ISD | Institute for Strategic Dialogue

gefördert durch:
Auswärtiges Amt

Online-Krisenprotokolle – Erweiterung des Politikinstrumentariums zum Schutz der Demokratie in Krisensituationen

Christian Schwieter

Über das Digital Policy Lab

Als zwischenstaatliche Arbeitsgruppe engagiert sich das Digital Policy Lab (DPL) dafür, politische Lösungen zur Verhinderung und Bekämpfung der Verbreitung von Desinformation, Hassrede sowie extremistischen und terroristischen Inhalten im Internet aufzuzeigen. Die Arbeitsgruppe besteht aus Vertreter:innen der zuständigen Ministerien und Aufsichtsbehörden ausgewählter liberal-demokratischer Länder. Die Arbeit des DPL zielt darauf ab, den regierungsübergreifenden Dialog zu fördern, politischen Entscheidungsträger:innen und Aufsichtsbehörden Zugang zu einschlägigem Fachwissen und Forschungsergebnissen zu verschaffen sowie eine internationale Arbeitsgemeinschaft zur Bewältigung der wichtigsten digitalpolitischen Herausforderungen aufzubauen. Wir danken dem Auswärtigen Amt für die Unterstützung des Projekts.

Über diesen Bericht

Im Rahmen des DPL organisierte das Institute for Strategic Dialogue (ISD) im Juli und September 2022 zwei Arbeitsgruppentreffen zum Thema Online-Krisenprotokolle. Die Arbeitsgruppe bestand aus Mitgliedern des DPL, die die nationalen Ministerien und Abteilungen sowie die Regulierungsbehörden aus Kanada, Neuseeland, der Slowakei, der Schweiz, dem Vereinigten Königreich und den USA repräsentierten. Ebenfalls teilgenommen haben Vertreter:innen der Wissenschaft und der Zivilgesellschaft. Auch wenn die Teilnehmenden an diesem Bericht mitgewirkt haben, spiegeln die darin geäußerten Ansichten nicht unbedingt die Ansichten aller Teilnehmenden oder der an diesem Projekt beteiligten Regierungen wider.

Herausgeberische Verantwortung:
Huberta von Voss (Executive Director, ISD Germany)

Autor

Christian Schwieter ist Project Manager bei ISD Germany und arbeitet in den Bereichen digitale Analyse und Digitalpolitik. Er erforscht die Auswirkungen von Online-Regulierung auf extremistische Akteur:innen und leitet das vom Bundesministerium der Justiz geförderte Forschungsprojekt »Radikalisierung in rechtsextremen Online-Subkulturen entgegentreten«. Vor seiner Tätigkeit für ISD forschte Christian Schwieter am Oxford Internet Institute und war Fachberater für den Digital-Untersuchungsausschuss des britischen Unterhauses. Christian hat einen MSc in Social Science of the Internet von der University of Oxford und einen BA vom Leiden University College in Den Haag, Niederlande. Er ist Co-Autor der Begleitpapiere des Digital Policy Lab 2020 sowie der ISD-Forschungsberichte »Stützpfeiler Telegram: Wie Rechtsextreme und Verschwörungsideolog:innen auf Telegram ihre Infrastruktur ausbauen«, »Um- und Abwege: Online-Strategien zur Verbreitung rechtsextremer Inhalte«, »Deutschland und der angebliche Klimalockdown« und »Auf Odysee: Die Rolle von Blockchain-Technologie für die Monetarisierung im rechtsextremen Onlinemilieu«.

Danksagungen

Wir bedanken uns bei allen den Teilnehmer:innen und Referent:innen für ihre wertvollen Beiträge als Teil der Arbeitsgruppe. Besonderer Dank gilt Denis Sparas und Julia van Best (Generaldirektion Kommunikationsnetze, Inhalte und Technologien, Europäische Kommission), Kristina Kirk (Department of the Prime Minister & Cabinet, Neuseeland), Dr. Erin Saltman (Global Internet Forum to Counter Terrorism, GIFCT), Antonis Samouris (Europol), Diego Naranjo (European Digital Rights, EDRi), Michael Meyer-Resende (Democracy Reporting International, DRI), und Iverna McGowan (Center for Democracy & Technology, CDT). Der Bericht wurde mit Unterstützung von Henry Tuck und Helena Schwertheim verfasst. Der Autor dankt ihnen herzlich für ihre wertvollen Hinweise.

ISD Institute for Strategic Dialogue

www.isdgermany.org

Inhaltsverzeichnis

Zusammenfassung

Das Gesetz über digitale Dienste der EU (eng. *Digital Services Act*, DSA) hat eine Vielzahl neuer Regulierungsinstrumente mit sich gebracht, die durch den Schutz der Grundrechte der Nutzer:innen und den Kampf gegen die Verbreitung illegaler und schädlicher Inhalte auf Online-Plattformen einen »sichereren digitalen Raum« schaffen sollen. Während es viele Diskussionen um die Haftung von Plattformen, algorithmische Audits und jährliche Risikobewertungen gegeben hat, wurde den in letzter Minute hinzugefügten DSA-Instrumenten – nämlich Mechanismen und Protokolle, die die DSA-Regulierungsbehörden in sogenannten »Krisenfällen« unterstützen sollen – weniger Aufmerksamkeit geschenkt. Um diese wenig erforschten, aber potenziell wichtigen neuen Regulierungsinstrumente zu beleuchten, hat ISD eine Arbeitsgruppe eingesetzt, um a) die Lehren aus bestehenden Online-Krisenprotokollen im Bereich der Bekämpfung von gewalttätigem Extremismus (engl. *countering violent extremism, CVE*) zu überprüfen und b) Empfehlungen zu sammeln, wie künftige Krisenprotokolle und Reaktionsmechanismen konzipiert und umgesetzt werden können, um die Grundrechte zu schützen, anstatt sie zu untergraben.

Die wichtigsten Erkenntnisse aus bestehenden Krisenprotokollen im CVE-Bereich:

1. Kleinere Plattformen und Dienste spielen eine Schlüsselrolle bei der Verbreitung gewalttätiger extremistischer Inhalte.
2. Beweise für die Strafverfolgung müssen gesichert werden, während gleichzeitig die zeitnahe Entfernung illegaler Inhalte zu gewährleisten ist.
3. Um Fähigkeiten zu verbessern, Kapazitäten aufzubauen und Doppelarbeit zu vermeiden, muss Fachwissen aus dem CVE-Bereich mit den wichtigsten Stakeholdergruppen, die mit der Gestaltung und Umsetzung der umfassenderen Krisenmechanismen des DSA beauftragt sind, geteilt werden.
4. Die Grundrechte müssen während und nach Krisenereignissen durch verfahrenstechnische Rechenschaftsmechanismen, einschließlich regelmäßiger Konsultationen mit Akteur:innen der Zivilgesellschaft, gewahrt werden.

Hauptanliegen und Empfehlungen zum Schutz der Grundrechte im Zusammenhang mit dem DSA-Krisenreaktionsmechanismus:

1. Die EU-Kommission muss klarstellen, ob sie ein Krisenereignis gemäß DSA als »Notstand« im Sinne der internationalen Menschenrechtsnormen (engl. *international human rights law,* IHRL) auslegen wird.
2. Bei Krisenereignissen ist ein ergänzender Schnellreaktionsmechanismus erforderlich, über den zivilgesellschaftliche Organisationen (ZGO) Plattformen direkt auf unzulässige Löschungen hinweisen können, möglicherweise koordiniert über das Europäische Gremium für digitale Dienste (im Folgenden das »Gremium«).
3. Delegierte Rechtsakte gemäß DSA müssen die Rolle der Kommission einschränken, indem sie dem Gremium mehr Befugnisse übertragen und seine Fähigkeit verbessern, als unabhängiges Aufsichtsorgan zu handeln.
4. Es müssen verlässliche und zeitnahe Bestimmungen für den Zugriff auf Daten für unabhängige Forscher:innen getroffen werden, um die Wirksamkeit und Verhältnismäßigkeit der Krisenreaktion zu bewerten, möglicherweise im Rahmen einer menschenrechtlichen Folgenabschätzung.

Einführung

In den letzten zwei Jahren kam es zu einer beispiellosen globalen Gesundheitskrise, gefolgt vom Ausbruch eines neuen Krieges in Europa, ausgelöst durch den russischen Angriff auf die Ukraine. Von der Verbreitung gesundheitsbezogener Fehlinformationen über die Radikalisierung der Anti-Lockdown-Bewegung bis hin zur strategischen Verbreitung von Kriegspropaganda – im Informationszeitalter tragen Aktivitäten auf Online-Plattformen wie Facebook, Instagram, YouTube, Twitter, TikTok oder Telegram zur weiteren Verschärfung dieser Krisen bei.

Während die politischen Entscheidungsträger:innen begonnen haben, ein Regelwerk zur Eindämmung der Macht von »Big Tech« zu entwickeln, haben einige Gesetzgeber:innen argumentiert, dass die Regulierungsbehörden in Krisenzeiten mit zusätzlichen Notfallbefugnissen ausgestattet werden sollten – auch bei der Regulierung von Online-Plattformen. Das EU-Gesetz über digitale Dienste (DSA) ist ein Beispiel für diesen Ansatz, da der Entwurf von Krisen wie der globalen Pandemie geprägt war und in den ersten Wochen nach der Invasion der Ukraine abgeschlossen wurde. Artikel 48 des DSA (Artikel 37 in früheren Fassungen) empfiehlt, »die Ausarbeitung von freiwilligen Krisenprotokollen zur Bewältigung von Krisensituationen einzuleiten, die strikt auf außergewöhnliche Umstände beschränkt sind, die die öffentliche Sicherheit oder Gesundheit beeinträchtigen«.

Darüber hinaus wurde im März 2022 ein zusätzlicher »Krisenreaktionsmechanismus« im Abschnitt »Risikobewertungen« (Artikel 36, 27a in früheren Versionen) eingeführt, der die Europäische Kommission ermächtigt, in Krisenzeiten zusätzliche Risikoanalysen von sehr großen Online-Plattformen zu verlangen. Die Covid-19-Pandemie und der Krieg in der Ukraine sind Beispiele für Fälle, in denen solche Krisenprotokolle und -mechanismen zur Anwendung kommen würden.

Befürworter:innen von Krisenprotokollen und Krisenreaktionsmechanismen argumentieren, dass diese Instrumente notwendig sind, um die Grundrechte in Krisensituationen, die die öffentliche Gesundheit oder die öffentlichen Sicherheit bedrohen, zu schützen. Beispielhaft für eine Maßnahme als Teil dieser Instrumente ist die Veröffentlichung verifizierter Informationen an prominenter Stelle auf der ersten Seite des Dienstes, um der Verbreitung von Fehlinformationen entgegenzuwirken. Kritiker:innen, auch aus der Zivilgesellschaft, argumentieren jedoch, dass die mangelnde Klarheit bei der Umsetzung dieser Notfallmaßnahmen die Rechtsstaatlichkeit bedroht, insbesondere bei Unklarheit darüber, welche Ereignisse als Krisenereignisse gelten und wer entscheidet, wann ein Notfall bzw. eine Krise ausgerufen wird. In einer öffentlichen Erklärung zum Krisenreaktionsmechanismus, die von den Menschenrechtsorganisationen Article 19 und European Digital Rights (EDRi) veröffentlicht wurde, warnten 23 zivilgesellschaftliche Organisationen davor, dass »Entscheidungen, die die Meinungsfreiheit und den Zugang zu Informationen berühren, insbesondere in Krisenzeiten, nicht rechtmäßig allein durch die Exekutivgewalt getroffen werden können«.[1]

Der DSA wurde zwar Oktober 2022 offiziell verabschiedet, aber die konkrete Ausgestaltung und Umsetzung der Krisenprotokolle und der damit verbundenen Mechanismen müssen noch weiter ausgearbeitet werden, um sicherzustellen, dass sie die Grundrechte schützen und nicht untergraben. Die Covid-19-Pandemie sowie der Krieg in der Ukraine und die entsprechenden Reaktionen der Plattformen bieten wichtige Anlässe, um zu erörtern, wie diese Mechanismen in Kraft gesetzt werden können. Dieser Bericht soll zu diesem Prozess beitragen, indem er Folgendes untersucht: a) die derzeitige Bedeutung von Online-Krisenreaktionsprotokollen und -mechanismen in Diskussionen über die Governance von Plattformen; b) welche Lehren aus bestehenden Protokollen und Mechanismen gezogen werden können; und c) wie Grundrechte in diesem Zusammenhang geschützt werden können. Im Sommer 2022 berief das ISD im Rahmen des Digital Policy Lab (DPL) eine Reihe von Arbeitsgruppen ein, an denen politische Entscheidungsträger:innen, Regulierungsbehörden und Expert:innen aus der Zivilgesellschaft teilnahmen, um diese Fragen zu erörtern.

Von Ad-hoc-Krisenreaktionen zu Protokollen und Mechanismen

Im Grunde besteht ein Krisenprotokoll aus einer Reihe von Regeln und Abläufen, die dazu dienen, die Auswirkungen eines unvorhergesehenen Notfalls zu mildern. Protokolle ermöglichen es den Stakeholdergruppen, Rollen, Verantwortlichkeiten und Vorgehensweisen vor einem Krisenereignis zu klären und die Notwendigkeit von Ad-hoc-Reaktionen zu minimieren. Eine Reihe von Stakeholdergruppen, darunter Regierungen, Unternehmen, Schulen und zivilgesellschaftliche Organisationen, können in verschiedenen Phasen an der Ausarbeitung und Umsetzung solcher Protokolle beteiligt sein, insbesondere wenn die voraussichtliche Krise ihre Sicherheit oder ihr Wohlergehen gefährden würde. Krisenprotokolle werden vor allem in Hochrisikosituationen, z. B. bei akuter Gefahr für Leib und Leben, angewandt.

In den vergangenen zwei Jahren wurden auf den durch die Covid-19-Pandemie ausgelösten globalen Gesundheitsnotstand verschiedene Ad-hoc-Reaktionen von unterschiedlichen Stakeholdergruppen, einschließlich Social-Media-Plattformen wie Facebook, YouTube oder Twitter, durchgeführt. Dazu gehört zum Beispiel die Priorisierung seriöser oder verifizierter Quellen zu Themen im Zusammenhang mit Covid-19, indem sie prominent auf der Startseite angezeigt werden, oder das Hinzufügen von Warnhinweisen zu Beiträgen mit falschen oder irreführenden Informationen über Impfstoffe.

In jüngster Zeit hat der russische Angriff auf die Ukraine auch zu einer Reihe von Sofortmaßnahmen geführt, die sowohl von Regierungen als auch von Plattformen ergriffen wurden. So hat die EU am 2. März 2022 die Ausstrahlung russischer Kriegspropaganda verboten, wovon die Konten der russischen Staatsmedien auf einer Vielzahl von Online-Plattformen wie YouTube, Facebook und Twitter betroffen sind.[2]

Das Gesetz über digitale Dienste der EU (DSA) – neue Instrumente zur Bewältigung von Krisensituationen

Die im Juni 2022 veröffentlichte vorläufige Einigung über das EU-Gesetz über digitale Dienste (siehe auch Berichtigung vom September) führte zwei zusätzliche Krisenreaktionsinstrumente ein: den verbindlichen Krisenreaktionsmechanismus (Artikel 36, zuvor 27a) sowie die freiwilligen Krisenprotokolle (Artikel 48, zuvor 37).[3, 4] Artikel 36 sieht einen verbindlichen Krisenreaktionsmechanismus vor, mit dem die Kommission in Krisenzeiten sehr große Online-Plattformen auffordern kann, zusätzliche, spezifische Ad-hoc-Risikobewertungen vorzunehmen und Risikominderungsmaßnahmen zu ergreifen. Artikel 48 beschreibt, wie die Kommission die Ausarbeitung von freiwilligen Krisenprotokollen zur Bewältigung von Krisensituationen unter außergewöhnlichen Umständen, die die öffentliche Sicherheit oder öffentliche Gesundheit bedrohen, einleiten kann. Sowohl die verbindlichen Bestimmungen von Artikel 36 als auch die freiwilligen Protokolle von Artikel 48 würden von der Kommission auf Empfehlung des Gremiums, das sich aus den nationalen Koordinatoren für digitale Dienste zusammensetzt, ausgelöst werden. Der verbindliche Artikel 36 wurde eingeführt, da die Mitgesetzgeber:innen befürchteten, dass die bereits im DSA vorgesehenen jährlichen Risikobewertungen nicht ausreichen würden, um diese Krisensituationen zu bewältigen. Die in Artikel 36 vorgesehenen zusätzlichen krisenbedingten Ad-hoc-Risikobewertungen würden daher die jährlichen Risikobewertungen ergänzen und stärken.

In den vergangenen Jahren wurden verschiedene Ad-hoc-Maßnahmen zur Bewältigung von Online-Krisen ergriffen. In jüngster Zeit hat die Diskussion über die Ausgestaltung krisenbedingter Sofortmaßnahmen angesichts des DSA neuen Auftrieb erhalten. Ziel des DSA ist es u.a., Protokolle und Mechanismen zu entwickeln, die die Vorgehensweisen und die Verantwortlichkeit der Stakeholdergruppen vor Eintreten eines Krisenereignises klären.

Der DSA ist nicht das erste Mal, dass Krisenprotokolle für Social-Media-Plattformen entwickelt und eingesetzt wurden. Innerhalb der CVE-Gemeinschaft wurden bereits seit mindestens 2017 Krisenprotokolle entwickelt, um die Verbreitung terroristischer Propaganda zu bekämpfen.[5] Im nächsten Abschnitt werden drei solcher Krisenprotokolle näher untersucht, um Erkenntnisse für die Gestaltung künftiger Online-Krisenprotokolle, die im Bereich der Plattformregulierung eingesetzt werden sollen, zu gewinnen.

Krisenprotokolle zur Bekämpfung von gewalttätigem Extremismus im Internet

Während die Bekämpfung der terroristischen Nutzung des Internets spätestens seit den Terroranschlägen von ISIS ganz oben auf der politischen Agenda stand, war der rechtsextreme Terroranschlag auf zwei Moscheen am 15. März 2019 in Christchurch, Neuseeland, der Hauptauslöser für die Entwicklung von Online-Krisenprotokollen.Der Anschlag zeigte sowohl die Schwachstellen vieler Online-Plattformen als auch die mangelnde Koordination zwischen Plattformen, Regierungen und Strafverfolgungsbehörden auf. Der Anschlag wurde live auf Facebook gestreamt und Aufnahmen in der Folge auf einer Vielzahl von Plattformen, darunter auch YouTube, hochgeladen.

Der Christchurch Call wurde von der neuseeländischen Premierministerin Jacinda Ardern und dem französischen Präsidenten Emmanuel Macron nach dem Anschlag initiiert.[6] Aus den internationalen Beratungen, an denen die CEO des ISD, Sasha Havlicek beteiligt war, entstanden 24 Selbstverpflichtungen zur Unterstützung von Regierungen und Unternehmen aus der Internetbranche mit dem Ziel, terroristische und gewalttätige extremistische Online-Inhalte zu beseitigen und gleichzeitig die Grundrechte und ein freies, offenes und sicheres Internet zu schützen. Eine besondere Verpflichtung besteht darin, gemeinsam an Verfahren zu arbeiten, die eine schnelle, koordinierte und wirksame Reaktion ermöglichen, wenn terroristische und gewalttätige Inhalte als Teil eines realen Angriffs verbreitet werden. Die Unterstützer:innen des Calls haben eine Reihe von ineinandergreifenden, freiwilligen Protokollen entwickelt, um dieser Verpflichtung nachzukommen.

Diese Protokolle a) definieren, was eine Krise ist und wann sie als beendet gilt, b) legen die Rollen der verschiedenen Stakeholdergruppen und die Maßnahmen fest, die sie in einer Krise ergreifen werden, und c) schaffen Kommunikationskanäle zwischen diesen Stakeholdergruppen, um schnelle und angemessene Reaktionen zu gewährleisten. Zu diesen Protokollen gehören das Christchurch Call Crisis Response Protocol, das Content Incident Protocol (CIP) des Global Internet Forum to Counter Terrorism (GIFCT), das EU-Krisenprotokoll EUCP sowie die Crisis Protocol Policy der Terrorist Content Analytics Platform (TCAP) (weitere Einzelheiten zu Umfang und Art dieser verschiedenen Protokolle siehe Tabelle 1). Ergänzend zu diesen internationalen Initiativen gibt es auch auf nationaler Ebene den neuseeländischen Online Crisis Response Process und den durch den australischen Online Safety Act eingeführten Prozess für Online-Krisenereignisse.[7]

Das Content Incident Protocol (CIP) des Global Internet Forum to Counter Terrorism (GIFCT)

Nach dem Christchurch Call wurde das bereits bestehende GIFCT, eine Initiative der Internetbranche, in eine unabhängige gemeinnützige Organisation umgewandelt und ihre Ressourcen und Mitgliederzahl erheblich erweitert.[8] Außerdem wurde das Content Incident Protocol (CIP) entwickelt. Der CIP-Prozess besteht aus drei Stufen: 1) Incident (Ereignis), 2) Content Incident (inhaltliches Ereignis) und 3) Content Incident Protocol (Protokoll zum inhaltlichen Ereignis). Auf der ersten Ebene gibt es wöchentliche Briefings, die mit den Plattformen geteilt werden, um sie für Ereignisse zu sensibilisieren, die nicht unbedingt sofortige Maßnahmen erfordern. Diese Briefings fließen auch in die verschiedenen Transparenzberichte ein, die von den GIFCT-Mitgliedern veröffentlicht werden. Die zweite Ebene, die Ebene des inhaltlichen Ereignisses, wird erreicht, wenn Inhalte eines Täters oder Komplizen eines gewalttätigen extremistischen Angriffs entdeckt werden, die von mutmaßlichen Täter:innen oder Kompliz:innen stammen.[9] Diese Inhalte werden dann gehasht und in eine Datenbank aufgenommen, auf die die GIFCT-Mitglieder zugreifen können, um sicherzustellen, dass die Inhalte auf ihren eigenen Plattformen erkannt und entfernt werden.[10] Die dritte Stufe, das Protokoll, wird nur im Falle einer live gestreamten und laufenden realen Bedrohung aktiviert.[11] Das vollständige Protokoll wurde nach der Schießerei in Halle (Deutschland) im Jahr 2019 und den Schießereien in Glendale (Arizona) und Buffalo (New York) im Jahr 2022 in den USA aktiviert.

Das EU-Krisenprotokoll (eng. *EU Crisis Protocol,* EUCP)

Das EU-Krisenprotokoll wurde 2019 vom EU-Internetforum als freiwilliger Handlungsrahmen angenommen, um einen schnellen und koordinierten grenzüberschreitenden Reaktionsmechanismus zur Bekämpfung der Verbreitung terroristischer Inhalte im Internet zu ermöglichen.[12] Wichtig ist, dass das Protokoll kein alltägliches Instrument zur Bekämpfung terroristischer Inhalte im Internet ist, da es spezifische und hochschwellige Kriterien in Verbindung mit der Art des Terroranschlags erfordert, bevor es aktiviert werden kann. Daher wurde das Protokoll seit seiner Verabschiedung nur einmal, nach dem islamistischen Mordanschlag auf den Lehrer Samuel Paty in einem Pariser Vorort im Oktober 2020, aktiviert. Im Gegensatz zum Content Incident Protocol (CIP) des Global Internet Forum to Counter Terrorism (GIFCT) soll das EU-Krisenprotokoll nicht nur die Verbreitung terroristischer Online-Inhalte einschränken, sondern auch die Ermittlungen der Strafverfolgungsbehörden aktiv unterstützen. Der Prozess ist nicht automatisiert, sondern hängt stark von der Koordination zwischen den nationalen Behörden, Plattform-Unternehmen und Europol ab. Vor diesem Hintergrund wird derzeit eine neue Plattform entwickelt, um die Koordination zwischen allen Stakeholdergruppen zu verbessern und die relevanten Informationen und Kommunikationskanäle zu konsolidieren. Der gesamte Prozess seitens Europol unterliegt der Aufsicht durch den Europäischen Datenschutzbeauftragten, um den Schutz des Grundrechts auf wirksamen Datenschutz zu gewährleisten.[13]

Geografie	Name	Art	Primär beteiligte Organisationen	Erfasste Plattformen	Art des erfassten Inhalts
Global	Content Incident Protocol (CIP) des Global Internet Forum to Counter Terrorism, (GIFCT)	Freiwillig, Initiative der Internetbranche	GIFCT	Airbnb, Amazon, Discord, Dropbox, Facebook, Instagram, JustPaste.it, LinkedIn, Mailchimp, Mega.nz Microsoft, Pinterest, Tumblr, Twitter, WhatsApp, WordPress.com & YouTube	terroristische und gewalttätige extremistische Inhalte
Global	Crisis Protocol Policy der Terrorist Content Analytics Platform (TCAP)	Freiwillig, hat zum Ziel, insbesondere kleinere Plattformen zu unterstützen	Tech Against Terrorism, finanziert von Public Safety Canada	Für alle Plattformen verfügbar	terroristische und gewalttätige extremistische Inhalte
EU	EU-Krisenprotokoll (eng. *EU Crisis Protocol*, EUCP)	Freiwillig, aber siehe Verordnung 2021/784 für verbindliche Verpflichtungen zur Entfernung und Speicherung von Inhalten	Europol	Meta, Twitter, Google, Microsoft, Dropbox, JustPaste.it, Dailymotion, Telegram, TikTok, Yubo, Discord, Vimeo & Snap	terroristische und gewalttätige extremistische Inhalte
Global	Christchurch Call Online Crisis Response Protocol	Freiwillige zwischenstaatliche Koordination	Christchurch Call-Regierungen und Unterstützer:innen aus Internetbranche, angeführt von den Regierungen Frankreichs und Neuseelands; auch Civil Society Advisory Network	Internetunternehmen, die den Call unterstützen, sind Amazon, Meta, Google, YouTube, Zoom, Dailymotion, Microsoft, Qwant, JV, LINE, Twitter, Roblox, Mega & Clubhouse	terroristische und gewalttätige extremistische Inhalte
Neuseeland	Online Crisis Response Process	Freiwillig, aber siehe Films, Videos, and Publications Classification Act 1993 (Aktualisierung 2019) für verbindliche Entfernungsverpflichtungen	Innenministerium	Alle, einschließlich Internetdienstanbieter	terroristische und gewalttätige extremistische Inhalte

Fortsetzung nächste Seite

Geografie	Name	Art	Primär beteiligte Organisationen	Erfasste Plattformen	Art des erfassten Inhalts
Australien	Abhorrent Violent Conduct Powers in einem Online-Krisenfall	Nicht gesetzlich verankert, aber siehe Online Safety Act 2021 & Criminal Code Amendment Act 2019	eSafety Commissioner	Alle, einschließlich Internetdienst-anbieter	Abscheuliche (eng. *abhorrent*) Inhalte zu gewalttätigem Verhalten
Vereinigtes Königreich	Crisis Response Protocol	Freiwillige Vereinbarung, unterstützt durch den Terrorism Act 2006, der einen Informationsaustausch zwischen der Regierung und der Internetbranche sowie Entfernungsanweisungen für Internetdienstanbieter vorsieht.	Innenministerium und Counter Terrorism Policing, einschließlich der Counter Terrorism Internet Referral Unit (CTIRU)	Alle	Online-Inhalte im Zusammenhang mit einem terroristischen Akt

Tabelle 1: Überblick über bestehende Online-Krisenprotokolle (nicht vollständig).

Lehren aus bestehenden Online-Krisenprotokollen

Nach Rücksprache mit den wichtigsten an der Gestaltung und Umsetzung der Krisenprotokolle beteiligten Stakeholdergruppen hat sich ein Konsens über vier Schlüsselbereiche herauskristallisiert, die den politischen Entscheidungsträger:innen bei der Entwicklung neuer Protokolle oder der Verbesserung bestehender Protokolle als Orientierung dienen sollten. Auch wenn die im DSA ermittelten Risiken bzw. mögliche Krisensituationen umfassender sind als die der bestehenden Online-Krisenprotokolle aus dem CVE-Bereich, sind die folgenden Lehren allumfänglich anwendbar.

1. **Kleinere Plattformen und »Alt-Tech«-Dienste spielen eine Schlüsselrolle bei der Verbreitung von gewalttätigen extremistischen Inhalten.** Besonders deutlich wurde dies nach dem rechtsextremen Terroranschlag in Buffalo (USA) im Mai 2022, als Teile des Livestreams von Millionen Menschen auf kleineren oder Nischenplattformen wie Streamable gesehen wurden.[14] Diese Arten von Plattformen sind häufig keine Mitglieder des GIFCT (oder Unterzeichner:innen des Christchurch Call) und fallen vermutlich nicht unter den Schwellenwert des DSA für sehr große Online-Plattformen. Dieses Problem hat zwei Dimensionen: Erstens fehlen vielen kleineren Plattformen sowohl die Kapazitäten als auch die Fähigkeiten, um schnell auf ein Krisenereignis zu reagieren, auch wenn sie diese Art von Inhalten nicht hosten wollen; zweitens gibt es Plattformen (unterschiedlicher Größe), die aus grundsätzlichen Erwägungen aktiv gegen solche Maßnahmen kämpfen und daher nicht bereit sind, gewalttätiges extremistisches Material zu entfernen (oft mit dem Argument der Meinungsfreiheit).[15] Krisenprotokolle müssen auf diese Tatsache reagieren, indem sie zum Beispiel sicherstellen, dass zumindest Verlinkungen zu solchen Inhalten auf Randplattformen von den größeren Plattformen entfernt werden.

2. **Beweise müssen gesichert und gleichzeitig die rechtzeitige Entfernung von Inhalten von den Plattformen gewährleistet werden.** Hier bedarf es Mechanismen mit denen gewährleistet wird, dass die Inhalte für rechtmäßige Zwecke wie Strafverfolgung, internationale Ermittlungen, Gerichtsverfahren, Journalismus und Forschung aufbewahrt werden. Jene Plattformen, die oft als erste mit dem Material in Be-

rührung kommen, müssen Verfahren einführen mit denen die Beweise aufbewahrt und sicher gespeichert werden, damit sie im Einklang mit den lokalen Gesetzen an die relevanten Stakeholdergruppen weitergegeben werden können, während gleichzeitig die öffentliche Verbreitung dieser Inhalte blockiert wird.[16] Im Zusammenhang mit den Gräueltaten während des syrischen Bürgerkriegs kritisierten zivilgesellschaftliche Organisationen bereits 2017, dass YouTube Videos auf seiner Plattform löschte, die »für eine mögliche Strafverfolgung von Kriegsverbrechen verwendet werden könnten.«[17] Alle an der Umsetzung eines Online-Krisenprotokolls beteiligten Stakeholdergruppen müssen sicherstellen, dass ihre Maßnahmen die Strafverfolgungsbehörden und Staatsanwaltschaften nicht daran hindern, die Täter:innen zur Rechenschaft zu ziehen.[18]

3. **Fachwissen aus dem CVE-Bereich muss mit den wichtigsten Stakeholdergruppen geteilt werden, die mit der Gestaltung und Umsetzung der umfassenderen Krisenmechanismen des DSA beauftragt sind, einschließlich der Koordinator:innen für digitale Dienste (eng. *Digital Services Coordinators*, DSCs), um Fähigkeiten zu verbessern, Kapazitäten aufzubauen und Doppelarbeit zu vermeiden.** Die DSCs werden wahrscheinlich aus unterschiedlichen Bereichen kommen (z. B. Medienregulierung) und benötigen daher möglicherweise zusätzliches Fachwissen über die (technische) Umsetzung von Krisenreaktionen. Es ist von entscheidender Bedeutung, dass die Krisenprotokolle und -mechanismen, die im Rahmen dieser Gesetzgebung eingeführt werden, darauf abzielen, die bestehenden Protokolle und Prozesse zu ergänzen und auf den Erfahrungen im CVE-Bereich aufbauen. Doppelte oder parallele Kommunikationskanäle und -prozesse können eine rasche Koordinierung während eines Krisenereignisses behindern.

4. **Die Grundrechte müssen während und nach Krisenereignissen geschützt werden.** Alle Arbeitsgruppenteilnehmer:innen waren sich einig, dass Krisenprotokolle und -mechanismen mit dem Ziel konzipiert und umgesetzt werden müssen, die Grundrechte zu schützen, und dass wirksame Schutzmaßnahmen, einschließlich des Zugangs zu Rechtsmitteln, im Prozess verankert sein sollen. Dazu gehört auch, dass von Anfang an ein breites Spektrum von Stakeholdergruppen, einschließlich der Zivilgesellschaft (wie im DSA vorgesehen), einbezogen wird und dass regelmäßige Überprüfungsprozesse und Folgenabschätzungen durchgeführt werden (siehe z. B. die GIFCT-Arbeitsgruppen).[19] Da die Notwendigkeit eines schnellen Handelns die Transparenz in Echtzeit behindern kann, ist die Transparenz in der Gestaltungs- und Bewertungsphase dieser Protokolle umso wichtiger.

Demokratische Garantien – Definitionen, Zuständigkeiten und Vorgehensweisen

Bei ihrem zweiten Treffen konzentrierte sich die Arbeitsgruppe auf den Schutz der Grundrechte im Zusammenhang mit den Online-Krisenprotokollen und -mechanismen, die im DSA vorgesehen sind. Wie in einer öffentlichen Erklärung von Article 19, EDRi, Access Now und 20 weiteren unterzeichnenden Organisationen beschrieben, war ein Hauptanliegen die mangelnde Transparenz während des Trialogprozesses, der zu den zusätzlichen Krisenprotokollen und -mechanismen im DSA führte. Einige Mitglieder der Arbeitsgruppe empfanden diese als nicht demokratisch legitimiert.[20] Neben diesen verfahrenstechnischen Bedenken wurden vor allem folgende Kernpunkte aufgeworfen:

1. **Die Kommission muss klarstellen, ob sie ein Krisenereignis gemäß DSA als »Notstand« im Sinne der internationalen Menschenrechtsnormen auslegen wird.** Menschenrechtsexpert:innen haben auf die unklare Formulierung in den derzeitigen DSA-Bestimmungen hingewiesen, aus der nicht hervorgeht, ob Artikel 36 einen Ausruf eines solchen Notstands darstellt. Folglich ist unklar, ob der Mechanismus eine potenzielle Abweichung von den Grundrechten im Sinne der Siracusa-Grundsätze vorsieht, oder ob der Artikel lediglich die Erwartungen an die Plattformen in Krisensituation formalisiert, ohne einen vollständigen Notstand auszurufen.[21] Mit anderen Worten: Ändert der Krisenzustand etwas am Schutz der Grundrechte? Die internationale Menschenrechtsnormen besagen, dass eine Ausnahme von den Grundrechten nur dann geltend gemacht werden kann, wenn eine »Bedrohung für das Leben einer Nation« vorliegt.[22] Ein solcher Notstand kann nur von den Mitgliedstaaten erklärt werden, nicht von der Kommission. In ähnlicher Weise sollte geklärt werden, wie der DSA-Krisenreaktionsmechanismus die bestehende Integrierte

Krisenreaktion des Rates der EU (eng. *Integrated Political Crisis Response,* IPCR) unter Leitung der Ratspräsidentschaft sowie das bestehende EU-Krisenprotokoll unter Leitung von Europol ergänzt.

2. **Bei Krisenereignissen ist ein ergänzender Schnellreaktionsmechanismus erforderlich, über den zivilgesellschaftliche Organisationen Plattformen direkt auf unzulässige Löschungen hinweisen können.** Insbesondere in Krisensituationen können die Anreize für Plattformen, die in den Anwendungsbereich des DSA fallen, risikoscheu zu handeln, zu einer unzulässigen Entfernung von Inhalten führen. Doch gerade in solchen Situationen sind die Meinungsfreiheit und der Zugang zu (korrekten) Informationen essenziell. Ein Krisenmechanismus, der die Grundrechte schützen soll, muss daher auch Möglichkeiten zur raschen Wiedergutmachung vorsehen, um das Risiko einer übermäßigen Sperrung (eng. *overblocking*), die zu Grundrechtsverletzungen führen könnte, einzudämmen.

3. **Delegierte Rechtsakte gemäß DSA müssen die Rolle der Kommission einschränken, indem sie dem Gremium für digitale Dienste mehr Befugnisse übertragen und seine Fähigkeit verbessern, als unabhängiges Aufsichtsorgan zu handeln.** In der verabschiedeten Fassung sieht der DSA nur wenige Kontrollmechanismen für die Exekutivgewalt bei Online-Krisenereignissen vor.[23] In den noch folgenden delegierten Rechtsakten muss die Rolle des Gremiums näher erläutert werden, und es muss sichergestellt werden, dass das Gremium über ausreichende Fachkenntnisse und Kapazitäten verfügt, um unabhängig von der Kommission als Aufsichtsorgan zu agieren.

4. **Es müssen verlässliche und zeitnahe Bestimmungen für den Zugriff auf Daten für unabhängige Forscher:innen getroffen werden, um die Wirksamkeit und Verhältnismäßigkeit der Krisenreaktion zu bewerten, möglicherweise im Rahmen einer menschenrechtlichen Folgenabschätzung (eng. *human rights impact assessment*, HRIA).** Anhand dieser Erkenntnisse kann beurteilt werden, ob Artikel 37 ein angemessenes Instrument zur Bewältigung von Krisenereignissen ist oder ob die in dem DSA enthaltenen allgemeinen Risikomanagementverfahren robust genug sind, außergewöhnliche Situationen zu bewältigen. Darüber hinaus können solche Daten Klarheit darüber verschaffen, welche Grundrechte von der Krise betroffen waren und wie die Reaktion diese Auswirkungen entweder abgemildert oder die Rechte untergraben hat. Anhaltspunkte für die Gestaltung solcher Datenzugriffsbestimmungen finden sich in dem kürzlich veröffentlichten Bericht der Arbeitsgruppe der Europäischen Beobachtungsstelle für digitale Medien (eng. *European Digital Media Observatory*, EDMO) über Datenzugriffsregeln für Plattformen und Forscher:innen.[24]

Fazit

Dieser Bericht hatte zum Ziel, die aktuellen Diskussionen über die Rolle von Online-Krisenreaktionsmechanismen zur Bekämpfung illegaler und schädlicher Inhalte zu kontextualisieren. Er stützte sich dabei auf die Debatte über Online-Krisenreaktionsprotokolle, die nach dem Terroranschlag in Christchurch 2019 eingeführt wurden. Während aus den bestehenden Online-Krisenprotokollen, die zur Bekämpfung von Terrorismus und gewalttätigem Extremismus im Internet entwickelt und umgesetzt wurden, viele Lehren gezogen werden können, besteht die Hauptaufgabe nun darin, diese Erkenntnisse in konkrete Maßnahmen umzusetzen, die über den CVE-Bereich hinaus angewandt werden können – wobei der Schutz der Grundrechte das übergeordnete Ziel bleiben muss. Die politischen Entscheidungsträger:innen müssen sicherstellen, dass bestehende zwischen- und innerstaatliche Koordinierungsmechanismen sowie Initiativen der Internetbranche berücksichtigt werden, damit Doppelarbeit, die wirksame Maßnahmen behindern könnte, vermieden wird. Es muss auch strenge Transparenzanforderungen geben, um eine unabhängige Prüfung der Wirksamkeit und Verhältnismäßigkeit der im Rahmen der Krisenreaktionsmechanismen ergriffenen Maßnahmen zu gewährleisten, wobei ein besonderer Schwerpunkt auf den Auswirkungen auf die freie Ausübung der Grundrechte im Internet liegen muss. Dazu gehört auch die öffentliche und legislative Kontrolle der Maßnahmen, die während eines Krisenereignisses sowohl von Plattformen als auch von Regierungen, Regulierungsbehörden und internationalen Organisationen ergriffen werden.

Endnoten

1 Article 19. (13. April 2022). *EU: Digital Services Act crisis response mechanism must honour human rights.* https://www.article19.org/resources/eu-digital-services-act-crisis-response-must-respect-human-rights/.

2 Eine ISD-Studie hat aufgezeigt, wie die Sperrung einschlägiger Nachrichtenkanäle umgangen wird. Kata Balint, Jordan Wildon, Francesca Arcostanzo und Kevin D. Reyes (6. Oktober 2022). *Effectiveness of the Sanctions on Russian State-Affiliated Media in the EU – An investigation into website traffic & possible circumvention methods.* https://www.isdglobal.org/isd-publications/effectiveness-of-the-sanctions-on-russian-state-affiliated-media-in-the-eu-an-investigation-into-website-traffic-possible-circumvention-methods-2/. Siehe auch Sara Bundtzen und Mauritius Dorn (5. April 2022). *Banning RT and Sputnik Across Europe: What Does it Hold for the Future of Platform Regulation?* https://www.isdglobal.org/digital_dispatches/banning-rt-and-sputnik-across-europe-what-does-it-hold-for-the-future-of-platform-regulation/.

3 Europäisches Parlament. (Oktober 2022). *REGULATION (EU) 2022/... OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of ... on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC.* https://www.europarl.europa.eu/meetdocs/2014_2019/plmrep/COMMITTEES/IMCO/DV/2022/06-15/DSA_2020_0361COD_EN.pdf.

4 Europäisches Parlament. (September 2022). *Berichtigung.* https://www.europarl.europa.eu/doceo/document/TA-9-2022-0269-FNL-COR01_EN.pdf.

5 2017 richtete die britische Regierung als Reaktion auf eine Reihe von Terroranschlägen im Inland eines der ersten Online-Krisenprotokolle dieser Art ein. Für weitere Informationen siehe Global Internet Forum to Counter Terrorism. (2022). *Introducing 2022 GIFCT working group outputs* (S. 10–11). https://gifct.org/wp-content/uploads/2022/07/GIFCT-22WG-CRP-MapGap-1.1.pdf/.

6 Christchurch Call. *Christchurch Call text.* https://www.christchurchcall.com/about/christchurch-call-text/.

7 eSafety Commissioner. (Dezember 2021). *Abhorrent violent conduct powers: regulatory guidance (eSC RG 5).* https://www.esafety.gov.au/sites/default/files/2022-03/Abhorrent%20Violent%20Conduct%20Powers%20Regulatory%20Guidance.pdf.

8 Global Internet Forum to Counter Terrorism. *Membership.* https://gifct.org/membership/.

9 Global Internet Forum to Counter Terrorism. *Content incident protocol.* https://gifct.org/content-incident-protocol/.

10 Global Internet Forum to Counter Terrorism. *Tech innovation.* https://gifct.org/tech-innovation/.

11 Global Internet Forum to Counter Terrorism. *Content incident protocol.* https://gifct.org/content-incident-protocol/.

12 Europäische Kommission. (Oktober 2019). *A Europe that protects EU Crisis Protocol: responding to terrorist content online.* https://home-affairs.ec.europa.eu/system/files/2019-10/20191007_agenda-security-factsheet-eu-crisis-protocol_en.pdf.

13 Europol. (November 2021). *Data protection and transparency.* https://www.europol.europa.eu/about-europol/data-protection-transparency.

14 Kellen Browning und Ryan Mac. (Mai 2022). *After Buffalo shooting video spreads, social platforms face questions.* New York Times https://www.nytimes.com/2022/05/15/business/buffalo-shooting-social-media.html.

15 Siehe Tech Against Terrorism. https://www.techagainstterrorism.org/ für Ressourcen und Schulungsangebote für kleinere Internetunternehmen.

16 Die Mitglieder der Arbeitsgruppe räumten ein, dass ein solcher Mechanismus in autoritären Kontexten missbraucht werden könnte. Daher sollten diese Maßnahmen durch ausreichende demokratische Garantien ergänzt werden. Siehe Global Internet Forum to Counter Terrorism. (2022). *Introducing 2022 GIFCT working group outputs* (p15). https://gifct.org/working-groups/.

17 Malachy Browne. (August 2017). *YouTube removes videos showing atrocities in Syria.* New York Times. https://www.nytimes.com/2017/08/22/world/middleeast/syria-youtube-videos-isis.html.

18 Business for Social Responsibility. (Juli 2021). *Human rights impact assessment: Global Internet Forum to Counter Terrorism.* https://www.bsr.org/en/our-insights/report-view/human-rights-impact-assessment-global-internet-forum-to-counter-terrorism. Darin wird darauf hingewiesen, wie wichtig all diese Verwendungsmöglichkeiten von Inhalten sind, um sicherzustellen, dass Opfer von Menschenrechtsverletzungen Zugang zu wirksamen Rechtsmitteln haben.

19 Global Internet Forum to Counter Terrorism. *Working groups 2022.* https://gifct.org/working-groups/.

20 Article 19. (13. April 2022). *EU: Digital Services Act crisis response mechanism must honour human rights.* https://www.article19.org/resources/eu-digital-services-act-crisis-response-must-respect-human-rights/.

21 Internationale Juristenkommission. (1. Juli 1984). *Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights.* https://www.icj.org/siracusa-principles-on-the-limitation-and-derogation-provisions-in-the-international-covenant-on-civil-and-political-rights/.

22 Es wird darauf hingewiesen, dass einige Mitglieder der Arbeitsgruppe Bedenken geäußert haben, dass die IHRL-Definition eines »Notfalls« weiter gefasst und weniger klar definiert ist als die im DSA enthaltene Definition einer Krise.

23 Die derzeitige Kontrollmechanismen in Bezug auf die Befugnisse der Kommission für den verbindlichen Krisenreaktionsmechanismus (Artikel 36) besagen, dass die Kommission auf Empfehlung des Gremiums handeln muss, wenn sie krisenspezifische Maßnahmen von sehr großen Online-Plattformen (engl. *very large online platforms*, VLOP) und sehr großen Onlinesuchmaschinen (engl. *very large online search engines*, VLOSE) fordert. Darüber hinaus muss die Kommission sicherstellen, dass VLOP- und VLOSE-Maßnahmen »erforderlich, gerechtfertigt und verhältnismäßig« sind, die Grundrechte respektieren und auf eine Dauer von höchstens drei Monate begrenzt sind. Außerdem muss die Kommission ihren Beschluss »öffentlich zugänglich« machen und »das Gremium von dem Beschluss in Kenntnis setzen«. Nach der Einführung von VLOP- bzw. VLOSE-Maßnahmen muss die Kommission dem Gremium im Rahmen ihrer VLOP- und VLOSE-Überwachung »mindestens monatlich« Bericht erstatten. Die Kommission kann auch ihren ursprünglichen Beschluss ändern (z. B. indem sie den Beschluss widerruft oder den Krisenreaktionszeitraum um weitere drei Monate verlängert), wiederum nur auf Empfehlung des Germiums. Schließlich muss die Kommission dem Europäischen Parlament und dem Rat »jährlich, in jedem Fall jedoch drei Monate nach Ende der Krise« über die Anwendung spezifischer VLOP- und VLOSE-Krisenmaßnahmen Bericht erstatten. Kontrollmechanismen in Bezug auf die Befugnisse der Kommission bei freiwilligen Krisenprotokollen (Artikel 48) sehen vor, dass das Gremium der Kommission empfehlen kann, »die Ausarbeitung von freiwilligen Krisenprotokollen [...] einzuleiten« und dass die Kommission »gegebenenfalls die Behörden der Mitgliedstaaten [...] und auch die Einrichtungen und sonstigen Stellen der Union einbeziehen« muss. Darüber hinaus kann die Kommission »gegebenenfalls auch Organisationen der Zivilgesellschaft [...] einbeziehen«.

24 European Digital Media Observatory. (Mai 2022). *Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access*. https://edmo.eu/wp-content/uploads/2022/02/Report-of-the-European-Digital-Media-Observatorys-Working-Group-on-Platform-to-Researcher-Data-Access-2022.pdf.

ISD | Institute for Strategic Dialogue

Amman | Berlin | London | Paris | Washington DC

www.isdgermany.org

gefördert durch:
Auswärtiges Amt