationen, die an Bedeutung gewannen, als der IS in die Defensive gedrängt wurde.

Veränderungen bei der internen Zusammensetzung des IS haben vermutlich ebenfalls zu diesem Rückgang der Medienproduktion beigetragen. Parallel zu dem Charakter seiner Aufstandsaktivitäten änderten sich verständlicherweise auch seine Outreach-Prioritäten: Der Fokus lag weniger auf Rekrutierung und mehr auf der Bindung der lokalen Anhängerschaft.[38]

Dies ist für unsere Zwecke relevant, weil unser Algorithmus Unternehmen potenziell helfen kann, die Wirksamkeit ihrer Entfernungsbemühungen zu beobachten. Tatsache ist, dass gewisse Inhalte schwieriger zu entfernen sind oder sich einfacher verstecken lassen. Zahlreiche Studien haben gezeigt, dass der IS und seine Anhänger die Notwendigkeit verstehen, sowohl sich selbst als auch ihre Inhalte auf den Mainstream-Plattformen zu

35 Siehe zum Beispiel: Milton, „Communication Breakdown“; Milton, „Down, but Not Out“; Winter, „Apocalypse, later“; Nanninga, „Branding a Caliphate in Decline“.

36 Aaron Y. Zelin, „The Islamic State's Territorial Methodology“, The Washington Institute for Near East Policy, 2016. Abgerufen: https://www.washingtoninstitute.org/policy-analysis/view/the-islamic-states-territorial-methodology.

37 Für eine Darstellung dieser Entwicklung siehe Ahmed S. Hashim, „The Islamic State's Way of War in Iraq and Syria: From Its Origins to the Post Caliphate Era“, Perspectives on Terrorism, Band 13:1, 2019: 23–32.

38 Ein Hinweis darauf ist die Tatsache, dass der IS seit Januar 2019 keine neue Video- oder Magazin-Inhalte über das AlHayat Media Center, seine am stärksten nach außen gerichtete, fremdsprachige Propagandastiftung, veröffentlicht hat.

verhüllen.[39] Zum Großteil sind diese Bemühungen nicht erfolgreich. Dennoch kann unser Tool die Identifizierung resilienter, über längere Zeit verwendeter linguistischer Marker unterstützen, was wiederum den Nutzen einer solchen temporalen Dimension der Analyse unterstreicht.

Geografische Merkmale

Da algorithmische Tools in der Lage sind, Abzeichen, Logos und andere Formen der Markenbildung zu erkennen, haben wir unseren Datenbestand auch nach geografischen Merkmalen durchsucht. Es sei auch hier darauf hingewiesen, dass Technologieunternehmen bei einem Echtzeit-Einsatz in der Lage wären, ihre vorhandenen Instrumente zu verfeinern und stärker auf Inhalte einer bestimmten Art auszurichten. Besonders nützlich wäre dies im Zusammenhang mit einer Mobilisierung ausländischer Kämpfer, wie dies speziell zwischen 2013 und 2015 bei dem IS der Fall war: Wenn Inhalte aus der Gegend einer solchen Mobilisierung (im genannten Fall Syrien) priorisiert und entfernt werden könnten, wäre es den Moderatoren möglich, Propaganda über die angeblichen Vorteile einer Mitwirkung erheblich zu untergraben.

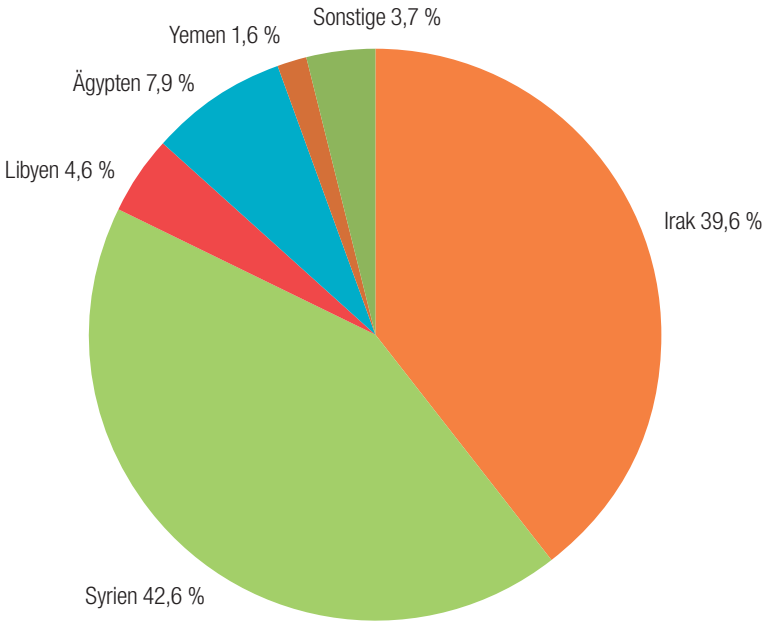
Zu diesem Zweck haben wir die Materialien nach ihrem angeblichen Herkunftsort katalogisiert, und zwar anhand der IS-Medieneinheit, die jeweils für deren Produktion verantwortlich war. In den von unserem Archiv abgedeckten Jahren betrieb der IS ein dreistufiges System für die Medienproduktion. Es umfasste zentrale Stellen wie die al-Furqan Foundation und das Al Hayat Media Center, sekundäre Medienagenturen wie die Amaq News Agency und das Furat Media Center sowie provinzielle Medieneinheiten wie die Medienbüros Wilayat al-Sham und Wilayat al-Iraq.[40] Vor dem Sommer 2018 war das IS-Mediennetz in Syrien und dem Irak in 23 regionale Medienbüros unterteilt, eins für Wilayat al-Raqqah, ein anderes für Wilayat Halab und so weiter.[41]

In unserem Archiv lassen sich 3.831 Einträge auf provinzielle Medieneinheiten zurückführen. Weitere 554 konnten einer zentralen Medieneinheit zugeordnet werden, vorwiegend der al-Furqan Foundation, dem AlHayat Media Center, der al-Itisam Foundation, al-Bayan Radio oder al-Naba. Bei den restlichen 1.905 Einträgen, zumeist von der Amaq News Agency produziert, war keine geografische Klassifizierung möglich, da im Archiv-Index kein Medienbüro mit spezifischem Standort angegeben war.

Wie Abb. 9 zeigt, stammen 42,6 % der Einträge, zu denen ein ortsspezifisches Medienbüro bekannt ist, aus Syrien. Weitere 39,6 % wurden von den IS-Medienbüros im Irak produziert und 7,9 % vom Medienbüro für Wilayat Sayna' in Ägypten. Die restlichen Einträge stammen aus Libyen, Jemen, Afghanistan, Nigeria, Russland, Algerien, Indonesien, Pakistan, Saudi-Arabien, Türkei, Tunesien,

39 Siehe Berger und Jonathon Morgan, „The ISIS Twitter census".
40 Abu Abdullah al-Masri, „The Isis papers: A masterplan for consolidating power", The Guardian, 7. Dezember 2015. Abgerufen: https://www.theguardian.com/world/2015/dec/07/islamic-state-document-masterplan-for-power.
41 BBC Monitoring, „Analysis: The Islamic State restructures its ‚provinces' a year on from 2017 defeats", 17. Oktober 2018. Abgerufen: https://monitoring.bbc.co.uk/product/c200bdcn.

Abb. 9: Inhalt nach Land



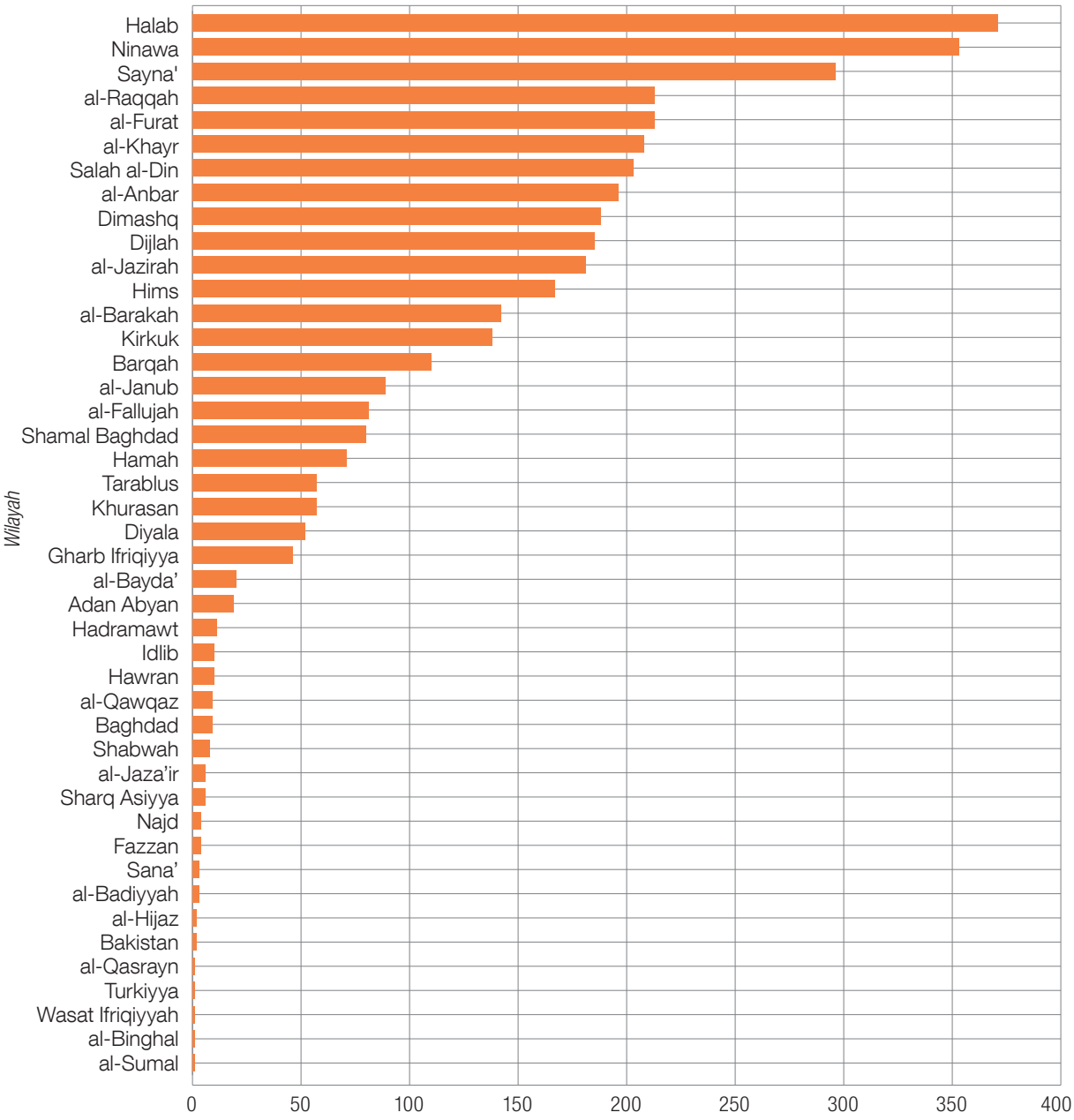
der Demokratischen Republik Kongo, Bangladesch und Somalia (in absteigender Reihenfolge genannt).

Abb. 10 stellt die Einträge nach *wilayah* statt Land dar. Daraus geht hervor, dass die meisten Inhalte von nur drei IS-Medieneinheiten produziert wurden: dem Wilayat Halab Medienbür in Syrien, dem Wilayat Naynawa Medienbüro im Irak und dem Wilayat Sayna' Medienbüro in Ägypten.

Das Vorherrschen von Inhalten aus Syrien und dem Irak steht im Einklang mit dem generellen geografischen Charakter des IS in den fraglichen Jahren.[42] Materialien vom Wilayat Sayna' Medienbüro sind jedoch im Archiv überrepräsentiert, was darauf hindeutet, dass ihr Ersteller entweder mehr Interesse an Materialien über IS-Aktivitäten in Ägypten hatte oder besser darauf zugreifen konnte. Dies sagt mehr über die Herkunft des Erstellers aus als über die Ursprünge des Materials selbst.

42 Siehe Winter, „Apocalypse, later".

Abb. 10: Inhalte von Medienbüros nach *Wilayah*



5 Schlussfolgerungen

Im Rahmen dieser Forschungsarbeit wurden automatisierte Textverarbeitungsmethoden eingesetzt, um eine Reihe von Analyse-Instrumenten zu entwickeln, mit denen sich große Mengen an IS-Propaganda analysieren und klassifizieren lassen. Obwohl wir uns ausschließlich mit IS-Material befasst haben, sind die hier entwickelten Ideen im Prinzip auf jede Form von gewalttätigem Extremismus anwendbar. Anhand eines Archivs von 6.290 Einträgen, das uns als Materialsammlung diente, haben wir versucht, die temporale, geografische und thematische Kategorisierung gewalttätiger extremistischer Inhalte zu automatisieren. Unser Tool hat den Datenbestand erfolgreich aufgeschlüsselt und damit die Identifizierung der potenziell schädlichsten Inhalte gegenüber weniger dringenden, aber dennoch höchst problematischen Inhalten ermöglicht. Durch die breitere Anwendung unseres Tools konnten zudem eine Reihe anderer, vielfach bestätigter Dynamiken identifiziert werden, sei es im Zusammenhang mit abnehmender Produktion, geografischen Merkmalen oder thematischer Vereinfachung.

Unser Hauptziel war die Entwicklung von Methoden, deren Anwendung auf ähnliche (auch sehr viel größere) Materialsammlungen den Prozess ihrer Aufschlüsselung sowie gegebenenfalls die Priorisierung ihrer Moderation und/oder Weiterleitung beschleunigen würde.

Dabei war unsere Absicht von Anfang an, ein Tool zu entwickeln, das Inhalte aufschlüsseln könnte, um so die Priorisierung für eine menschliche Prüfung zu unterstützen. Dieses Instrument ist natürlich nicht für die Verwendung in Isolation gedacht. Uns ist bewusst, dass Technologieunternehmen bereits über eine Reihe ausgereifter Systeme zur Identifizierung und Entfernung von Inhalten verfügen. In den allermeisten Fällen geschieht dies vollkommen automatisch. Aber auch menschliche Prüfer werden bei der Bewertung von problematischerem Material weiterhin eine wichtige Rolle spielen, und hier ist unser Tool in der Lage, bestehende Prozesse zu optimieren.

Wir sind überzeugt, dass Instrumente wie dieses angesichts der zunehmend vielfältigen Herausforderungen für Technologieunternehmen – darunter staatlich unterstützte Desinformationskampagnen, sogenanntes „coordinated inauthentic behavior“ (CIB) und die Verbreitung von Verschwörungstheorien – eher noch an Bedeutung gewinnen werden. Gewalttätige extremistische Organisationen konstruieren ihr Propagandamaterial aus vielen sich wiederholenden Elementen: Sie verwenden eine begrenzte Sammlung an Logos, die gleichen einleitenden Sequenzen und die gleichen Audiospuren. Bei der automatischen Erkennung sind solche Inhalte also leicht auszumachen. Wenn es allerdings darum geht, Offline-Schaden zu begrenzen und gleichzeitig die Redefreiheit zu bewahren, reicht die Erkennung gefährdender Inhalte allein nicht aus. Diese Materialien müssen ebenfalls sowohl in ihrem breiteren Kontext

als auch hinsichtlich der Motivation für ihre Herstellung untersucht und verstanden werden. Gegenwärtig sind dazu nur menschliche Moderatorinnen und Moderatoren in der Lage.

Die im Laufe dieser Untersuchung entwickelten Instrumente würden einiges dazu beitragen, einen solchen Prozess der Priorisierung für die menschliche Untersuchung zu straffen.

Die politische Landschaft

Dieser Abschnitt wurde von Armida van Rij und Vivienne Moxham-Hall, beide Research Associates am Policy Institute des King's College London, verfasst. Er bietet einen Überblick über den politischen Kontext des Berichtsthemas.

Einleitung

Der Missbrauch des Internets durch Terroristen und andere Personen mit extremistischen Ansichten ist im Laufe des letzten Jahrzehnts zu einem wachsenden Problem geworden. Plattformen von Social-Media- und anderen Technologieunternehmen werden benutzt, um terroristische Propaganda zu verbreiten, Mitglieder für terroristische Organisationen zu rekrutieren oder zu Gewalthandlungen anzustiften. Technologieunternehmen genauso wie politische Entscheidungsträger stehen vor der schwierigen Aufgabe, solche schädlichen Inhalte zu klassifizieren – oder, noch einfacher, zu bestimmen, was illegal ist und von der Plattform entfernt werden sollte. Dieser Prozess erfordert notwendigerweise Entscheidungen darüber, was als „extremistisch" oder „terroristisch" gilt, welche Ressourcen und Kapazitäten nötig sind, um all die neuen, stündlich hochgeladenen Inhalte zu moderieren, und wie man schädliche Inhalte sperren kann, ohne gleichzeitig die Rede-, Gedanken- und Diskussionsfreiheit zu gefährden. Aufgrund dieser Kombination von Faktoren ist der Kampf gegen terroristische Online-Inhalte und die Einstufung dessen, was zulässig ist und was nicht, zu einer enorm schwierigen Herausforderung geworden, die jedoch sehr reale Konsequenzen hat. Der Live-Stream eines Terroranschlags auf zwei Moscheen in Christchurch (Neuseeland) auf Facebook wurde bis zu seiner Entfernung mindestens 4.000 Mal angeschaut. Doch auch nach der Löschung wurde das Filmmaterial erneut hochgeladen und auf Social-Media-Websites, einschließlich Facebook, weiter verbreitet.[43]

Um besser zu verstehen, wie einzelne Länder mit terroristischem Online-Material umgehen, wie dieses sich am besten kategorisieren lässt und welche staatlichen Maßnahmen es gibt, um schädliche Inhalte zu entfernen, haben wir die politischen Rahmenbedingungen von neun Gesetzgebern hinsichtlich terroristischer Online-Inhalte untersucht. Diese neun Beispiele wurden nicht willkürlich ausgewählt, sondern erscheinen in unserer Liste, weil sie Mitglieder des unabhängigen Beratungsausschusses (Independent Advisory Committee, IAC) des Global Internet Forum to Counter Terrorism (GIFCT) sind. Der IAC selbst besteht aus 21 Mitgliedern, darunter Vertreter von sieben Regierungen, zwei internationalen Organisationen und 12 Organisationen der Zivilgesellschaft,

43 „Facebook: New Zealand attack video viewed 4,000 times", BBC News, 19. März 2019. Abgerufen: https://www.bbc.co.uk/news/business-47620519.

die insgesamt ein breites Spektrum an Fachkompetenzen einbringen. Die relevanten Gesetzgeber:

- Regierung von Kanada
- Regierung von Frankreich
- Regierung von Ghana
- Regierung von Japan
- Regierung von Neuseeland
- Regierung des Vereinigten Königreichs
- Regierung der Vereinigten Staaten
- Europäische Union
- Counter-Terrorism Committee Executive Directorate (CTED) des Sicherheitsrates der Vereinten Nationen

Dieser Bericht befasst sich jeweils mit den folgenden Fragen: (i) Wer sind die wichtigsten Stakeholder bei jedem dieser Gesetzgeber? (ii) Mit welchen Herausforderungen sind sie konfrontiert? (iii) Was sind die politischen Entwicklungen und die wichtigsten Gesetze in jeder Rechtsordnung? (iv) Was planen die Stakeholder für die Zukunft?

Hinsichtlich der ersten Frage ist klar, dass es eine Vielzahl von Anspruchsgruppen gibt, von Regierungsstellen und Internetdienstanbietern (ISP) bis hin zu Social-Media-Unternehmen, Organisationen der Zivilgesellschaft und der Öffentlichkeit. Wir versuchen in diesem Bericht nicht, alle diese Anspruchsgruppen im Zusammenhang mit einem bestimmten Land oder einer bestimmten Organisation anzuführen. Vielmehr konzentrieren wir uns ausschließlich auf die wichtigsten Stakeholder mit klaren Verantwortlichkeiten für die Bekämpfung extremistischer Online-Inhalte.

Vor der Bewertung der einzelnen Länder lassen sich eine Reihe gemeinsamer Herausforderungen feststellen. Eine Herausforderung, die insbesondere die westlichen Länder teilen, ist die Notwendigkeit, das Recht auf Redefreiheit mit dem Schutz der Bevölkerung zu vereinbaren. Regierungen wurden von Verteidigern der Redefreiheit mehrfach dafür kritisiert, die Entfernung extremistischer Online-Videos per Gesetz zu regeln, und gewarnt, dies sei potenziell eine Einschränkung der Redefreiheit und letztlich Zensur.

Dazu kommt, dass extremistische Inhalte zunehmend auf kleineren Plattformen erscheinen, diese aber nicht über die nötigen Kapazitäten verfügen, um illegale Inhalte zu überwachen, zu überprüfen und zu entfernen. Während die größeren Social-Media- und IT-Unternehmen über mehr Ressourcen verfügen und solche Herausforderungen besser bewältigen können, hat sich dies für kleinere Organisationen oft als schwierig erwiesen.

Kanada

Das Department of Public Safety and Emergency Preparedness (Public Safety Canada) hat die Entwicklung der Terrorist Content Analytics Platform unterstützt. Statistics Canada, das nationale statistische Amt, hat die Aufgabe, Terrorismus in Kanada zu verfolgen. Das Canada Centre for Community Engagement and Prevention of Violence leitet die Aktivitäten des Landes

zur Bekämpfung der Radikalisierung und arbeitet dabei mit der Regierung, der Zivilgesellschaft, Strafverfolgungsbehörden und internationalen Organisationen zusammen.

Im Jahr 2017 hat das Canada Centre die nationale Strategie zur Bekämpfung von Radikalisierung und Gewaltbereitschaft (National Strategy on Countering Radicalisation to Violence) vorgestellt, die auf frühzeitige Prävention, die Prävention unter gefährdeten Gruppen und die Befreiung von gewalttätigen Ideologien ausgerichtet ist.[44] Kanada setzt aktuell bereits Kategorisierungsmethoden ein, um terroristische Aktivitäten zu verfolgen. Statistics Canada verfolgt den Terrorismus auf der Grundlage von 13 Codes des UCR-Programms (Uniform Crime Reporting).[45] Online-Aktivitäten bilden in diesem System derzeit keine spezifische Kategorie. Eine Befragung von 13 städtischen Polizeibehörden in Kanada 2016 ergab allerdings, dass 40 % von ihnen die UCR-Codes nicht kannten und rund 50 % Probleme hatten, in speziellen Fällen einen passenden Code zu finden.[46] Dies unterstreicht die Problematik für Staaten und Strafverfolgungsbehörden, terroristische Aktivitäten zu definieren.

Der kanadische Premierminister Justin Trudeau schloss sich 2019 dem Christchurch Call to Action an, einer weltweiten Initiative zur Bekämpfung von terroristischen und gewalttätigen extremistischen Inhalten im Netz. Diesbezüglich hat Kanada die Organisation Tech Against Terrorism mit der Entwicklung der Terrorist Content Analytics Plattform beauftragt.[47] Dabei handelt es sich um eine zentralisierte Plattform mit dem ersten und größten Bestand an verifizierten terroristischen Inhalten. Sie automatisiert die Erkennung und Analyse solcher Materialien,[48] um kleinen Technologieunternehmen zu helfen, terroristische Inhalte zu identifizieren und effektiver dagegen vorzugehen, sowie die Entscheidungsfindung bei der Moderation von Inhalten zu unterstützen. Auf sekundärer Ebene wird sie zudem eine sichere akademische Forschung ermöglichen und so zu einem besseren Verständnis der Bedrohung beitragen, die von Terrorismus und extremistischen Inhalten ausgeht. Dies steht im Einklang mit der Verpflichtung gemäß dem Christchurch Call to Action, kleine Online-Plattformen bei der Entwicklung von Kapazitäten zur Bekämpfung terroristischer Online-Inhalte zu unterstützen.[49]

44 Siehe „Canada: Extremism & counter-extremism“, Counter-Extremism Project, 23. Juni 2020. Abgerufen: https://www.counterextremism.com/countries/canada.

45 Patrick McCaffery et al., „Classification and Collection of Terrorism Incident Data in Canada“, Perspectives on Terrorism, Band 10:5, 2016: 43. Abgerufen: https://www.universiteitleiden.nl/binaries/content/assets/customsites/perspectives-on-terrorism/2016/issue-5/505-classification-and-collection-of-terrorism-incident-data-in-canada-by-patrick-mccaffery-lindsy-richardson-jocelyn-j.-belanger.pdf.

46 Ibid.

47 „Press release: Tech Against Terrorism award grant by the Government of Canada to build Terrorist Content Analytics Platform“, Tech Against Terrorism, 27. Juni 2019. Abgerufen: https://www.techagainstterrorism.org/2019/06/27/press-release-tech-against-terrorism-awarded-grant-by-the-government-of-canada-to-build-terrorist-content-analytics-platform/.

48 „Update: Initial version of the Terrorist Content Analytics Platform to include far-right terrorist content“, Tech Against Terrorism, 2. Juli 2020. Abgerufen: https://www.techagainstterrorism.org/2020/07/02/update-initial-version-of-the-terrorist-content-analytics-platform-to-include-far-right-terrorist-content/.

49 Government of Canada, Public Safety Canada, „Government of Canada announces initiatives to address violent extremist and terrorist content online“, Pressemitteilung, 26. Juni 2019. Abgerufen: https://www.canada.ca/en/public-safety-canada/news/2019/06/government-of-canada-announces-initiatives-to-address-violent-extremist-and-terrorist-content-online.html.

Europäische Union

In der EU gibt es gegenwärtig mehrere Gesetze im Zusammenhang mit extremistischen Inhalten im Internet. Die Richtlinie (EU) 2017/541 zur Terrorismusbekämpfung zielt darauf ab, die Rechtsvorschriften der Mitgliedstaaten zur Kriminalisierung terroristischer Straftaten zu harmonisieren. Artikel 21 dieser Richtlinie verpflichtet die Mitgliedsstaaten zu Maßnahmen, die eine unverzügliche Entfernung von Online-Inhalten sicherstellen, wie z. B. Inhalte, die eine Aufforderung zum Terrorismus darstellen, Materialien zur Ausbildung und Anwerbung und andere terroristische Aktivitäten.[50] Die Art der Maßnahmen liegt dabei im Ermessen der einzelnen Staaten. Einige Mitgliedstaaten haben im Rahmen ihrer nationalen Gesetzgebung Mitteilungs- und Aktionsverfahren („Notice and Action", N&A) für Online-Plattformen erlassen,[51] darunter Frankreich, Deutschland und Spanien. Laut Artikel 14.3 der EU-Richtlinie über den elektronischen Geschäftsverkehr müssen Mitgliedstaaten Verfahren für die Entfernung einer Information oder die Sperrung des Zugangs zu ihr festlegen.[52] Die Auslegung dieses Artikels liegt jedoch im freien Ermessen jedes Staates, weshalb er EU-weit unterschiedlich angewendet wird. Außerdem einigte sich die Europäische Kommission 2016 mit Microsoft, Twitter, Facebook und YouTube auf einen freiwilligen Verhaltenskodex zur Bekämpfung illegaler Hassreden im Internet.[53]

Was die Klassifizierung und wirksame Bekämpfung extremistischer Online-Inhalte betrifft, steht die EU vor vielen Herausforderungen, ganz voran die Notwendigkeit, gemeinsame Standards und Verfahren für die (nach der Brexit-Übergangszeit) 27 nationalen Rechtsgebiete zu vereinbaren und zu implementieren. Diese für die EU so typische Problematik hat zu einer fragmentierten Landschaft unter den Mitgliedsstaaten geführt.

So gibt es z. B. keine Regeln auf EU-Ebene bezüglich Mitteilungs- und Aktionsverfahren für illegale Inhalte auf Online-Plattformen. Stattdessen haben nur einige wenige Mitgliedsstaaten ein Regelwerk für Mitteilungs- und Aktionsverfahren eingeführt. Manche (wie Frankreich, das Vereinigte Königreich und Ungarn) folgten dabei dem Geist der EU-Richtlinie über den elektronischen Geschäftsverkehr, während andere (wie Spanien) den Weg eines eigenständigen Rechtsinstruments wählten.[54] In einigen Ländern stehen Maßnahmen zur Messung, Sperrung, Filterung und Entfernung von Online-Inhalten nicht im Einklang mit Artikel 10 der Europäischen Menschenrechtskonvention, wonach Einschränkungen der Redefreiheit „rechtmäßig, legitim und notwendig" sein müssen.[55]

50 Europäische Kommission, „Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Verhinderung der Verbreitung terroristischer Online-Inhalte", 2018/0331 (COD), 2018(a): 3. Abgerufen: https://eur-lex.europa.eu/resource.html?uri=cellar:dc0b5b0f-b65f-11e8-99ee-01aa75ed71a1.0003.02/DOC_1&format=PDF.

51 Europäische Kommission, „Impact Assessment accompanying the document Proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online", 2018(b): 122. Abgerufen: https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-swd-408_en.pdf.

52 Ibid.: 10

53 Sicherheitsrat der Vereinten Nationen, Counter-Terrorism Committee Executive Directorate, „More support needed for smaller technology platforms to counter terrorist content", CTED Trends Alert, November 2018: 4. Abgerufen: https://www.un.org/sc/ctc/wp-content/uploads/2019/01/CTED-Trends-Alert-November-2018.pdf.

54 Europäische Kommission, 2018(b): 122.

55 Europarat, „Comparative study on blocking, filtering and take-down of illegal internet content", 20. Dezember 2015. Abgerufen: https://edoc.coe.int/en/internet/7289-pdf-comparative-study-on-blocking-filtering-and-take-down-of-illegal-internet-content-.html#.

Abb. 11: Bestehende Initiativen in den EU-Mitgliedstaaten zu Mitteilungs- und Aktionsverfahren[56]

MS	Gesetz	Gesetzgebung in Vorbereitung	Abgedeckte illegale Inhalte
BE	Keins	Notice & Action (N&A)	
DE	Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken (NetzDG)		Hassrede
DK	Keins		
ES	Königliches Dekret 1889/2011 über die Arbeitsweise der Kommission für geistiges Eigentum – abgeändert durch Gesetz 21/2014.		Urheberrechts-verletzungen
FI	Gesetz 2002/458 über die Bereitstellung von Diensten der Informationsgesellschaft		Urheberrechts-verletzungen
FR	Gesetz 2004-575 vom 21. Juni 2004 über das Vertrauen in die digitale Wirtschaft		Nur für eindeutig illegale Inhalte
HU	Gesetz CVIII von 2001 über bestimmte Aspekte des elektronischen Handels und über Dienste der Informationsgesellschaft		Verletzung von Rechten des geistigen Eigentums
IT	AGCOM-Verordnung zum Schutz des Urheberrechts in elektronischen Kommunikationsnetzen, 680/13/CONS, 12. Dezember 2013		Urheberrechts-verletzungen
LT	Verordnung über die Verweigerung des Zugangs zu Informationen, die auf illegale Weise erworben, erstellt, geändert oder verwendet wurden, genehmigt durch den Regierungsbeschluss Nr. 881 vom 22. August 2007		Horizontal
PL	Keins	Arbeitet potenziell an einer N&A-Initiative	
PT	Gesetzesdekret Nr. 7/2004 vom 7. Januar 2004, Lei do Comércio Electrónico, 7. Januar 2004		Außergerichtliche vorläufige Streitbeilegung
SE	Gesetz über die Verantwortung für elektronische Bulletin Boards		Urheberrechts-verletzungen, rassistische Inhalte
UK	Electronic Commerce Regulations S.I. 2002/2013		Horizontal – legt die Anforderungen hinsichtlich einer Meldung nieder

56 Abbildung entnommen aus: Europäische Kommission, „Impact Assessment accompanying the document Proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online“, 2018(b): 123. Abgerufen: https://ec.europa.eu/commission/sites/beta-political/files/soteu2018-preventing-terrorist-content-online-swd-408_en.pdf.

Im Jahr 2018 veröffentlichte die Europäische Kommission einen Gesetzesentwurf, nach dem extremistische Inhalte innerhalb einer Stunde nach dem Hochladen entfernt werden müssen. Zudem würde das Gesetz den Plattformen eine Sorgfaltspflicht gegenüber ihren Nutzern auferlegen[57] und Mitgliedsstaaten verpflichten sicherzustellen, dass ihre Behörden und Strafverfolgungsorgane über die erforderlichen Kapazitäten zur Bekämpfung terroristischer Online-Inhalte verfügen.[58] Dies erwies sich jedoch bei einigen Mitgliedsstaaten und europäischen Parlamentariern gleichermaßen als unpopulär. Man hielt die Definition von Terrorismus in diesem Gesetzesentwurf für zu weit gefasst und die Forderung nach Entfernung innerhalb einer Stunde für zu strikt, was potenziell zu einer Kultur der Zensur führen könnte.[59] Das Gesetz wurde im Europäischen Parlament abgeändert, um den Hauptanliegen seiner Kritiker Rechnung zu tragen.[60] Die Verhandlungen über die EU-Verordnung zur Verhinderung der Verbreitung terroristischer Online-Inhalte zwischen dem Rat, der Kommission und dem Parlament wurden aufgrund der Corona-Pandemie vorerst eingestellt. Im Juli 2020 kündigte die Europäische Kommission unverbindliche Leitlinien im Rahmen der ursprünglich 2018 verabschiedeten Richtlinie über audiovisuelle Mediendienste an, laut denen Online-Plattformen ihre Nutzer vor Hassreden und Minderjährige vor schädlichen Inhalten schützen müssen.[61] Darüber hinaus arbeitet die EU aktuell an einem Gesetz über digitale Dienste, das darauf abzielt, „das Online-Ökosystem in vielen Bereichen zu regulieren, darunter ... anstößige Inhalte“.[62]

Frankreich

Beamte der Nationalpolizei, die mit der Bekämpfung digitaler Verbrechen betraut sind, tragen die Verantwortung für die Durchsetzung von Gesetzen. Die Überprüfung, ob Plattformen, die zuvor aufgrund extremistischer Inhalte gesperrt wurden, weiterhin über diese Inhalte verfügen, ist Aufgabe des L’Office central de lutte contre la criminalité liée aux technologies de l’information et de la communication (Zentralbüro zur Bekämpfung von Kriminalität in Verbindung mit Informations -und Kommunikationstechnologie, im Grunde also die französische Zentrale für Cyberkriminalität).[63] Zudem werden illegale Inhalte dieser Zentrale gemeldet.

In Frankreich regelt Artikel 6 des Gesetzes Nr. 2004-575 vom 21. Juni 2004 die Haftung für Hosting-Plattformen. Demnach haften Unternehmen nur dann für Aktivitäten oder Informationen, die auf ihrer Plattform veröffentlicht werden, wenn sie Kenntnis von illegalen Inhalten haben und diese nicht entfernen. Sie sind jedoch verpflichtet, auf Meldungen über illegale Inhalte zu reagieren. Plattformen müssen über einen Berichtsmechanismus verfügen, der es jedem ermöglicht, Inhalte zu melden. Im Falle einer Meldung muss der illegale Inhalt innerhalb

57 Europäische Kommission, 2018(b): 3.
58 Ibid.: 4.
59 Faiza Patel, „EU ‚Terrorist Content‘ Proposal Sets Dire Example for Free Speech Online“, Just Security, 5. März 2019. Abgerufen: https://www.justsecurity.org/62857/eu-terrorist-content-proposal-sets-dire-free-speech-online/.
60 Al Cuddy, „EU struggles over law to tackle spread of terror online“, BBC News, 17. April 2019. Abgerufen: https://www.bbc.co.uk/news/world-europe-47962394.
61 „Facebook, YouTube, Twitter to face same EU rules on hateful content as broadcasters“, EURACTIF, 3. Juli 2020. Abgerufen: https://www.euractiv.com/section/digital/news/facebook-youtube-twitter-to-face-same-eu-rules-on-hateful-content-as-broadcasters/.
62 Samuel Stolton, „Platform clamp down on hate speech in run up to Digital Services Act“, EURACTIF, 23. Juni 2020. Abgerufen: https://www.euractiv.com/section/digital/news/platforms-clamp-down-on-hate-speech-in-run-up-to-digital-services-act/.
63 Europäische Kommission, 2018(b): 117.

von 24 Stunden gelöscht werden; andernfalls können die Behörden die elektronische Adresse des Inhalts an die Plattform melden, die den Zugang dann umgehend sperren muss.[64] Nach dem Terroranschlag auf das Redaktionsbüro von *Charlie Hebdo* verabschiedete Frankreich zudem im Februar 2015 ein neues Gesetz, das der Nationalpolizei die Befugnis gibt, Websites mit illegalen Inhalten ohne Gerichtsbeschluss zu schließen.[65] Ein weiterer Erlass, Nr. 2015-253 vom März 2015, bezieht sich spezifisch auf die direkte Provokation und/oder Anstiftung zum Terrorismus sowie die Verherrlichung von Terrorismus.[66]

Ein jüngst in Frankreich erlassenes Gesetz verpflichtet Technologieunternehmen, extremistische Inhalte innerhalb einer Stunde nach einer entsprechenden polizeilichen Anweisung zu entfernen. Andernfalls hat die französische Aufsichtsbehörde Conseil Supérieur de l'Audiovisuel die Befugnis, Bußgelder von bis zu 4 % des weltweiten Umsatzes zu verhängen. Für einige Plattformen ist das erneute Hochladen zuvor identifizierter und entfernter Inhalte ein erhebliches Problem. In Frankreich tendiert der Oberste Gerichtshof jedoch offenbar zu der Meinung, dass Plattformen nicht verpflichtet sind, das erneute Auftreten bereits entfernter Inhalte zu verhindern. Er signalisiert damit eine eher begrenzte Sorgfaltspflicht für Social-Media-Unternehmen.[67] Die Mitteilungs- und Aktionsverfahren sind auf illegale Inhalte begrenzt, was extremistisches Material beinhalten kann, aber nicht unbedingt alles erfasst.

Insbesondere äußern Gruppen zum Schutz der Meinungsfreiheit eine gewisse Besorgnis über die Folgen dieser Pflicht, Inhalte zu entfernen. Sie identifizieren zwei Problemfelder: Erstens ist es eine schwierige Aufgabe für kleinere Technologiefirmen, die nicht über die nötigen Ressourcen verfügen, rund um die Uhr große Mengen an Material zu überwachen.[68] Um der Gesetzgebung nachzukommen, müssen sie eventuell auf Zensur zurückgreifen, statt Geldstrafen zu riskieren. Zweitens wird befürchtet, dass dieses Gesetz missbraucht werden könnte, um politischen Aktivismus allgemein zu zensieren. Speziell dieser Punkt unterstreicht, wie schwierig es ist, extremistischen Inhalt zu definieren, da die Grenzen oft sehr fließend sind.[69]

Ghana

In Ghana gibt es eine spezielle Strafverfolgungsbehörde für Cyberkriminalität, mit der Europol, Interpol und ISPs kooperieren.[70] Aus öffentlich zugänglichen Informationen geht jedoch nicht hervor, welche Anstrengungen Ghana in der Terrorbekämpfung tatsächlich unternimmt. Ebenso unklar ist, ob es fortlaufende Bemühungen zur Bekämpfung terroristischer Online-Inhalte gibt, und wenn ja, welche Herausforderungen und Debatten sich daraus ergeben haben.

64 „Online terrorist propaganda: France and UK put internet giants in the cross-hairs", Jones Day, Juli 2017. Abgerufen: https://www.jonesday.com/en/insights/2017/07/online-terrorist-propaganda-france-and-uk-put-internet-giants-in-the-cross-hairs.
65 Europäische Kommission, 2018(b): 117.
66 Ibid.
67 Ibid.: 10.
68 „France gives online firms one hour to pull ‚terrorist' content", BBC News, 14. Mai 2020. Abgerufen: https://www.bbc.co.uk/news/technology-52664609.
69 Benedict Wilkinson und Armida van Rij, „An analysis of the Commission for Countering Extremism's call for evidence: Report 1 – Public understanding of extremism", Policy Institute, King's College London, 9. Dezember 2019.
70 Kristina Cole et al., „Cybersecurity in Africa: An Assessment" https://www.researchgate.net/profile/Seymour_Goodman/publication/267971678_Cybersecurity_in_Africa_An_Assessment/links/54e93dca0cf25ba91c7ef580/Cybersecurity-in-Africa-An-Assessment.pdf.

Japan

Das 2015 geschaffene National Centre of Incident Readiness and Strategy for Cybersecurity verpflichtet Infrastrukturunternehmen, wie z. B. für Versorgungsleistungen (Gas, Wasser, Elektrizität), Transportnetzwerke und Finanzinstitutionen, ihre Cybersicherheitsmaßnahmen proaktiv zu verstärken.

Im Jahr 2017 wurde in Japan ein umstrittenes neues Gesetz verabschiedet, das Verschwörungen zur Planung von Terrorismus und anderen schweren Vergehen bestraft. Es führt 277 verschiedene Delikte an, darunter das Kopieren von Musik und das Pilzesammeln in Naturschutzgebieten.[71] Gegner des Gesetzes kritisieren die Beschneidung von Freiheitsrechten und die unkonkrete Anwendung. Die Erklärung der Staats- und Regierungschefs auf dem G20-Gipfel in Osaka 2019 ruft Technologieunternehmen dazu auf, ihre Plattformen nicht für terroristische Zwecke missbrauchen zu lassen.[72]

Neuseeland

Die übergreifenden Maßnahmen gegen Terrorismus in Neuseeland werden zwischen mehreren Regierungsstellen, Kommunen und Organisationen des privaten Sektors koordiniert; sie stehen unter der Leitung des Cabinet External Relations and Security Committee sowie des Security and Intelligence Board. Die Gesamtstrategie ist in dem im Februar 2020 veröffentlichten Strategieplan zur Terrorismusbekämpfung beschrieben.[73]

Die Gruppe „Digital Safety“ am Department of Internal Affairs ist zuständig für die Überwachung von Online-Inhalten wie Filmen, Videos und anderen Publikationen, die potenziell darauf ausgerichtet sind, „Schaden zu verursachen“. Die Definition erfasst alle Online-Materialien, die „Dinge wie Sex, Horror, Verbrechen, Grausamkeit oder Gewalt derart beschreiben, darstellen, ausdrücken oder anderweitig behandeln, dass die Verfügbarkeit der Publikation dem öffentlichen Wohl schaden dürfte“.[74] Zur Identifizierung solchen Materials scheint man sich weitgehend auf bereits vorhandene externe Meldesysteme oder interne Inhalts-Sortiersysteme der Social-Media-Plattformen zu verlassen; die Regierung selbst führt offenbar keinen eigenen Screening-Prozess durch. Maßnahmen gegen identifizierte anstößige Inhalte sind in Neuseeland im Films, Videos, Publications and Classifications Act von 1993 niedergelegt, aktualisiert durch eine Änderung nach dem Terroranschlag von Christchurch im Jahr 2019, der gefilmt und online veröffentlicht wurde. Mitschnitte des Angriffs wurden von ISPs und Social-Media-Plattformen sofort freiwillig entfernt, als sie von der Gruppe „Digital Safety“ auf anstößiges Online-Material aufmerksam gemacht wurden. Verstöße gegen das Gesetz können

71 „Japan passes controversial anti-terror conspiracy law“, BBC News, 15. Juni 2017. Abgerufen: https://www.bbc.co.uk/news/world-asia-40283730; Robin Harding, „Japan passes pre-emptive anti-terrorism law“, The Financial Times, 15. Juni 2017. Abgerufen: https://www.ft.com/content/75130598-5181-11e7-bfb8-997009366969.

72 Regierung von Japan, „G20 Osaka Leaders’ statement on preventing exploitation of the internet for terrorism and violent extremism conducive to terrorism (VECT)“. Abgerufen: https://www.mofa.go.jp/policy/economy/g20_summit/osaka19/en/documents/final_g20_statement_on_preventing_terrorist_and_vect.html.

73 Government of New Zealand, Officials’ Committee for Domestic and External Security Coordination, Counter-Terrorism Coordination Committee, „Countering terrorism and violent extremism national strategy overview“, Februar 2020. Abgerufen: https://dpmc.govt.nz/sites/default/files/2020-02/2019-20%20CT%20Strategy-all-final.pdf.

74 Government of New Zealand, Department of Internal Affairs, „Objectionable and restricted material“. Abgerufen: https://www.dia.govt.nz/Censorship-Objectionable-and-Restricted-Material.

mit Haftstrafen von bis zu 14 Jahren und Geldbußen von bis zu 200.000 NZD geahndet werden.

Nach dem beispiellosen, live im Internet übertragenen Mord an 50 Menschen in Christchurch 2019 gründeten Neuseeland und Frankreich gemeinsam die Initiative „Christchurch Call to Action", um „terroristische und gewalttätige extremistische Inhalte im Internet zu eliminieren".[75] 48 Länder, darunter 31 neue Länder, kamen nachfolgend im Rahmen des Christchurch Call zusammen, um das Global Internet Forum to Counter Terrorism (GIFCT) neu zu strukturieren. Das GIFCT entwickelte 2019 ein gemeinsames Protokoll für die Krisenbewältigung, das von Google in Neuseeland getestet wurde, um eine koordinierte Handhabung der Online-Auswirkungen extremistischer Anschläge zu ermöglichen.[76]

Vereinigtes Königreich

Im Vereinigten Königreich ist die Terrorismusbekämpfung auf gesetzgeberischer und politischer Ebene eine Aufgabe des Home Office (Innenministerium). Dieses arbeitet dabei eng mit dem 2016 geschaffenen National Security Centre, Teil des Government Communications Headquarters, zusammen. Das Department for Digital, Culture, Media and Sport (DCMS) trägt die Verantwortung für die Aufrechterhaltung eines sicheren und offenen Internets.[77] Das Home Office arbeitet zudem eng mit Drittorganisationen zusammen, um spezifische Technologien zur Unterbindung und Bekämpfung von gewalttätigen extremistischen Online-Inhalten zu entwickeln. Eine Gruppe solcher Mitwirkenden sind Technologie- und KI-Unternehmen, die Instrumente gegen extremistische Inhalte entwickeln, wie z. B. Faculty (ehemals ASI Data Science). Zur zweiten Gruppe gehören Plattformen, die durch das Hochladen und Teilen illegaler Inhalte missbraucht werden, wie Facebook, Twitter und Microsoft. Darüber hinaus gibt es unabhängige Stellen, wie die Commission on Countering Extremism (Teil des Home Office) und das UK Council for Internet Safety, die gegen schädliche Auswirkungen des Internets einschließlich Hasskriminalität und Extremismus vorgehen.[78]

Das Home Office hat die Police Counter-Terrorism Internet Referral Unit (CTIRU) eingerichtet, an die jeder verdächtige Inhalte melden kann. Die CTIRU war bereits für die Entfernung von mehr als 300.000 einzelnen terroristischen Inhalten verantwortlich.[79] Im Februar 2018 gab die britische Regierung die Entwicklung einer neuen Technologie bekannt, die Videos präzise analysiert, um festzustellen, ob sie möglicherweise IS-Propaganda sind.[80] Diese Technologie nutzt maschinelles Lernen und war speziell für kleinere Technologieunternehmen bestimmt, die im Gegensatz zu größeren

75 Siehe https://www.christchurchcall.com/.
76 GIFCT, Joint Tech Innovation. Abgerufen: https://www.gifct.org/joint-tech-innovation/.
77 Houses of Parliament, Clare Lally und Rowena Bermingham, „Online extremism", UK Parliament POST, 6. Mai 2020: 3. Abgerufen: https://post.parliament.uk/research-briefings/post-pn-0622/.
78 HM Government, Home Office und Department for Digital, Culture, Media & Sport. „Online harms – White Paper", April 2019: 36. Abgerufen: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf.
79 HM Government, Home Office, „The United Kingdom's strategy for countering terrorism", Juni 2018: 35. Abgerufen: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/716907/140618_CCS207_CCS0218929798-1_CONTEST_3.0_WEB.pdf.
80 HM Government, Home Office, „New technology revealed to help fight terrorist content online", Pressemitteilung, 13. Februar 2018. Abgerufen: https://www.gov.uk/government/news/new-technology-revealed-to-help-fight-terrorist-content-online.

Unternehmen wie Facebook und YouTube nicht über ausreichende Kapazitäten zur Entwicklung eigener Tools verfügen. Die Klassifizierung des Inhalts erfolgt im Upload-Stream einer Plattform, d. h. ein Video wird abgelehnt, bevor es je die Plattform erreicht. Mit einem solchen präventiven Ansatz ist es möglich, Maßnahmen zu ergreifen, bevor das potenziell schädliche Video im Internet erscheint. Die damalige Innenministerin Amber Rudd schloss nicht aus, Unternehmen, denen andere wirksame Überwachungsressourcen fehlen, zukünftig per Gesetz zur Nutzung des Instruments zu verpflichten. Das Vereinigte Königreich hat zudem Gesetze erlassen, die mit einem Mitteilungs- und Aktionsverfahren gegen illegale Inhalte vorgehen.[81]

In dem vom Home Office und DCMS gemeinsam erstellten Whitepaper „Online Harms“ von 2019 schlug die Regierung die Einrichtung einer unabhängigen Aufsichtsstelle vor, die sich der Klärung der Verantwortung für die Regulierung von Online-Inhalten widmen würde.[82] Das Gesetz würde es zudem ermöglichen, Plattformen für die Verbreitung schädlicher Inhalte auf ihren Websites zur Rechenschaft zu ziehen. Gewissen Berichten zufolge dürfte sich die Verabschiedung des Gesetzes zur Einrichtung einer solchen Aufsichtsstelle im Parlament allerdings bis 2023 oder 2024 verzögern.[83] Stattdessen wurden der Medienaufsichtsbehörde Ofcom zusätzliche Befugnisse gegeben, um Technologieunternehmen für den Schutz der Bürger vor schädlichen, darunter extremistischen, Inhalten verantwortlich zu machen.[84] Außerdem kündigte das Vereinigte Königreich im September 2019 finanzielle Unterstützung für die Entwicklung von Technologien zur automatischen Erkennung von Videos an, die manipuliert wurden, um bestehende Erkennungsmethoden zu umgehen.[85] Ziel ist, dieses Tool kostenlos allen Technologieunternehmen zur Verfügung zu stellen.

USA

In den Vereinigten Staaten hat das Bureau of Counterterrorism innerhalb des State Department unter der Leitung eines Koordinators für Terrorismusbekämpfung die Aufgabe, „koordinierte Strategien und Ansätze zu entwickeln, um den Terrorismus im Ausland zu besiegen und die Zusammenarbeit internationaler Partner bei der Terrorismusbekämpfung [zu sichern]“.[86] Das Bureau of Counterterrorism arbeitet zudem mit Technologieunternehmen zusammen, um den Informationsaustausch zu verbessern.[87] Das Department of Homeland Security steht in enger Zusammenarbeit mit Bündnispartnern der USA sowie mit Organisationen wie Tech Against Terrorism, um speziell extremistische Online-Inhalte zu bekämpfen. Die USA engagieren sich auch in dem blockübergreifenden Global Counterterrorism Forum mit dem Ziel, einen langfristigen Ansatz zur Bekämpfung der vom Terrorismus ausgehenden Bedrohungen zu entwickeln.

81 Europäische Kommission, 2018(b): 20.
82 Houses of Parliament, 2020; Home Office und Department for Digital, Culture, Media & Sport, 2019.
83 „Online harms bills: warning over ‚unacceptable‘ delay“, BBC News, 29. Juni 2019. Abgerufen: https://www.bbc.co.uk/news/technology-53222665.
84 Ibid.
85 HM Government, Home Office, „UK to help develop new tech to stop sharing of terrorist content“, Pressemitteilung, 24. September 2019. Abgerufen: https://www.gov.uk/government/news/uk-to-help-develop-new-tech-to-stop-sharing-of-terrorist-content.
86 US Government, Department of State, Bureau of Counterterrorism. Abgerufen: https://www.state.gov/bureaus-offices/under-secretary-for-civilian-security-democracy-and-human-rights/bureau-of-counterterrorism/.
87 US Government, Department of State, „Country reports on terrorism 2019“, Bureau of Counterterrorism, Juni 2020. Abgerufen: https://www.state.gov/wp-content/uploads/2020/06/Country-Reports-on-Terrorism-2019-2.pdf.

Laut der US-Strategie zur Terrorismusbekämpfung von 2018 ist die „Bekämpfung des Online-Einflusses von Terroristen" ein Prioritätsbereich. Weiter heißt es, die USA seien bestrebt, in Zusammenarbeit mit Partnern die terroristische Nutzung des Online-Raums zur Rekrutierung, Geldbeschaffung und Radikalisierung von Personen zu bekämpfen.[88] Die USA verfolgen bei der Bekämpfung des Online-Einflusses von Terroristen drei Prioritäten: 1) Einbindung des privaten Sektors, 2) Unterstützung von Technologieunternehmen und Organisationen der Zivilgesellschaft bei der Verbreitung von Gegenbotschaften und 3) Schutz der Rechte unter dem 1. Verfassungszusatz.[89]

Der 1. Zusatzartikel zur Verfassung stellt politische Entscheidungsträger in den USA vor eine erhebliche Herausforderung, was die Regulierung extremistischer Online-Inhalte betrifft. Der Kontext ist in den USA etwas anders als in Europa, weil so viele Aussagen unter dem 1. Verfassungszusatz geschützt sind. Gewisse Inhalte gelten vielleicht in Frankreich oder dem Vereinigten Königreich als illegal, sind jedoch in den USA durchaus zulässig.[90] Diese Unterscheidung zwischen schädlichen und dennoch legalen Inhalten, anstößigen und dennoch legalen Inhalten sowie schlichtweg illegalen Inhalten macht es ISPs schwer, anstößige Inhalte zu entfernen. Paragraf 230 des Communications Decency Act von 1996 erlaubt Technologieunternehmen jedoch, anstößige Inhalte zu moderieren, auch wenn sie legal sind.[91] Im Laufe der Jahrzehnte haben Urteile des Supreme Court zu einer nuancierteren Handhabung des Konflikts zwischen Meinungsfreiheit einerseits und der Förderung von Gewalt andererseits geführt.[92]

Counter-Terrorism Committee Executive Directorate der Vereinten Nationen

Das Counter-Terrorism Committee Executive Directorate (CTED) der Vereinten Nationen wurde vom UN-Sicherheitsrat mit der Resolution 1535 (2004) eingerichtet, um als Expertengremium das Counter-Terrorism Committee (CTC) des Sicherheitsrats zu unterstützen.[93] Sein anfängliches Ziel bestand darin, die Implementierung von Resolutionen des Sicherheitsrats zur Terrorismusbekämpfung durch die UN-Mitgliedstaaten zu bewerten und diese Bemühungen im Wege eines Dialogs zu unterstützen. Das CTED steht in enger Zusammenarbeit mit dem Sicherheitsrat, den großen Technologieunternehmen im GIFCT sowie Organisationen der Zivilgesellschaft. Zudem hat es die Initiative „Tech Against Terrorism" gegründet, ein öffentlich-privates Partnerschaftsprojekt mit dem Auftrag, „den globalen Technologiesektor unter Wahrung

88 US Government, White House, „National strategy for counterterrorism of the United States of America", Oktober 2018: 22. Abgerufen: https://www.whitehouse.gov/wp-content/uploads/2018/10/NSCT.pdf.
89 US Government, Department of Homeland Security, „Strategic framework for countering terrorism and targeted violence", September 2019: 24. Abgerufen: https://www.dhs.gov/sites/default/files/publications/19_0920_plcy_strategic-framework-countering-terrorism-targeted-violence.pdf.
90 Daphne Keller et al., „Regulating Online Terrorist Content: A Discussion With Stanford CIS Experts About New EU Proposals", SLS Blogs, 25. April 2019. Abgerufen: https://law.stanford.edu/2019/04/25/regulating-online-terrorist-content-a-discussion-with-stanford-cis-directors-about-new-eu-proposals/.
91 Ibid.
92 Victoria L. Killion, „Terrorism, Violent Extremism, and the Internet: Free Speech Considerations", Congressional Research Service, 6. Mai 2019: i. Abgerufen: https://fas.org/sgp/crs/terror/R45713.pdf.
93 Naureen Chowdhury Fink, „Meeting the Challenge: A Guide to United Nations Counterterrorism Activities", International Peace Institute, 2012: 45. https://www.ipinst.org/wp-content/uploads/publications/ebook_guide_to_un_counterterrorism.pdf.

der Menschenrechte bei der Reaktion auf die terroristische Nutzung des Internets zu unterstützen“.[94]

Gegenwärtig gibt es mehrere UN-Resolutionen über den Missbrauch des Internets für terroristische Zwecke. Resolution 2129 (2013) des Sicherheitsrats verweist auf die sich weiterentwickelnde Verflechtung von Terrorismus und IKT sowie auf die Nutzung von Technologien wie dem Internet zur Begehung und Förderung von Terrorakten, indem sie die Anstiftung, Anwerbung, Geldbeschaffung oder Planung solcher terroristischer Handlungen ermöglichen.[95] Diese Resolution bekräftigt zudem das Mandat des CTED. Resolutionen 2354 (2017), 2395 (2017) und 2396 (2017) rufen Mitgliedstaaten auf, durch Kooperation untereinander sowie mit dem privaten Sektor und der Zivilgesellschaft zu verhindern, dass das Internet von Terrororganisationen missbraucht wird.[96] Darüber hinaus hat das CTED in Zusammenarbeit mit Südkorea, dem GIFCT und Tech Against Terrorism im Jahr 2017 seine Online-Plattform gestartet, um den Austausch von Informationen und Good Practice zu fördern.[97] Die Plattform soll kleineren Technologieunternehmen helfen, Online-Inhalte zu überwachen und gegen gewalttätigen Extremismus vorzugehen.[98]

In seinem Trends Alert vom November 2018 erkannte das CTED die Herausforderung an, vor der kleinere Plattformen und ISPs bei der Regulierung illegaler und terroristischer Inhalte stehen und zu deren Bewältigung die Plattform für den Wissensaustausch beitragen soll. Seine Publikationen beleuchten die sehr unterschiedlichen Ansätze von UN-Mitgliedstaaten und Technologieunternehmen. Es gibt eine Vielzahl an Vorgehensweisen –[99] von der Selbstregulierung bei Technologieunternehmen (speziell den größeren wie Facebook und Twitter) bis hin zu Ländern, die noch keine eigenen Vorkehrungen für Mitteilungs- und Aktionsverfahren getroffen haben (wie beispielsweise die Niederlande).

Schlussfolgerungen

In diesem Bericht bieten wir einen Überblick über die politischen Rahmenbedingungen von neun Gesetzgebern hinsichtlich terroristischer Online-Inhalte. Länder sind sich der Tatsache bewusst, dass der Missbrauch des Internets durch Extremisten und Terroristen bekämpft werden muss, sowohl auf staatlicher Ebene als auch im privaten Sektor. Alle von uns analysierten Länder haben bzw. erarbeiten Gesetze und Instrumente, um dieses Problem anzugehen. Und obwohl all diese Gesetzgeber (mit einer Ausnahme) die Bedrohung, die von der Nutzung des Online-Raums durch Extremisten und Terroristen ausgeht, öffentlich anzuerkennen scheinen, herrscht weniger Einigkeit darüber, welche Schritte ein Land von Technologieunternehmen verlangen sollte, um diese Bedrohung zu bekämpfen. Wir stellen auch fest, dass Länder im Kampf gegen extremistische Online-Inhalte gewisse Herausforderungen, Partner und Gesetze gemeinsam haben. Allerdings bestehen auch Unterschiede zwischen diesen neun Gesetzgebern

94 „March 2020 update“, Tech Against Terrorism, 3. März 2020. Abgerufen: https://www.techagainstterrorism.org/2020/04/03/march-2020-update/.
95 UN-Sicherheitsrat, Counter-Terrorism Committee, „Public-private efforts to address terrorist content online: A year of progress – what’s next?“, 14. September 2018. Abgerufen: https://www.un.org/sc/ctc/news/event/public-private-efforts-address-terrorist-content-online-year-progress-whats-next/.
96 Ibid.
97 Ibid.
98 UN-Sicherheitsrat, Counter-Terrorism Committee Executive Directorate, 2018: 2.
99 Siehe zum Beispiel: UN-Sicherheitsrat, Counter-Terrorism Committee Executive Directorate, 2018.

hinsichtlich der Mittel, die sie gegen extremistische Online-Inhalte einsetzen, und der Rücksicht auf die gesetzlich niedergelegte Rede- und Meinungsfreiheit. Multilaterale Organisationen und Institutionen wie die EU und das CTED müssen gelegentlich mit dem Problem kämpfen, dass ihr Mandat von diversen Ländern ausgeht (CTED) oder dass sich Mitgliedstaaten nicht über notwendige Maßnahmen einig sind (EU). Kooperationen mit globalen Initiativen wie dem GIFCT sowie die Verwendung von Instrumenten, die von Organisationen wie Tech Against Terrorism entwickelt wurden, sind offenbar populär.

Global Network
on Extremism & Technology

KONTAKTANGABEN

Im Falle von Fragen oder zur Anforderung weiterer Exemplare wenden Sie sich bitte an:

ICSR
King's College London
Strand
London WC2R 2LS
United Kingdom

T. **+44 20 7848 2098**
E. **mail@gnet-research.org**

Twitter: **@GNET_research**

Wie alle anderen GNET-Publikationen kann auch dieser Bericht kostenlos von der GNET-Website unter www.gnet-research.org heruntergeladen werden.

© GNET