



Neue Technologien und Friedenseinsätze: Hate Speech – Gift für die Mandatsumsetzung

Als „Gift“ und „globale Gefahr“ hat UN-Generalsekretär (GS) António Guterres die Hassrede (*Hate Speech*) bezeichnet, die auch im Umfeld von internationalen Friedenseinsätzen schädlich wirkt. Der in Zusammenhang mit der COVID-19-Pandemie zu beobachtende „*Tsunami of Hate*“ (UN) hat die Aufmerksamkeit für die digitale Verbreitung schädlicher Informationen und die Notwendigkeit wirksamer Gegenmaßnahmen noch einmal verstärkt. Auf Grundlage seiner ersten Strategie/Aktionsplan gegen Hassrede (Mai 2019) veröffentlichte Guterres im September 2020 einen detaillierten Leitfaden zur Umsetzung für Feldpräsenzen. Neue Technologien sollen dabei unterstützend zum Einsatz kommen. Dieses ZIF kompakt ist der zweite Teil einer Serie zu neuen Technologien und Friedenseinsätzen.

Hate Speech im Umfeld von Friedenseinsätzen: Status quo

Hate Speech schränkt Betroffene nicht nur in der Ausübung ihrer Grund- und Menschenrechte ein und verstärkt Spannungen und Konflikte zwischen Gruppen, sondern kann auch den Boden für physische Übergriffe gegen Angehörige der jeweiligen Gruppen bereiten. So waren zum Beispiel die Gräueltaten gegen die Rohingya in Myanmar 2017 auch die Folge von Hassrede und der Aufforderung zu Gewalt im Internet ([DW](#)). *Hate Speech* ist **kontextspezifisch** und häufig eng verbunden mit **Desinformation**¹ ([Disinformation-Amplified Hate Speech](#)). In Einsatzländern von Friedensmissionen ist das Phänomen zum Teil gut dokumentiert (u.a. Südsudan, Libyen, s. [Hate Speech Lexicons](#) von PeaceTech Lab).

Staaten schaffen zunehmend Gesetzesgrundlagen um gegen *Hate Speech* vorzugehen und weisen Social Media Plattformen Verantwortlichkeiten zu. In der Praxis führen Regelungen der Onlinekommunikation allerdings nicht nur zur Verteidigung, sondern auch zur Einschränkung des **Rechts auf freie Meinungsäußerung** ([DW](#)). 2020 wurden in 29 Ländern – darunter vier Länder mit Friedenseinsätzen – 155 Internet Shutdowns dokumentiert, in 19 Fällen davon mit dem offiziellen Grund die Verbreitung von *Hate Speech* zu bekämpfen ([Internet Shutdown Report](#), [Digital Rights Landscape Report](#)).

UN-Friedenseinsätze mit einem Mandat zum **Schutz der Zivilbevölkerung** (*Protection of Civilians, POC*) sind **grundsätzlich** damit beauftragt, im Rahmen ihrer Aktivitäten im Bereich „Schutz durch Dialog und Engagement“ der Verbreitung von *Hate Speech* mit Information und strategischer Kommunikation zu begegnen ([UN POC Policy](#), Nr.53). Einige Einsätze haben außerdem ein **explizites Mandat** in Bezug auf *Hate Speech*, darunter UNMISS in Südsudan ([Res. 2459/2019](#)) und MINUSCA in der Zentralafrikanischen Republik ([Res. 2499/2019](#)). Die Menschenrechtskomponente von MINUSCA zum Beispiel beobachtet *Hate Speech* in den sozialen Medien und arbeitet mit Facebook zusammen, um schädliche Inhalte zu entfernen. Ein weiteres Beispiel ist UNSMIL, die *Special Political Mission* in Libyen. Wie in der Zentralafrikanischen Republik haben auch in Libyen problematische Botschaften in sozialen Medien die Gräben zwischen den Gemeinschaften vertieft. Die Menschenrechtskomponente von UNSMIL hat mit libyschen Journalist:innen und Facebook zusammengearbeitet, um ein gemeinsames Verständnis von Hassrede zu entwickeln und verantwortungsvollen Journalismus zu stärken ([OHCHR](#), 10/2020).

Die neue Strategie der UN: 13 Key Commitments

Mit seiner UN-weiten Strategie/Aktionsplan gegen Hassrede ([Strategy and Plan of Action on Hate Speech](#)) will Generalsekretär António Guterres die Anstrengungen der Organisation intensivieren. Zum einen sollen die tieferen Ursachen und Treiber bekämpft werden, zum anderen will er die UN dazu befähigen, effektiv auf die Auswirkungen von Hassrede auf Gesellschaften zu reagieren. Die 13 *Key Commitments* (s. Kasten) adressieren *Hate Speech* als komplexes **gesellschaftliches Problem**. Eine wichtige Rolle bei der Umsetzung spielen UN-Friedenseinsätze.

¹ Desinformation ist eine Information, die falsch ist und erstellt wurde um einer Person, sozialen Gruppe, Organisation oder einem Staat Schaden zuzufügen – s. die Analyse von Wardle und Derakhshan (2017) zu [Information Disorder](#), die als weitere Phänomene Misinformation und Malinformation (darunter *Hate Speech*) beschreibt. Malinformation ist eine auf Wahrheit beruhende Information, die genutzt wird um einer Person, sozialen Gruppe, Organisation oder einem Staat Schaden zuzufügen.

Guterres' Begriffsklärung

Für den Terminus Hassrede gibt es in der internationalen Gemeinschaft **keine einheitliche rechtliche Definition**, was auch als ein wesentliches Problem ihrer Regulierung angesehen wird. Laut Art.20 Abs.2 des Internationalen Pakts über bürgerliche und politische Rechte (1966) haben Staaten jedoch „jedes Eintreten für nationalen, rassistischen oder religiösen Hass, durch das zu Diskriminierung, Feindseligkeit oder Gewalt aufgestachelt wird“, gesetzlich zu verbieten. Eine zentrale Rolle für die Definition spielt in der Fachliteratur der Aspekt der gruppenbezogenen Menschenfeindlichkeit. Der UN-Generalsekretär stellt seiner Strategie folgendes Verständnis voran:

*“Any kind of communication in speech, writing or behaviour, that **attacks or uses pejorative or discriminatory language** with reference to a person or a group on the basis of who they are, in other words, based on their religion, ethnicity, nationality, race, colour, descent, gender or other **identity factor**. This is often rooted in, and generates, intolerance and hatred, and in certain contexts can be demeaning and divisive.”*

Umsetzung der Strategie

Der Leitfaden des GS zur Implementierung seiner Strategie ([Detailed Guidance on Implementation for United Nations Field Presences](#)) bietet Friedenseinsätzen ein „Menü“ an Optionen, die kontextabhängig im Einklang mit dem jeweiligen Mandat priorisiert, in **landesbezogene Aktionspläne** gegossen und gemeinsam mit staatlichen und nicht-staatlichen Akteuren umgesetzt werden sollen. 107 Einzelmaßnahmen präzisieren die 13 Handlungsfelder der Strategie.

Zur Einordnung des **Schweregrades** von *Hate Speech* (*Top*, *Intermediate* und *Bottom Level*) sollen Friedenseinsätze die sechs Kriterien des sog. [Rabat Threshold Test](#) (2012) anwenden. Im Falle strafrechtlich relevanter Hassrede werden die direkte Adressierung von Plattformbetreibern und die Empfehlung von Maßnahmen (*Commitment 6*) sowie die Förderung einer unabhängigen, unparteiischen Untersuchung und von strategischen Gerichtsverfahren (*Commitment 3*) angeraten. Direkte Interventionsmöglichkeiten bieten sich zudem im Rahmen der strategischen Kommunikation (*Commitment 10*) – die Online-Kanäle eröffnen auch Friedenseinsätzen diverse Möglichkeiten, Botschaften schnell an ein breites Publikum zu verteilen.

Einsatz neuer Technologien

An verschiedenen weiteren Stellen empfiehlt der GS Friedenseinsätzen den Einsatz neuer Technologien im Engagement gegen *Hate Speech*, insbesondere bei der **Beobachtung und Analyse** (*Commitment 1*). Das Verständnis von Inhalten (*Message*), Akteuren und ihrer Motivation (*Messenger*) sowie den Verteilungswegen schädlicher Informationen (*Messaging*) ist zentral, um ihren Einfluss auf das Konfliktumfeld einschätzen und geeignete Gegenmaßnahmen definieren zu können. Technologien der Künstlichen Intelligenz (KI) für die [automatisierte Textanalyse](#) können Friedenseinsätze dabei unterstützen. Der Unterschied zwischen einer manuellen und einer automatisierten Social-Media-Analyse besteht in der Fähigkeit, Tausende von Kommentaren und Nutzern zu analysieren und ihre Reaktionen, Gesprächsthemen und Stimmungen zu quantifizieren.² Wichtig ist eine sorgfältige Adaption des Systems an das Umfeld; die unter Umständen mögliche Voreingenommenheit (*Bias*) KI-generierter Analysen ist eine weitere Herausforderung ([A/HRC/44/57](#)). Friedenseinsätze können eine Reihe *in-house* entwickelter Tools verwenden, unter anderem hat die UN-Initiative [Global Pulse](#) in den vergangenen Jahren Anwendungen erprobt (z.B. [Qatalog](#)). Diese Instrumente werden noch nicht systematisch eingesetzt, eine der Vorreiter:innen ist MINUSMA in Mali (*Radio Mining Hate Speech*).

² s. dazu auch den Beitrag [„Take back the analysis: Five things you can actually learn about a conflict context from social media“](#) auf dem ZIF-Blog *“100 Voices on Technology and Peace Operations“*.

Die 13 Key Commitments der Strategie

1. Beobachtung und Analyse von *Hate Speech*
2. Auseinandersetzung mit den tieferen Ursachen, Treibern und Verursachern von *Hate Speech*
3. Unterstützung der Opfer von *Hate Speech*
4. Zusammenführung relevanter Akteure
5. Zusammenarbeit mit neuen und traditionellen Medien
6. Einsatz von Technologie
7. Einsatz von Bildung als Mittel zur Bekämpfung von *Hate Speech*
8. Förderung friedlicher, inklusiver und gerechter Gesellschaften, um die tieferen Ursachen und Treiber von *Hate Speech* zu bekämpfen
9. Engagement in der Advocacy-Arbeit
10. Entwicklung von Richtlinien für die externe Kommunikation
11. Nutzung von Partnerschaften
12. Aufbau der Fähigkeiten von UN-Mitarbeiter:innen
13. Unterstützung der Mitgliedstaaten

Ein weiterer wichtiger Ansatzpunkt im digitalen Raum ist neben der [Governance](#) von schädlichen Inhalten (Moderation, Oversight und Regulation) die Förderung von Initiativen, die online **Gegenrede** (*Counter Speech*) einsetzen um die schädliche Wirkung von Hassrede zu minimieren und die Debattenkultur insgesamt zu verändern (*Commitment 2*).

Herausforderungen

Das Engagement gegen Hassrede ist komplex und erfordert Maßnahmen auf verschiedenen Ebenen von unterschiedlichen Akteuren. Der GS hat sehr konkrete Handlungsempfehlungen formuliert, die eine gute Grundlage für die Ausarbeitung landesbezogener Aktionspläne sind. Die Zusammenarbeit mit Partnern aus der Tech-Industrie, insbesondere bei der Beobachtung von *Hate Speech*, ist ein wichtiger Baustein – der von Guterres vorgeschlagene *Contact Point* als Verbindungsstelle unterstreicht dies. Für ein präventives Handeln der Friedenseinsätze ist die Fähigkeit entscheidend, die Mengen an Informationen im Netz effizient auszuwerten und Trends in die Konfliktanalyse einzubeziehen. Neue Technologien sollten hier systematisch eingesetzt werden, müssen aber immer von “menschlichen“ Analysekapazitäten begleitet sein, um die Ergebnisse im Kontext zu deuten.