ISD | Powering solutions to extremism and polarisation

Erforschung des sich im Wandel begriffenen Online-Ökosystems: Hindernisse, Methoden und zukünftige Herausforderungen

Jakob Guhl, Oliver Marsh & Henry Tuck

Autoren
Jakob Guhl ist Senior Research Manager beim ISD, wo er in der Digital Research Unit und für das ISD Germany arbeitet. Seine Forschungsschwerpunkte sind Rechtsextremismus, islamistischer Extremismus, Hassrede, Desinformation und Verschwörungsideologien. Als ISD-Experte wurde Jakob mehrfach eingeladen, seine Forschungen dem Bundesministerium der Justiz vorzustellen und dem Bundesministerium des Innern und für Heimat seine Handlungsempfehlungen zur Prävention gegen Rechtsextremismus auszusprechen.

Oliver Marsh ist Gründer des Beratungsunternehmens The Data Skills Consultancy, das die Arbeit an der Schnittstelle von Datenkompetenz und Soft Skills unterstützt. Zuvor war er als Regierungsbeamter am Aufbau der Rapid Response Unit der britischen Regierung und im Ministerium für Digitales, Kultur, Medien und Sport an der Schaffung von Datenzuverlässigkeit nach dem Brexit beteiligt. Er ist Fellow der Denkfabrik Demos, Policy Fellow der Royal Academy of Engineering und Honorary Research Associate der Abteilung für Wissenschaft und Technologie an der University College London.

Henry Tuck ist Head of Digital Policy beim ISD, wo er die Arbeit zur digitalen Regulierung und zu den Maßnahmen der Technologieunternehmen gegen Terrorismus, Extremismus, Hass sowie Des- und Misinformation im Internet leitet. Henry betreut das Projekt Digital Policy Lab (DPL) des ISD und führt Beratungen zu wichtigen Vorschlägen im Bereich der digitalen Regulierung in Europa und den Five-Eyes-Ländern durch. Er arbeitet mit der Digital Analysis Unit des ISD zusammen, um Forschungsergebnisse in umsetzbare Empfehlungen für die digitale Politik zu übersetzen.

Überblick
In diesem Bericht werden die Ergebnisse der ersten Scoping-Phase eines von Omidyar Network geförderten und vom Institute for Strategic Dialogue (ISD) und CASM Technology ins Leben gerufenen Projekts vorgestellt. Ziel des Projekts ist es, Online-Räume jenseits der Mainstream-Plattformen zu identifizieren, die zunehmend von Individuen und Gemeinschaften zur Verbreitung von Extremismus, Hass und Desinformation genutzt werden, die sich von Mainstream-Plattformen entfernen. Der Bericht skizziert die wichtigsten Hindernisse, die diese Plattformen bei der Erforschung und Bekämpfung schädlicher Inhalte und Verhaltensweisen aufweisen, und identifiziert bestehende Forschungsmethoden und Tools, um diese Hindernisse zu bewältigen. Schließlich stellt er mögliche Zukunftsszenarien für das im Wandel begriffene Online-Ökosystem vor und gibt eine Reihe von Empfehlungen für politische Entscheidungsträger:innen, Plattformen und die Forschungsgemeinschaft.

Danksagungen
Dieser Bericht wäre ohne die finanzielle Unterstützung durch das Omidyar Network nicht möglich gewesen. Wir möchten uns bei Wafa Ben-Hassine, Anamitra Deb und Emma Leiken für ihre Vision, ihre kontinuierliche Unterstützung und ihr aufschlussreiches Feedback bedanken.

Die Autoren möchten auch dem gesamten Projektteam für seine Beiträge danken, die diesen Bericht möglich gemacht haben: Francesca Visser, Jacob Davey, Lea Gerster, Daniel Maki, David Leenstra und Francesca Arcostanzo vom ISD sowie Nestor Prieto Chavana und Carl Miller von CASM.

Schließlich bedanken wir uns auch bei Eduardo Ustaran und Nick Westbrook von der Wirtschaftskanzlei Hogan Lovells für ihre wertvolle Zeit und Unterstützung beim Verständnis der in diesem Bericht behandelten rechtlichen Herausforderungen.

ISD | Powering solutions to extremism and polarisation

Amman | Berlin | London | Paris | Washington DC

www.isdglobal.org

Contents

Glossar

Mit dem Begriff **Alt-Tech** bezeichnet man Social-Media-Plattformen, die von Gruppen und Einzelpersonen genutzt werden, die der Meinung sind, dass ihre politischen Ansichten auf den großen Social-Media-Plattformen nicht erwünscht sind. Dazu gehören Plattformen, die zur Förderung bestimmter politischer Ziele eingerichtet wurden; libertäre Plattformen, die ein breites Spektrum politischer Positionen tolerieren, darunter auch hasserfüllte und extremistische Meinungen; sowie Plattformen, die ursprünglich für ganz andere, unpolitische Zwecke wie z. B. Gaming bestimmt waren.

Eine Programmierschnittstelle, kurz **API** (vom Englischen „application programming interface"), ist ein Programmteil, der es zwei Anwendungen ermöglicht, miteinander zu kommunizieren. Es gibt eine Vielzahl von Verwendungsmöglichkeiten für APIs. Im Zusammenhang mit diesem Bericht ermöglichen APIs Forscher:innen den Zugriff auf bestimmte Daten von einigen Online-Plattformen durch an die Plattformen gestellte Anfragen. Als Schnittstelle bieten APIs auch eine zusätzliche Sicherheitsebene, indem sie keinen direkten Zugriff auf die Daten zulassen und den Umfang und die Häufigkeit der Anfragen protokollieren, verwalten und kontrollieren.

Das ISD definiert **Desinformation** als falsche oder irreführende Informationen, die mit der Absicht verbreitet werden, zu täuschen, wirtschaftliche und/oder politische Vorteile zu erzielen und Schaden zu verursachen. Wenn wir uns Informationen beziehen, die unabsichtlich verbreitet werden, verwenden wir den Begriff **Misinformation**.

Extremismus definiert das ISD als die Befürwortung einer Weltanschauung, die die Überlegenheit und Dominanz einer identitätsbasierten Eigengruppe über alle Fremdgruppen propagiert. Extremismus beinhaltet somit eine Dehumanisierung der „Anderen", die mit Pluralismus und universellen Menschenrechten unvereinbar ist.

Wir definieren **fragmentierte Plattformen** als solche, bei denen auf Online-Inhalte theoretisch ohne technologische oder ethische Hindernisse zugegriffen werden kann, die aber dennoch nicht schnell oder systematisch durchsucht werden können, z. B. über eine API. Relevante Inhalte müssen daher manuell aus einer großen Menge anderer Inhalte herausgesucht werden.

Unter **Hass** versteht das ISD Überzeugungen oder Praktiken, die eine ganze Gruppe von Menschen aufgrund geschützter Merkmale wie ethnische Zugehörigkeit, Religion, Geschlecht, sexuelle Orientierung oder Behinderung angreifen, verleumden, delegitimieren oder ausgrenzen. Hassakteur:innen sind Einzelpersonen, Gruppen oder Gemeinschaften, die sich aktiv und offen an den oben genannten Aktivitäten beteiligen, ebenso wie diejenigen, die implizit Menschengruppen angreifen, z. B. durch die Verwendung von Verschwörungsideologien und Desinformation. Hasserfüllte Aktivitäten laufen dem Pluralismus und der universellen Anwendung der Menschenrechte zuwider.

Offene Plattformen sind Social-Media-Plattformen, auf denen Inhalte für allgemeine Nutzer:innen ohne weitere Überprüfung sichtbar und oft über Suchmaschinen zugänglich sind. Im Gegensatz dazu sind Inhalte auf **geschlossenen Plattformen** nicht ohne weiteres über Suchmaschinen zugänglich und erfordern oft eine zusätzliche Authentifizierung oder eine Einladung. Plattformen enthalten oft sowohl offene als auch geschlossene Elemente, z. B. gibt es bei Facebook öffentliche (offene) und private (geschlossene) Gruppen.

Unter **schädlichen Inhalten und Verhaltensweisen** verstehen wir ein breites Spektrum von Online-Aktivitäten, die negative Auswirkungen auf die Menschenrechte, die Gesellschaft und/oder die Demokratie haben können. Diese können von der gezielten Belästigung von Einzelpersonen über die Anstiftung zu Gewalt gegen eine bestimmte Gruppe bis hin zur Verbreitung von Desinformationen und schädlichen Verschwörungsideologien reichen. In einigen Fällen ist das Schadensrisiko bereits mit dem Inhalt selbst verbunden, wobei die Gefahr durch dessen Verbreitung noch verstärkt wird. In anderen Fällen wird die Schadensgefahr eher durch aggregierte Verhaltensmuster als durch die Art des Inhaltes selbst verursacht. Je nach geografischem und rechtlichem Kontext können verschiedene Formen schädlicher Inhalte oder Verhaltensweisen illegal sein oder nicht. Und je nach Plattform können schädliche Inhalte oder Verhaltensweisen auch durch die Community-Guidelines, Standards oder Regeln eines Unternehmens abgedeckt sein oder nicht.

Unter **Verschlüsselung** versteht man den Prozess der Chiffrierung von Informationen, so dass sie für jede und jeden außer den angegebenen Empfänger:innen unverständlich sind.

Bei **Verschwörungsideologien** handelt es sich um Erklärungsversuche für bestimmte Phänomene, indem ein finsteres, von mächtigen Akteur:innen inszeniertes Komplott heraufbeschworen wird. Verschwörungen werden als geheim oder esoterisch dargestellt, wobei sich die Anhänger:innen einer Theorie als die wenigen Eingeweihten sehen, die Zugang zu verborgenem Wissen haben. Die Anhänger:innen von Verschwörungsideologien sehen sich in der Regel in direkter Opposition zu den Mächten, die das vermeintliche Komplott inszenieren.

Einleitung

Viele Personen und Gemeinschaften, die Extremismus, Hass und Desinformation verbreiten, ziehen sich von den etablierten Plattformen der sozialen Medien zurück. Stattdessen nutzen sie ein breiteres und vielfältigeres Spektrum von Online-Räumen mit noch weniger Moderation, bzw. Plattformen, die mehr Privatsphäre, Sicherheit und Anonymität bieten. In diesem Bericht werden die Ergebnisse der ersten Scoping-Phase eines vom Omidyar Network geförderten und vom Institute for Strategic Dialogue (ISD) und CASM Technology ins Leben gerufenen Projekts vorgestellt, um diese Online-Räume zu identifizieren und geeignete Forschungsmethoden zu entwickeln.

In Phase II des Projekts werden die Erkenntnisse aus der Scoping-Phase auf die Untersuchung von drei kleinen Plattformen in englischer, französischer und deutscher Sprache angewendet, um das Verständnis der Forschung darüber zu erweitern, welche Methoden (mit bestehendem Datenzugriff) auf diese Online-Räume anwendbar sind. In der dritten und letzten Phase des Projekts teilt das ISD die Erkenntnisse aus den Phasen I und II mit politischen Entscheidungsträger:innen. Um die Forschungsergebnisse sowie die Auswirkungen auf die Transparenz der Plattformen und den Datenzugriff mit Vertreter:innen der Regierung, der Regulierungsbehörden, der Forschung und des privaten Sektors zu diskutieren beruft es einen runden Tisch mit Expert:innen ein,. Auf der Grundlage unserer Ergebnisse werden wird zu überlegen seint, wie die rechtlichen und regulatorischen Rahmenbedingungen angepasst werden müsste, um mit der zunehmenden Bandbreite und technologischen Vielfalt der Online-Plattformen Schritt zu halten und gleichzeitig die grundlegenden Rechte auf Privatsphäre, Sicherheit und Anonymität im Internet zu respektieren und zu schützen.

Unser Ziel ist es, die Verbreitung schädlicher Inhalte und Verhaltensweisen im Internet zu verstehen und zu bekämpfen. Die Verbreitung schädlicher Inhalte hat mithilfe verschiedener Kanäle schon immer viele Formen angenommen. Durch die Etablierung und Nutzung des Internets hat es in den letzten Jahrzehnten jedoch eine wichtige technologische Revolution gegeben: nämlich die zunehmende Möglichkeit, Kommunikationsdaten systematisch zu sammeln, zu speichern und präzise zu durchsuchen. Ursprünglich erforderte dies einen speziellen Zugriff auf Daten und war daher weitgehend auf ausgewählte Gruppen (z. B. Eigentümer:innen von Kommunikationstechnologien oder Nachrichtendienste) beschränkt. Die zunehmende Popularität öffentlicher Online-Räume, insbesondere einiger weniger dominanter Social-Media-Plattformen, hat es einem breiten Spektrum von Forscher:innen ermöglicht, verschiedene Formen von Online-Gefahren zu verfolgen, zu analysieren und – hoffentlich – zu bekämpfen. Doch dieser Trend könnte sich nun umkehren. Mehrere soziale und technologische Entwicklungen – das Wachstum von Plattformen, die sich ideologisch gegen Moderation stellen, das Aufkommen neuer Technologien, z. B. Blockchain, Augmented und Virtual Reality (AR/VR) sowie künstliche Intelligenz und die zunehmende Nutzung verschlüsselter Plattformen für private Nachrichtenübermittlung – können in einer Art und Weise zusammenwirken, die es schwieriger macht, schädliche Online-Aktivitäten zu bekämpfen.

Der vorliegende Bericht befasst sich mit diesen Herausforderungen sowie den Methoden und Werkzeugen, die den Forscher:innen zur Verfügung stehen. Nach einer allgemeinen Einführung in die Aufgabe, schädliche Inhalte und Verhaltensweisen im Internet zu identifizieren, werden in Kapitel 1 der Ablauf und die Ergebnisse unserer Rahmenuntersuchung beschrieben. Der Fokus liegt darauf, die aktuelle Landschaft der Online-Plattformen und -Apps zu erfassen, die bei Gemeinschaften beliebt sind, die Extremismus, Hass und Desinformation verbreiten. Darauf aufbauend stellen wir in Kapitel 2 drei Arten von Hindernissen für die Forschung bzw. den Datenzugriff vor, die diese Plattformen aufweisen; außerdem betrachten wir die aktuellen und (potenziellen) künftigen Auswirkungen dieser Hindernisse für die Forschungsgemeinschaft, für politische Entscheidungsträger:innen und Unternehmen. Kapitel 3 gibt einen Überblick über drei allgemeine Forschungsmethoden, potenziellen Stärken und Schwächen der einzelnen Methoden sowie über Tools. In Kapitel 4 schlagen wir Fallstudien zu Plattformen und potenzielle Forschungsansätze zur Überwindung dieser Hindernisse vor, die in Phase II des Projekts erprobt werden sollen. Kapitel 5 präsentiert mögliche Zukunftsszenarien – von pessimistisch bis optimistisch – für Forscher:innen und diejenigen, die schädliche Inhalte und Verhaltensweisen im Internet bekämpfen, und schlägt eine Reihe von ersten Empfehlungen für politische Entscheidungsträger:innen, Plattformen und die Forschungsgemeinschaft vor. In den Anhängen zu diesem Bericht werden schließlich die vollständigen Ergebnisse der Untersuchung vorgestellt und mögliche ethische,

rechtliche und sicherheitsrelevante Risiken im Zusammenhang mit der Erforschung dieser Online-Plattformen weiter untersucht.[1]

Schädliche Inhalte und Verhaltensweisen im Internet

Schädliche Inhalte und Verhaltensweisen können ein breites Spektrum von Aktivitäten umfassen, von Online-Belästigung und Anstiftung zu Gewalt bis hin zur Verbreitung von Desinformationen und Verschwörungsideologien. Das Risiko ergibt sich in manchen Fällen aus den Inhalten selbst; in anderen Fällen wird der Schaden eher durch Verhaltensmuster, als durch die Art des Inhalts verursacht. Bei schädlichen Verhaltensweisen im Internet sind einzelne Inhalte vielleicht nicht besonders schädigend an sich, aber die systematische Verbreitung von ungeprüften Informationen oder polarisierenden Darstellungen kann sich in der Summe als gefährlich erweisen. Belästigung ist ein besonders relevantes Beispiel dafür. Während einzelne provokative oder feindselige Inhalte nicht unbedingt zu erheblichem Schaden führen, kann Belästigung, wenn sie Teil eines Verhaltensmusters ist, das sich in großen Mengen oder über einen längeren Zeitraum gegen bestimmte Personen oder Gemeinschaften richtet, Journalist:innen, Aktivist:innen, Politiker:innen oder Mitglieder marginalisierter Gemeinschaften von der Teilnahme am öffentlichen Geschehen im Internet abhalten.

Bestimmte Inhalte und Verhaltensweisen können die Menschenrechte von marginalisierten Gruppen verletzen, das Vertrauen in demokratische Institutionen und Grundsätze untergraben und es schwierig machen, in politischen Debatten eine gemeinsame Basis zu finden. Dabei kann es sich um eindeutig ideologische Inhalte (z. B. gewalttätige extremistische Inhalte), allgemeinere gesellschaftliche Themen mit politischen Implikationen (z. B. frauenfeindliche „Incel“[i] -Inhalte) oder auch unpolitische, aber schädliche Inhalte (z. B. die Förderung von Selbstverletzungen) handeln. Je nach geografischem und rechtlichem Kontext können verschiedene Formen schädlicher Inhalte und Verhaltensweisen illegal sein oder nicht. Während einige Formen in den meisten Kontexten illegal sind (z. B. terroristische Inhalte oder Material zum sexuellen Missbrauch von Kindern), können die Gesetze für andere Formen wie z. B. Hassrede, von Land zu Land sehr unterschiedlich sein. Viele Akteur:innen, die schädliche Inhalte verbreiten, sind sich der rechtlichen Grenzen bewusst und achten darauf, eine verschlüsselte oder implizite Sprache zu verwenden, um nicht in die Illegalität abzugleiten. Die zunehmende Erkenntnis, dass viele Formen legaler Inhalte dennoch erheblichen Schaden anrichten können, hat zu Diskussionen darüber geführt, wie wie Des- und Misinformationen durch Regulierung bekämpft werden können, z. B. durch das Gesetz über digitale Dienste der EU[2] oder die Online Safety Bill[3] des Vereinigten Königreichs.

Auch private Unternehmen legen „Community Guidelines“ fest. Dabei handelt es sich um Standards oder Regeln, die erlaubte Inhalte und Verhaltensweisen auf der jeweiligen Plattformumreißen. Diese Richtlinien decken in der Regel zumindest illegale Inhalte oder Verhaltensweisen in jenen Rechtsgebieten ab, in denen das Unternehmen tätig ist. Viele große Social-Media-Plattformen gehen sogar noch weiter und verbieten andere Formen schädlicher, jedoch legaler Aktivitäten. Auch wenn die genauen Definitionen, Grenzwerte und Durchsetzungsmaßnahmen sich unterscheiden, haben sich die Richtlinien, Standards und Regeln vieler großer Unternehmen unter dem Druck von Werbetreibenden, der Zivilgesellschaft, des Gesetzgebers und der Nutzer:innen dahingehend angenähert, dass sie ein ähnliches Spektrum an legalen, aber potenziell schädlichen Aktivitäten untersagen.[4]

Im Gegensatz dazu haben wir bei unseren Untersuchungen erhebliche Unterschiede in den Gemeinschaftsrichtlinien, Standards und Regeln bei vielen der kleineren Plattformen festgestellt, die das breitere Online-Ökosystem ausmachen. Verschiedene Plattformen nehmen teilweise radikal unterschiedliche Positionen zu diversen Formen sogenannter „legaler, aber schädlicher“ Aktivitäten ein. Einige verbieten illegale Aktivitäten nur in jenem Rechtsgebiet, in dem sie ansässig sind, während andere noch weitergehen. Diese Abweichung lässt sich auf unterschiedliche Faktoren zurückführen. Einige Plattformen verfügen möglicherweise nicht über ausreichende Ressourcen, um umfassendere Regeln einzuführen und durchzusetzen (z. B. Plattformen, die nur geringe oder gar

i Incels (kurz für „unfreiwillig zölibatär“) bilden eine Online-Subkultur, deren überwiegend männliche Anhänger der Ansicht sind, dass sie von sexuellen Beziehungen ausgeschlossen werden bzw. dass Frauen sie als nicht begehrenswert empfinden. Incel-Gemeinschaften verbreiten oft sehr frauenfeindliche Ideen, und Anhänger der Subkultur haben bereits einige Anschläge mit mehreren Todesopfern verübt; siehe O'Donnell, Catharina und Shor, Eran, „"This is a political movement, friend": Why "incels" support violence“, The British Journal of Sociology, 73(2), Januar 2022, https://onlinelibrary.wiley.com/doi/10.1111/1468-4446.12923.

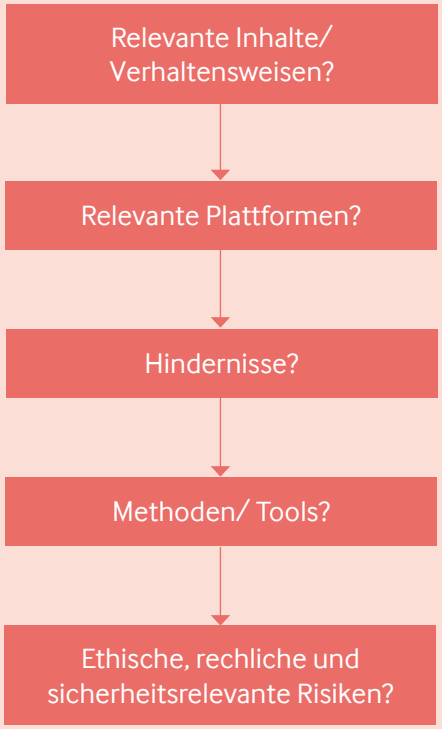
keine Einnahmen oder Gewinne erzielen). Andere Plattformen haben vielleicht ein eher uneingeschränktes Verständnis von Meinungsfreiheit. Darüber hinaus gibt es auch Plattformen, die eine ideologische Haltung einnehmen, z. B. solche, die speziell für extremistische Gemeinschaften eingerichtet wurden (u. a. rechtsextreme Foren wie Iron March oder Fascist Forge).[5]

Wenn schädliche Inhalte und Verhaltensweisen schnell genug erkannt werden, können rechtliche, technische oder andere Maßnahmen ergriffen werden, um den möglichen Schaden zu begrenzen. Beispielsweise können Plattformen eine Reihe von Maßnahmen ergreifen, um die entsprechenden Inhalte oder Konten zu entfernen oder einzuschränken, wenn gegen ihre Regeln verstoßen wird.[6] Die Urheber:innen könnten auch nach nationalem Recht angeklagt werden, wenn sie die Grenze zur Illegalität überschritten haben. Selbst wenn die Inhalte bereits im Umlauf sind, kann das Aufspüren schädlicher Inhalte dazu beitragen, wirksame Gegenargumente zu entwickeln und die Verbreitung so einzudämmen. Ganz allgemein kann das Wissen um schädliche Aktivitäten Trends, Techniken und Tools aufzeigen, die zur Entwicklung solcher Botschaften verwendet werden, und somit dazu beitragen, ebenjene Inhalte und Verhaltensweisen im Allgemeinen besser vorherzusagen, zu erkennen und zu bekämpfen.

Schädliche Inhalte und Verhaltensweisen aufspüren

Digitale Technologien haben wesentlich zum einfacheren Auffinden und Sammeln von Daten beigetragen. Eine Reihe von Werkzeugen hat die Suche erheblich erleichtert, darunter Suchmaschinen wie Google, plattformspezifische Technologien wie CrowdTangle[ii] oder Twitter Advanced Search, marketingorientierte Social-Listening-Programme für soziale Medien wie Brandwatch oder auch forschungsspezifische Technologien wie Method52.[iii] Die Beobachtung eines breiten Spektrums von Online-Räumen ist wesentlich unkomplizierter als die physische Infiltration zahlreicher extremistischer Gruppen.

Abb. 1: Wichtige Forschungsentscheidungen zur Untersuchung schädlicher Inhalte oder Verhaltensweisen im Internet



Mehrere Forschungsansätze ergänzen sich gegenseitig. Beispielsweise können Forscher:innen durch die Suche nach bestimmten Schlüsselwörtern oder Inhalten zu einem neuen Online-Raum geführt werden, wo sie neue Schlüsselwörter oder Themen für ihre Suche entdecken. Oft haben solche Inhalte ihren Ursprung in speziellen Räumen (z. B. in extremistischen Foren), bevor sie auf etablierte Plattformen gelangen, wo sie dann eine viel größere Verbreitung und ein neues Publikum finden. Ebenso ist belegt, dass Belästigungskampagnen häufig in solch spezialisierten Foren koordiniert werden.[7] Aus diesem Grund ist es wichtig, eine Kombination aus der Beobachtung von Online-Nischenräumen und der Suche nach Inhalten während ihrer Verbreitung zu wählen, um den Lebenszyklus schädlicher Inhalte zu verfolgen (und hoffentlich zu verkürzen). Hindernisse technischer, sozialer und/oder rechtlicher Natur können bei der Entdeckung bestimmter Inhalte und Verhaltensweisen oder beim Zugang zu breiteren Online-Räumen diesen positiven Kreislauf jedoch unterbrechen.

ii Ein von Meta übernommenes Analyse-Tool, das Zugriff zu einigen (zunehmend begrenzten) öffentlich verfügbaren Daten von Facebook und Instagram bietet.

iii Method52 ist ein von CASM und der Universität von Sussex entwickeltes Tool zur Analyse sozialer Medien. Für weitere Informationen siehe „Technology and Values", CASM, https://www.casmtechnology.com/pages/technology.

Manche Online-Plattformen sind absichtlich so gestaltet, dass der Zugriff auf Daten minimiert wird. Möglicherweise ist dies nur ein Nebeneffekt anderer Funktionen wie z. B. der Ende-zu-Ende-Verschlüsselung. Es sollte betont werden, dass Funktionen, die auf eine geschützte, private und sichere Kommunikation abzielen, im Hinblick auf den Schutz der Menschenrechte und Privatsphäre große Vorteile bieten. In autoritären Ländern (aber nicht nur dort) schützen sichere Kommunikationstechnologien Aktivist:innen und Dissident:innen vor Überwachung und staatlichen Eingriffen. Die Bekämpfung schädlicher Aktivitäten auf Plattformen, die solche Technologien nutzen, sollte nicht auf Kosten dieser Vorteile gehen.

Zu Beginn eines Projektes zur Erforschung schädlicher Inhalte oder Verhaltensweisen im Internet müssen eine Reihe wichtiger Entscheidungen getroffen werden. Erstens: Welche schädlichen Verhaltensweisen, Dynamiken oder Narrative sind von Interesse? Zweitens: Auf welchen Plattformen werden wir sie voraussichtlich finden? Drittens: Welche Hindernisse stellen die betreffenden Plattformen für die Forschung dar? Viertens: Welche Methoden und Tools stehen zur Verfügung, um diese Hindernisse zu überwinden? Und schließlich: Welche ethischen, rechtlichen und sicherheitsrelevanten Risiken können sich aus diesen Entscheidungen ergeben?[iv] Diese Fragen und Entscheidungen werden in den einzelnen Kapiteln dieses Berichts erörtert. Dabei gilt zu beachten, dass diese Prozesse oft nicht linear verlaufen, sondern parallel zueinander, und sich gegenseitig beeinflussen.

Einiges deutet darauf hin, dass die Hindernisse für die Erforschung schädlicher Inhalte und Verhaltensweisen im Internet zunehmen. Dieses Problem scheint besonders dringlich in jenen Online-Räumen, die weniger Moderation und/oder mehr Privatsphäre, Sicherheit und Anonymität bieten. Um einen Überblick über die aktuelle Landschaft von Plattformen und Apps zu erhalten, die bei schädlichen Gemeinschaften beliebt sind, haben wir eine Liste von Fallstudienplattformen aus drei aktuellen englisch-, französisch- und deutschsprachigen Datensätzen erstellt, die sich auf Extremismus bzw. Verschwörungsideologien konzentrieren. Das Verfahren zur Identifizierung dieser Plattformen sowie die Ergebnisse werden in Kapitel 1 beschrieben. Auf der Grundlage der Plattformen, die in dieser Phase des Scopings identifiziert wurden, werden in Kapitel 2 drei generelle Arten von Hindernissen für die Erforschung und Bekämpfung von schädlichen Inhalten und Verhaltensweisen vorgestellt.

iv Eine ausführliche Erörterung dieser Risiken für Forscher:innen und Organisationen ist in den Anhängen zu diesem Bericht zu finden.

Kapitel 1: Rahmenuntersuchung der Plattformen

Im Hinblick auf die in der Einleitung genannten Themen hat das ISD zunächst eine Liste von Plattformen und Apps zusammengestellt, um neue und aufsteigende Plattformen zu ermitteln – Orte, dievon verschiedenen Gemeinschaften im Jahr 2021 genutz wurden, umExtremismus, Hass oder Desinformation zu verbreiten. Außerdem sind die Hindernisse für die Suche nach schädlichen Inhalten auf diesen Plattformen erfasst und kategorisiert worden.

Als Grundlage für die Analyse verwendete das ISD eine sogenannte Seed-List von Akteur:innen und Gemeinschaften auf Facebook, Instagram, Twitter, YouTube, Reddit, 4chan, Telegram und Gab. Diese Liste wurde aus früheren Forschungsprojekten über Desinformation, Hass und extremistische Gruppen auf Französisch,[8] Englisch[9] und Deutsch zusammengestellt.[10] Diese im Jahr 2021 zusammengestellten Datensätze enthielten Listen von Akteur:innen und Gruppen, die Desinformationen und Verschwörungsideologien über COVID-19 und Impfstoffe verbreitet haben und/oder an rechtsextremen oder antisemitischen Aktivitäten beteiligt waren.[v] Anhand dieser Datensätze konnte das ISD alle URLs zu anderen Plattformen ermitteln, die von diesen Gruppen genutzt werden, und systematisch die häufigsten Plattformen auflisten, auf die die Gemeinschaften verlinkten.

Diese Methodik ist mit einigen Vorbehalten verbunden. Der Ausgangspunkt dieser Studie umfasste nur Plattformen und Gemeinschaften, die für Forscher:innen bereits zugänglich waren und beinhaltete keine geschlossenen oder verschlüsselten Plattformen und keine geschlossenen Messaging-Apps. Darüber hinaus lag der Schwerpunkt unserer Seed-Liste auf rechtsextremen und verschwörungsideologischen Akteur:innen; möglicherweise wandern andere Gemeinschaften und Gruppen (z. B. islamistische Extremist:innen) ebenfalls von Mainstream-Plattformen ab, aber zu einer Reihe von anderen Alt-Tech-Plattformen. Schließlich konzentrierte sich unsere Seed-Liste auf englisch-, französisch- und deutschsprachige Online-Gemeinschaften. Vermutlich gibt es in anderen Sprachen und Länderkontexten noch weitere neue, ebenfalls relevante Plattformen. Aus diesen Gründen sind die Ergebnisse nicht repräsentativ für die gesamte Desinformations- und Hasslandschaft im Internet, sondern beschränken sich auf die in die Analyse einbezogenen Gemeinschaften und Sprachen.

Ein alternativer Ansatz könnte eine breitere Auswahl an Plattformen als Ausgangspunkt nehmen, z. B. geschlossene oder verschlüsselte Messaging-Plattformen wie WhatsApp. Dieser Ansatz würde jedoch zusätzliche ethische und rechtliche Bedenken aufwerfen. Nutzer:innen verwenden diese Dienste in der Annahme, dass ihre Gespräche privat bleiben. Der Zugriff auf diese geschlossenen Bereiche könnte ethische und möglicherweise rechtliche Risiken mit sich bringen, da ein zusätzliches Maß an Einmischungerforderlich wäre. Alternativ hätte eine Auswahl von Plattformen aus der vorhandenen Literatur und bestehenden ethnografischen Untersuchungen zur Identifizierung von Alt-Tech-Plattformen, die anfällig für den Missbrauch durch extremistische Gruppen sind, abgeleitet werden können – obwohl dies die Gefahr mit sich bringt, dass nur die bekanntesten Plattformen berücksichtigt werden. Beide Ansätze könnten in Zukunft verwendet werden, um die von uns zusammengestellte Liste zu ergänzen; für die ersten Untersuchungen der ermittelten allgemeinen Bedrohungen war diese Liste jedoch mehr als ausreichend.

Im Zuge der Analyse wurden 35 Plattformen in französischsprachigen Ländern, 31 in deutschsprachigen und 21 in englischsprachigen Ländern identifiziert.[vi] Um verschiedene Arten von Hindernissen für die Forschung zu ermitteln, sind diese Plattformen anhand ihres Inhalts, technologischen Eigenschaften, Umfangs und ihrer relevanten Nutzungsregeln kategorisiert worden. Als Schlüsselelemente für die Kategorisierung haben wir auch die Einstellung der Plattformen zum Datenschutz und zur freien Meinungsäußerung betrachtet, die anhand der Äußerungen ihrer Gründer:innen, der Unternehmensrichtlinien und/oder der Art ihres Nutzerkreises bewertet wurde.

v Da die Datensätze aus in den letzten Jahren durchgeführten, aber unterschiedlichen Projekten stammten, umfassten sie unterschiedliche Zeiträume und waren von unterschiedlichem Umfang. Die englischsprachigen Daten umfassten 2,5 Millionen Beiträge zwischen dem 1. Januar 2021 und dem 30. November 2021, die deutschsprachigen Daten umfassten 659.000 Beiträge zwischen dem 1. Januar 2021 und dem 12. September 2021, und die französischsprachigen Daten umfassten 2 Millionen Beiträge zwischen dem 31. Juli 2020 und dem 31. Januar 2021.

vi Siehe Anhang: Plattform-Scoping-Daten – Linkzählungen für die vollständige Liste sortiert nach Sprachen.

Um diese erste Liste von Plattformen einzugrenzen und die für unsere Untersuchung relevantesten zu ermitteln, entwickelten wir einen Kodierungsboden und kodierten jede Plattform nach ihren Eigenschaften. Dieser Kodierungsbogen enthielt allgemeine Informationen zu jeder Plattform, wie z. B. die Anzahl der Nutzer:innen weltweit, den Zweck der Plattform, das Gründungsdatum und die Frage, ob sie klare Moderationsregeln verfolgt, insbesondere in Bezug auf Hassrede und Desinformation. Darüber hinaus ermittelten wir, ob die einzelnen Plattformen über Geschäftsbedingungen für die Datennutzung durch externe Parteien verfügen und geschlossene Gruppen anbietet.

Die technologischen Eigenschaften der einzelnen Plattformen wurden ermittelt, um etwaige Hindernisse für deren Erforschung zu bewerten. Zu diesen Eigenschaften gehört, ob die Plattform über eine Suchfunktion und/oder eine API verfügt, ob sie verschlüsselt ist oder neue Technologien wie AR/VR oder Blockchain einsetzt. Schließlich wurden Hindernisse für das Auffinden schädlicher Inhalte festgestellt und in drei Kategorien eingeteilt, auf die in den folgenden Kapiteln näher eingegangen wird:

- Technologische Eigenschaften, die den Zugriff auf Daten blockieren/beschränken
- Ethische und rechtliche Fragen, mit denen die Forscher:innen sich auseinandersetzen mussten
- Fragmentierung der Inhalte auf verschiedenen Plattformen, was eine effiziente und systematische Datenerfassung erschwerte

Da das Ziel dieser Untersuchung letztlich darin bestand, Hindernisse für die Forschung zu ermitteln, beschränkten wir die Auswahl der Plattformen auf diejenigen, die mindestens eines der drei Hindernisse aufwiesen. So kamen wir auf insgesamt 15 Plattformen in den drei Sprachen. Unter diesen Plattformen schlossen wir folgende ein:

- Traditionelle soziale Medien und Messaging-Apps mit geschlossenen Gruppen wie Facebook, VK, Telegram und WhatsApp, da das Vorhandensein von privaten Gruppen zusätzliche ethische Herausforderungen mit sich bringt.
- Discord, weil diese Plattform ethische Hindernisse in den geschlossenen Gruppen und Fragmentierungshindernisse in den öffentlichen Gruppen aufweisen, zumal die Forschung auf der Plattform nur Server für Server und nicht systematisch durchgeführt werden kann.
- Odysee, da sie sowohl Fragmentierung als auch technologische Hindernisse aufweist.
- Kik, weil der Inhalt von Chats mit den bestehenden Methoden und Werkzeugen nicht zugänglich ist und somit ein technologisches Hindernis darstellt.
- Eine Reihe weiterer Plattformen, die sowohl technologische als auch ethische Hindernisse besitzen (nandbox, Hoop Messenger, Riot, Minds und Rocket.Chat).
- Vimeo, DLive und Spotify, da die Beschränkungen bei der Analyse audiovisueller Inhalte (und im Fall von DLive die Verwendung der Blockchain-Technologie) technologische Hindernisse darstellen.

Englisch

	Telegram	Minds	Discord	Facebook	VK
Leitung	Pawel Durow (CEO)	Bill Ottman (CEO)	Jason Citron (CEO)	Mark Zuckerberg (CEO)	Wladimir Kirijenko (CEO)
Anzahl der Nutzer:innen weltweit	500 Mio.	2,5 Mio.	350 Mio.	2,89 Mrd.	460 Mio.
Klare Inhaltsrichtlinien?	Richtlinien nur gegen die Verbreitung von Gewalt und illegaler Pornografie	Ja	Ja	Ja	Richtlinien gegen Terror, Propaganda und Hassrede, aber nicht gegen Desinformation
Zweck	Alternative Chat-Plattform zur Vermeidung staatlicher Überwachung	Alternative zu Facebook, da diese eine große Menge an Daten sammelt	Kommunikation für Gamer:innen	Soziales Netzwerk	Soziales Netzwerk
Gründungsjahr	2013	2011 (Start 2015)	2015	2004	2006
Bedingungen für die Datennutzung?	Ja	Verbietet den Datenexport	Erlaubt kein Data-Mining bzw. Datenextraktion	Ja	Ja
Embedded Analytics?	Nein	Ja	Ja	Ja	Ja
Aufzeichnung der Domainregistrierung verfügbar?	Ja	Ja	Ja	Ja	Ja
Geschlossene Gruppen?	Ja (Ende-zu-Ende-Verschlüsselung bei Chats)	Nein	Ja	Ja	Ja
Fragmentierungshindernisse?	Nein	Nein	Ja	Nein	Nein
Ethische und rechtliche Hindernisse?	Ja, geschlossene Gruppen	Ja, geschlossene Gruppen	Ja, geschlossene Gruppen	Ja, geschlossene Gruppen	Ja, geschlossene Gruppen
Technologische Hindernisse?	Nein	Ja, Ende-zu-Ende-Verschlüsselung und Blockchain	Nein	Nein	Nein
Suchfeld?	Ja	Ja	Ja	Ja	Ja
API?	Ja	Ja	Ja	Ja	Ja
Link zur API	https://core.telegram.org/	https://gitlab.com/minds/engine	https://support.discord.com/hc/en-us/articles/212889058-Discord-s-Official-API	https://developers.facebook.com/docs/pages/	https://vk.com/dev
Verschlüsselt?	Gruppen und Kanäle verwenden Cloud-Verschlüsselung; Chats verwenden Ende-zu-Ende-Verschlüsselung	Ja	Ja, Standardverschlüsselung	Nein	Nein
Neue Technologien?	Nein	Ja, Blockchain	Nein	Ja, VR	Nein
Anmerkungen		Sammelt Statistiken über das Nutzerverhalten. Data-Mining der populärsten Konten und gelegentliches Veröffentlichen der Daten. Gibt keine persönlichen Informationen weiter.			

Deutsch

	DLive	Hoop Messenger	nandbox	Odysee	Riot/Element	Rocket.Chat	WhatsApp
Leitung	Justin Sun (CEO)	Sahand Adilipour (President)	Hazem A. Maguid (CEO)	Julian Chandra (CEO)	Matthew Hodgson (CEO und CTO)	Gabriel Engel (CEO)	Mark Zuckerberg (CEO)
Anzahl der Nutzer:innen weltweit	5 Mio.	Unklar	Unklar	8,7 Mio.	35 Mio.	12 Mio.	2 Mrd.
Klare Inhaltsrichtlinien?	Ja	Ja	Ja	Ja, aber nicht für Desinformation	Nein, bietet aber Leitlinien für Moderator:innen	Ja, aber nicht für Desinformation und Hassrede	Ja
Zweck	Live-Streaming	Sichere Nachrichtenübermittlung	Sichere Nachrichtenübermittlung	Dezentrale Plattform zur gemeinsamen Nutzung von Videos	Dezentralisierte, sichere Nachrichtenübermittlung	Sichere Nachrichtenübermittlung	Nachrichtenübermittlung
Gründungsjahr	2017	2014	2016	2020	2016 (als Riot)	2015	2009
Bedingungen für die Datennutzung?	Ja, gibt Daten an Dritte weiter	Gibt keine Daten weiter, es sei denn, dies ist gesetzlich vorgeschrieben	Gibt keine kommerziellen Daten weiter, kooperiert aber mit den Strafverfolgungsbehörden	Gibt keine persönlich identifizierbaren Daten weiter, stellt aber anonymisierte Daten zur Verfügung	Nur in Ausnahmefällen zur Einhaltung der gesetzlichen Bestimmungen	Nein	Gibt Daten an andere Meta-Unternehmen und an Dritte weiter
Embedded Analytics?	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Aufzeichnung der Domainregistrierung verfügbar?	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Geschlossene Gruppen?	Nein	Ja	Ja	Nein	Ja	Ja	Ja
Fragmentierungshindernisse?	Nein	Nein	Nein	Ja	Nein	Nein	Nein
Ethische und rechtliche Hindernisse?	Nein	Ja	Ja	Nein	Ja	Ja	Ja
Technologische Hindernisse?	Ja, audio-visuell	Ja, Verschlüsselung	Ja, Verschlüsselung	Ja, audio-visuell	Ja, Ende-zu-Ende-Verschlüsselung	Ja, Verschlüsselung	Ja
Suchfeld?	Ja	Ja	Ja	Ja	Ja (öffentliche Räume)	Nein	Nein
API?	Ja	Nein	Ja	Nein	Ja	Nes	Nein (größtenteils)
Link zu API	https://docs.dlive.tv/api/	Keine Angaben	https://api.nandbox.com/#nandbox-api	Keine Angaben	https://element.io/developers	https://developer.rocket.chat/reference/api	https://www.whatsapp.com/business/api
Verschlüsselt?	Nein	Ja	Ja	Nein	Ja	Ja	Ja
Neue Technologien?	Ja, Blockchain	Nein	Nein	Ja, Blockchain	Ja, dezentrales Protokoll	Nein	Nein
Anmerkungen		Kanäle können gelöscht werden					

Französisch

	Spotify	Vimeo	Kik
Leitung	Daniel Ek (CEO)	Anjali Sud (CEO)	Ted Livingston (CEO)
Anzahl der Nutzer:innen weltweit	173 Mio. (Premium-Abonnenten)	175 Mio.	300 Mio.
Klare Inhaltsrichtlinien?	Ja, aber nicht für Desinformation	Ja, einschließlich Desinformation zu ausgewählten Themen	Ja
Zweck	Audio-Streaming	Hosting und gemeinsame Nutzung von Videos	Nachrichtenübermittlung
Gründungsjahr	2006	2004	2010
Bedingungen für die Datennutzung?	Gibt anonymisierte Daten an Forscher:innen weiter	Nein	Nein
Embedded Analytics?	Ja	Ja	Ja
Aufzeichnung der Domainregistrierung verfügbar?	Ja	Ja	Ja
Geschlossene Gruppen?	Nein	Nein	Nein
Fragmentierungs-hindernisse?	Ja	Nein	Nein
Ethische und rechtliche Hindernisse?	Nein	Nein	Nein
Technologische Hindernisse?	Ja, Audiomaterial	Ja, audio-visuelles Material	Ja, Inhalt nicht zugänglich
Suchfeld?	Ja	Ja	Ja, aber für Nutzer:innen, nicht für Inhalte
API?	Ja	Ja	Ja
Link zu API	https://developer.spotify.com/documentation/web-api/	https://developer.vimeo.com/api/reference	https://kik.readthedocs.io/en/latest/api.html
Verschlüsselt?	Ja, Musik	Nein	Nein
Neue Technologien?	Nein	Nein	Nein

Kapitel 2: Drei Hindernisse

In diesem Paragraf stellen wir drei generelle Arten von Forschungshindernissen vor. Die einzelnen Hindernisse schließen sich dabei nicht gegenseitig aus. Obwohl wir uns in erster Linie auf die Auswirkungen der einzelnen Arten von Hindernissen, auf das Auffinden schädlicher Inhalte und Verhaltensweisen konzentrieren, bringt jedes dieser Hindernisse zusätzliche Herausforderungen für die Bekämpfung der Auswirkungen solcher Aktivitäten mit sich; einige dieser Herausforderungen werden wir ebenfalls kurz vorstellen. Es gibt nur sehr wenige Fälle, in denen diese Hindernisse die Durchführung von Recherchen auf einer bestimmten Plattform völlig unmöglich machen. Im nächsten Kapitel untersuchen wir eine Reihe von Methoden und Werkzeugen, die zur Überwindung dieser Hindernisse beitragen können.

Hindernis 1: Technologie

Die Technologie kann den Datenzugriff erheblich verbessern, sie kann ihn aber auch einschränken. Plattformen können absichtlich Technologien verwenden, die den Zugriff auf Daten einschränken, oder sie auch andere technologische Eigenschaften aufweisen, die unbeabsichtigt Hindernisse für Forscher:innen schaffen. Die technologischen Eigenschaften bestimmter Inhalte können die Fähigkeit von Forscher:innen einschränken, systematische, groß angelegte Datenanalysen durchzuführen.

Einige dieser Technologien sind vielleicht schon bekannt, stellen aber immer noch ein Hindernis dar, andere wiederum sind neu oder werden gerade entwickelt. Zu den Technologien gehören:

- **Verschlüsselung**: Bei der Verschlüsselung werden Inhalte für alle außer bestimmten Empfänger:innen unverständlich gemacht. Eine systematische Datenerfassung für Forscher:innen ist nicht möglich, wenn die Absender:innen bzw. Empfänger:innen keinen Zugriff gewähren.
- **Neue Formate**: Bestimmte Inhalts- oder Datenformate, insbesondere Audio- und audiovisuelle Formate, lassen sich (noch) nicht so gut systematisch durchsuchen und speichern wie Text. Die Art der Inhalte bzw. der Daten, die Forscher:innen auf einer Plattform sammeln und analysieren können, hat erhebliche Auswirkungen auf die Art der Analyse, die durchgeführt werden kann. Textdaten von traditionellen Social-Media-Plattformen wie Facebook, Instagram, Twitter und VK können relativ leicht untersucht werden, vor allem wenn eine systematische Suchfunktion zur Verfügung steht (z. B. über eine API); überwiegend audiovisuelle Plattformen wie YouTube oder Audio-Plattformen wie Spotify stellen jedoch zusätzliche Herausforderungen dar, da Video- und Toninhalte nicht auf dieselbe Weise durchsucht bzw. analysiert werden können. Audiovisuelle Inhalte für AR/VR-Technologien werden zunehmend entwickelt, und es gibt bereits Hinweise darauf, dass sie zur Verbreitung schädlicher Inhalte oder zur Belästigung anderer Nutzer:innen verwendet werden. Dies könnte sich in Zukunft noch erheblich verstärken, wenn sich diese Art von Technologien weiter durchsetzt.[11] Der Echtzeitcharakter und die flüchtige Natur von AR/VR-Aktivitäten stellt auch eine Herausforderung für systematischere Datenerfassungsansätze dar.
- **KI-generierte Inhalte**: Wie das Beispiel der „Deep Fakes" zeigt, werden die von künstlicher Intelligenz erzeugten Inhalte immer glaubwürdiger. Die Geschwindigkeit, mit der neue Inhalte entwickelt werden können, erschwert eine systematische Datenerhebung.
- **Dezentralisierung**: Dies ermöglicht den Plattformen, ohne zentrale Verwaltung zu operieren, und kann die Optionen der Administrator:innen einschränken, Inhalte zu entfernen oder Nutzer:innen zu sperren – insbesondere solche, bei denen schädliche Verhaltensmuster festgestellt wurden. Neben dezentralisierten Plattformen gibt es auch Projekte, die eine dezentralisierte Kommunikation zwischen Plattformen ermöglichen sollen.[vii] Die Dezentralisierung kann daher zu einer weiteren Fragmentierung führen und die Möglichkeiten für einen systematischeren Datenzugriff für Forscher:innen verringern.
- **Blockchain**: Hierbei handelt es sich um eine Technologie, mit der Ereignisse (z. B. wer wann welche Inhalte veröffentlicht hat) in einem unveränderlichen Ledger (englisch für „Hauptbuch")

vii Siehe den Ecosystem Review, der vor der Einführung des dezentralen Twitter-Protokolls Bluesky erstellt wurde.

aufgezeichnet werden. Dies befähigt dazu, den aktuellen, wahren Zustand eines Systems zu ermitteln, indem der aktuelle Stand des Ledgers konsultiert wird, ohne dass menschliche Vermittlung erforderlich ist. Eine Blockchain kann daher für die Dezentralisierung eingesetzt werden (z. B. bei Plattformen wie Riot). Die Blockchain-Technologie wird auch häufig zur Unterstützung von Zahlungen in Kryptowährungen verwendet, was zunehmend von Plattformen genutzt wird, um es Nutzer:innen zu ermöglichen, Inhalte direkt zu monetarisieren, anstatt sich auf Werbung zu verlassen (z. B. Odysee und LBRY). Diese finanziellen Anreize bergen die Gefahr, dass die Verbreitung schädlicher Inhalte zu einem Geschäftsmodell wird, das sich aufgrund seiner Abhängigkeit von der Blockchain-Technologie als besonders resistent gegenüber Regulierung oder Eindämmung erweisen könnte. Aus Sicht der Forschung ist die systematische Erfassung von Daten aus blockchainbasierten Plattformen noch relativ neues Territorium. Da teilweise blockchainbasierte Plattformen wie Odysee nicht über öffentliche APIs verfügen, ist auch unklar, welche Daten verfügbar werden könnten und ob im Laufe der Datenerhebung möglicherweise weitere Hindernisse auftauchen.

Viele dieser Technologien stellen Hindernisse für bestimmte Formen des Datenzugriffs dar und können von Plattform zu Plattform variieren. Betrachten wir zum Beispiel einen verschlüsselten Telegram- oder Signal-Chat im Vergleich zu einer privaten Facebook-Gruppe: Obwohl beide ähnliche ethische Fragen aufwerfen (wie zuvorerörtert), dürfte die Verschlüsselung von Telegram oder Signal für Forscher:innen, die einen ethnografischen Ansatz verfolgen, im Vergleich zu einer unverschlüsselten, privaten Facebook-Gruppe keine zusätzlichen Probleme aufwerfen; in beiden Fällen muss die Erlaubnis der anderen beteiligten Nutzer:innen eingeholt werden, um den Zugriff zu erhalten. Für die Unternehmen selbst, als auch für die Strafverfolgungsbehörden und Nachrichtendienste, stellt die Verschlüsselung jedoch zusätzliche technologische Hindernisse dar: ein privater und verschlüsselter Telegram- bzw. Signal-Chat ist ohne die Erlaubnis der beteiligten Nutzer:innen unzugänglich; Meta könnte jedoch gegen den Willen der Administrator:innen den Zugriff auf Daten aus der privaten, aber unverschlüsselten Facebook-Gruppe erzwingen.

Zusätzliche technologische Herausforderungen bei der Eindämmung schädlicher Inhalte und Verhaltensweisen

Die Bandbreite dieser Herausforderungen ist so groß wie die Bandbreite der neuen Technologien. Einige Beispiele sind:

- **Neue Formate**: Es ist denkbar, dass neue, möglicherweise auf AR/VR basierende Formen von Inhalten, sich als viel ansprechender und effektiver bei der Radikalisierung des Publikums erweisen und/oder schädlichen Inhalten zu einer größeren Verbreitung oder Wirkung verhelfen. Der Marktdruck könnte dazu führen, dass die Plattformen nicht bereit sind, die Einführung neuer Technologien zu verlangsamen, selbst wenn sie mit solchen Problemen konfrontiert werden.
- **KI-generierte Inhalte**: Dies könnte dazu führen, dass sich Inhalte schneller verbreiten, als sie bekämpft werden können. Die Automatisierung (z. B. durch „Bots") wird bereits jetzt genutzt, um schädliche Inhalte schnell zu vervielfältigen und zu verbreiten. Eine höherstehende KI könnte über die Vervielfältigung hinausgehen und die Mutation von Inhalten unter Beibehaltung ihrer ursprünglichen Bedeutung ermöglichen.
- **Blockchain**: Eine besonders strikte Nutzung der Blockchain-Technologie könnte die Löschung von Inhalten durch eine zentralisierte Behörde unmöglich oder nahezu unmöglich machen (z. B. eine Situation, in der zuwiderhandelnde Nutzer:innen der Löschung ihrer Inhalte zustimmen müssten), auch wenn sich die Frage stellt, wie dies in Verbindung mit rechtlichen Anforderungen funktionieren würde.[12]

Hindernis 2: Ethik und Recht

Der Zugriff auf Daten aus Online-Räumen, insbesondere die Sammlung und Verarbeitung dieser Daten, kann ethische Fragen aufwerfen, z. B. ob die Privatsphäre der Nutzer verletzt wird oder die Verwendung von Daten oder Inhalten ohne Zustimmung der Nutzer:innen legitim ist. Dies kann auch zu Verstößen gegen ethische Forschungspraktiken, Geschäftsbedingungen der Plattformen oder sogar gegen das Gesetz führen. Für akademische Forscher:innen, die oft strenge ethische Genehmigungsverfahren durchlaufen

und einschlägige rechtliche Anforderungen einhalten müssen, kann diese Herausforderung besonders groß sein. Die Strafverfolgungsbehörden (und in vielen Ländern die Nachrichtendienste) unterliegen ebenfalls zusätzlichen rechtlichen Beschränkungen für den Zugriff auf und die Nutzung von personenbezogenen Daten. Dies ist aus einer Vielzahl von Gründen wünschenswert, vor allem wegen des Menschenrechts auf Privatsphäre und der Gewährleistung eines ordnungsgemäßen Verfahrens. Das Recht auf Privatsphäre ist zwar nicht absolut, Ausnahmen müssen jedoch im Rahmen der Rechtsstaatlichkeit gerechtfertigt sein. Folglich können Datenschutzbestimmungen die Möglichkeit einschränken, schädliche Inhalte zu finden. Einige Forscher:innen haben argumentiert, dass die weltweite Zunahme von Datenschutzgesetzen – allen voran die Datenschutz-Grundverordnung (DSGVO)[viii] in der EU und die von der DSGVO abgeleiteten Gesetze in anderen Ländern – Plattformen einen zusätzlichen Anreiz geben könnte, Daten nicht weiterzugeben.[13]

Messaging-Apps wie WhatsApp sind ein drängendes, aktuelles Beispiel. Über WhatsApp wird eine große Menge an Inhalten ausgetauscht, darunter Formen der Desinformation, Aufrufe zur Gewalt und andere schädliche Inhalte. Für Forschende, die Mitglieder einer WhatsApp-Gruppe sind, ist das Sammeln von Daten unglaublich einfach: WhatsApp verfügt über eine einfache Funktion zum Exportieren des gesamten Chatverlaufs als Textdatei. Aber wie sind die Forscher:innen dieser Gruppe beigetreten? Haben sie die ausdrückliche Erlaubnis aller Mitglieder eingeholt, den Inhalt der Gruppe für Forschungszwecke zu nutzen (was die Teilnehmer:innen möglicherweise zur Selbstzensur veranlassen würde)? Wissen die Gruppenmitglieder überhaupt, dass es einen Forscher bzw. eine Forscherin in ihrem Chat gibt und haben sie ihre Zustimmung gegeben, an einer Forschung teilzunehmen? Haben sich die Forscher:innen möglicherweise durch Täuschung Zutritt zur Gruppe verschafft?

Ähnliche ethische Probleme können bei der Erforschung von Discord auftreten. Der API-Client von Discord ermöglicht es Forscher:innen, sich mit einem Server zu verbinden und Nachrichten aus den Kommunikationskanälen („Channels“) live zu erfassen sowie historische Nachrichten zu sammeln. Zwei Möglichkeiten stehen den Forscher:innen zur Verfügung, um sich mit einem Server verbinden, die jeweils ein unterschiedliches Maß an Täuschung erfordern. Im ersten Fall muss ein Bot-Konto manuell von einem Server-Administrator bzw. einer Server-Administratorin (z. B. von der Person, die den Server erstellt hat, oder von einer anderen Person, die über diese Rechte verfügt) zum Server hinzugefügt werden; diese Person hat die Möglichkeit, einen solchen Zugang zu verweigern. Außerdem wird der Bot in der Benutzerliste eindeutig als solcher gekennzeichnet, was insbesondere in Gemeinschaften, die sensible Themen diskutieren, Verdacht erregen könnte. Die zweite Möglichkeit besteht darin, einen Bot hinter einem normalen Benutzerkonto laufen zu lassen (ein so genannter „Self-Bot“). In diesem Fall treten die Forscher:innen dem Server als normale Benutzer:innen bei (z. B. mit einem Einladungslink), und der Bot gibt sich anschließend als dieser Benutzer bzw. diese Benutzerin aus; dieses betrügerische Verhalten verstößt jedoch gegen die Nutzungsbedingungen von Discord ,was eine zusätzliche ethische Herausforderung darstellt.

Bei Messaging-Apps, die als Hauptbestandteil ihres Marktangebots ausdrücklich mehr Privatsphäre und Sicherheit versprechen als Mainstream-Optionen wie WhatsApp, sind diese Probleme möglicherweise noch gravierender. Plattformen, die einen stärkeren Fokus auf die Privatsphäre ihrer Nutzer:innen versprechen, haben ebenfalls Gemeinschaften angezogen, die Extremismus, Hass oder Desinformationen verbreiten. So hat beispielsweise MeWe, das 2012 von Mark Weinstein gegründet wurde – einem Verfechter des Schutzes der Privatsphäre – großen Erfolg bei Verschwörungsideolog:innen und Rechtsextremist:innen.[14] Kik, ein anonymer Instant-Messaging-Dienst, wird Berichten zufolge genutzt, um die sexuelle Ausbeutung von Kindern zu erleichtern.[15] Wie im vorherigen Kapitel über technologische Hindernisse dargelegt, kommen bei diesen Plattformen häufig Verschlüsselungsverfahren zum Einsatz. Außerdem ist es unwahrscheinlich, dass solche Gruppen Forscher:innen willkommen heißen.

Zusätzliche ethische und rechtliche Herausforderungen bei der Bekämpfung von schädlichen Inhalten und Verhaltensweisen

Da viele Plattformen als Reaktion auf die zunehmenden Regulierungen und Moderationspraktiken in den

viii Die DSGVO ist eine Verordnung der Europäischen Union zur Regelung des Datenschutzes. Sie regelt die Erhebung, Speicherung und Übertragung personenbezogener Daten und hat daher wichtige Auswirkungen auf die Online-Forschung.

traditionellen sozialen Medien entstanden sind, werden diese Alt-Tech-Plattformen oft als Bollwerk der „freien Meinungsäußerung" dargestellt und ziehen Gemeinschaften und Ideologien an, die in anderen Bereichen gesperrt worden sind aufgrund von Verstößen gegen die Gemeinschaftsstandards und/oder Richtlinien gegen Hass, Desinformation und Belästigung. Daher ist es wahrscheinlich, dass die Plattformmoderation (und damit auch die allgemeinen Geschäftsbedingungen und die allgemeine Plattformaktivität) Maßnahmen ausdrücklich entgegensteht wie der Entfernung von Inhalten, Kontosperrungen oder sogar der Abwertung schädlicher Inhalte in algorithmischen Empfehlungen, Newsfeeds oder Suchergebnissen.

Die Plattform Minds positioniert sich beispielsweise ausdrücklich als Gegnerin der Zensur und Befürworterin der Meinungsfreiheit. Obwohl dies beispielsweise viele Impfgegner anzieht, ist es sehr unwahrscheinlich, dass die Moderator:innen die Löschung dieser Inhalte unterstützen würden (obwohl die Website auch illegale Inhalte löscht). Nichtsdestotrotz präsentiert sich das Unternehmen als Gegner von Fehlinformationen und argumentiert, dass diese durch Gegenrede bekämpft werden sollten.[ix]

Hindernis 3: Fragmentierung

Viele Online-Inhalte, darunter auch schädliche Inhalte, sind zwar theoretisch ohne Hindernisse durch technologische Strukturen bzw. durch ethische und rechtliche Aspekte online zugänglich; dennoch müssen Forscher:innen wissen, wo sie suchen müssen. Oft befinden sich relevante Inhalte in riesigen Mengen von Material, das nicht schnell und systematisch durchsucht werden kann, beispielsweise mithilfe einer plattformweiten Suchfunktion oder API. Wir bezeichnen Plattformen, auf denen theoretisch zugängliche Inhalte nicht schnell oder systematisch durchsucht werden können, als „fragmentiert". Da die Inhalte öffentlich sichtbar sind, ohne technologische bzw. ethische und rechtliche Hindernisse, können fragmentierte Plattformen als eine Unterkategorie offener Plattformen betrachtet werden.[x] Allerdings sind nicht alle offenen Plattformen fragmentiert, denn einige bieten Forscher:innen die Möglichkeit, Inhalte systematisch zu durchsuchen. Fragmentierte Plattformen sind auch von geschlossenen Plattformen zu unterscheiden. Während geschlossene Plattformen ebenfalls nicht systematisch durchsucht werden können, ist der Zugang zu ihnen ohne zusätzliche Informationen oder Genehmigungen (z.B. Passwörter oder andere Arten der persönlichen Identifizierung) nicht möglich.

Dies war im Laufe der Forschungsgeschichte oft der Fall, wie alle Historiker:innen bestätigen können, die schon einmal ein schlecht beschriftetes physisches Archiv durchsuchen mussten. Moderne Suchwerkzeuge, vor allem Google, aber auch plattformspezifische Technologien wie CrowdTangle[xi] oder die Twitter-API, haben erst in jüngster Zeit dafür gesorgt, dass es für Forscher:innen einfacher geworden ist, Inhalte schnell und systematisch aufzufinden. Diese Erleichterung kann jedoch überbewertet werden und wurde es auch schon oft. Ein großer Teil des Internets, möglicherweise über 90 %, erscheint nicht in einer Google-Suche. Dies ist das so genannte „Deep Web".[xii] Außerdem sind wichtige Formen der sozialen Medien und der Online-Kommunikation wie private und/oder verschlüsselte Nachrichten, E-Mails und geschlossene Gruppen für externe Forscher:innen schon immer unzugänglich gewesen. Nichtsdestotrotz ist die schnelle und systematische Suche als Methode zur Aufdeckung von schädlichen Inhalten und Verhaltensweisen wesentlich einfacher geworden. Zwei konvergierende Trends könnten jedoch die Wirksamkeit dieser Methoden einschränken.

Der erste Trend ist, dass viele Online-Plattformen, sowohl neue als auch etablierte, die Menge an Daten reduzieren, auf die über APIs oder andere Tools zugegriffen werden kann. Am auffälligsten ist dies bei Facebook. Bis 2014

ix Das ISD definiert Gegenrede bzw. Gegennarrative als Botschaften, die eine positive Alternative zu extremistischer Propaganda bieten und/oder darauf abzielen, extremistische Narrative zu dekonstruieren bzw. zu delegitimieren.

x Während geschlossene Plattformen auch nicht systematisch durchsucht werden können, ist der Zugang zu ihnen ohne zusätzliche Informationen (z. B. Passwörter oder andere Arten der persönlichen Identifizierung) nicht möglich. Siehe Glossar.

xi CrowdTangle ist ein Tool für das Durchsuchen öffentlicher Inhalte auf Facebook und Instagram. Es gehört zum Unternehmen Meta, das im Laufe der Zeit die verfügbaren Daten eingeschränkt hat. Nichtsdestotrotz ermöglicht CrowdTangle eine schnelle Abfrage von Schlüsselwörtern, um eine enorme Bandbreite an Material zu erhalten.

xii Genau genommen besteht das Deep Web aus Online-Material, das von Suchmaschinen nicht „indiziert" wird und daher bei einer Suche über Google, Bing, DuckDuckGo usw. nicht erscheint. Dazu gehört eine breite Palette von Inhalten, die viele Menschen täglich nutzen, z. B. alle Inhalte, für die ein Passwort erforderlich ist oder die sich hinter einer Bezahlschranke befinden. Das Deep Web ist nicht mit dem „Dark Web" zu verwechseln, auf das nur über bestimmte Browser zugegriffen werden kann und das häufig für illegale Aktivitäten genutzt wird.

konnten Forscher:innen die „Graph Search"-Funktion der Facebook-API verwenden, um nicht nur auf die Daten der Nutzer:innen zuzugreifen, sondern auch auf die Daten ihrer Freunde. Selbst nach der Ankündigung erheblicher neuer Beschränkungen im Jahr 2014[16] konnten Forscher:innen immer noch problemlos jeden Beitrag auf einer öffentlichen Facebook-Seite und jeden Kommentar zu diesem Beitrag sowie die zugehörigen Profilinformationen für jeden Beitrag und Kommentar herunterladen. Im Jahr 2018 wurde die API stark eingeschränkt. Der Zugriff auf die Daten erfolgt seitdem weitgehend über Facebook-Partnerbetriebe, insbesondere CrowdTangle.[17] Das bedeutet, dass viele Schlüsselbereiche der Plattformen wie z. B. private Gruppen oder Seiten außerhalb des Anwendungsbereichs der API liegen, was Forscher:innen dazu zwingt, ältere, arbeitsintensivere und weniger systematische Forschungsmethoden anzuwenden, wie z. B. das manuelle Suchen und Lesen von Material.

Der zunehmende regulatorische und öffentliche Druck hat zwar Vorteile in Bezug auf die Stärkung des Schutzes der Privatsphäre und der Datenrechte, dennoch könnte es sein, dass die Suchwerkzeuge und APIs der Plattformen von Haus aus restriktiver werden. Abhängig von der Größe der Plattform und der Rechtsprechung, in der sie tätig ist, können neue Vorschriften wie das EU-Gesetz über digitale Dienste auch zu einem umfassenderen Zugang für Forscher:innen oder die breite Öffentlichkeit führen. Viele der neueren Plattformen in unseren Fallstudien (siehe Kapitel 4) verfügen nicht über plattformweite Suchfunktionen, auch nicht als Teil ihrer APIs. Zwar ist es oft noch möglich, mit relativ alten Technologien auf relevante Daten zuzugreifen, doch handelt es sich dabei oft um eher ad-hoc „zusammengebastelte" Methoden, wie z. B. die Entwicklung von Bots, die menschliche Nutzer:innen imitieren und Text replizieren. Diese Technologien müssen für bestimmte Zwecke entwickelt und gepflegt werden, einschließlich der Erstellung von Daten in einem systematischen Format. Dies erfordert einen wesentlich höheren Aufwand als die Verwendung allgemeiner Such-APIs. In einigen Fällen kann die Verwendung solcher Technologien für den Datenzugriff auch gegen die Nutzungsbedingungen der Plattformen verstoßen, was zusätzliche ethische und rechtliche Herausforderungen mit sich bringt.

Ein zweiter potenzieller Trend ist die breitere Fragmentierung von hasserfüllten Online-Räumen. Die zunehmende Bereitschaft vieler großer Plattformen gegen schädliche Inhalte und Verhaltensweisen vorzugehen, könnte diese Gemeinschaften dazu veranlassen, eine Vielzahl von alternativen Räumen zu suchen oder aufzubauen. Insbesondere im Vorfeld und unmittelbar nach dem Sturm auf das US-Kapitol am 6. Januar 2021 beobachteten Forscher:innen, dass Trump-freundliche Akteur:innen von den zunehmend strengeren Plattformen Facebook und Twitter zu bereits bestehenden „Meinungsfreiheit"-Räumen wie Gab und Parler wechselten. Diese sind angeblich auch zur Koordinierung von Teilen des Aufstands genutzt worden.[18] Tatsächlich wurde Parler am 8. Januar 2021, nachdem Facebook und Twitter die Konten von Präsident Trump auf ihren Plattformen gesperrt hatten, die am häufigsten heruntergeladene App.[19] Nachdem Parler am 10. Januar 2021 der Zugang zu Amazons Cloud-Hosting-Dienst Amazon Web Services (AWS) verweigert wurde, schienen die Nutzer:innen zu Gab zu wechseln. In den folgenden zwei Monaten wurden laut Datenlecks 2,4 Millionen Konten auf Gab erstellt. Man an geht davon aus, dass Gab etwa 4 Millionen Konten beherbergt, obwohl die Zahl der aktiven Nutzer:innen eher auf 100.000 geschätzt wird.[20]

Technische Funktionalitäten könnten zu diesem Trend beitragen. Plattformen wie nandbox ermöglichen es den Nutzer:innen, mit wenig technischem Know-how neue Messenger-Apps zu erstellen. Diese Art von Diensten könnte die rasche Fragmentierung potenzieller Räume für die Aufnahme extremistischer Inhalte und Gemeinschaften erleichtern. Für Forscher:innen bieten solche Plattformen eindeutige Räume zum Auffinden und Erforschen schädlicher Inhalte, wenn ihre Popularität unter Extremisten weithin bekannt ist; wir können jedoch nicht davon ausgehen, dass die Fragmentierung immer zu solchen eindeutigen Räumen zum Auffinden schädlicher Inhalte führen wird. Es gibt eine Reihe von großen, fragmentierten Plattformen wie Discord, Spotify und DLive, auf denen schädliche Inhalte inmitten einer riesigen Menge anderer textlicher oder audiovisueller Inhalte unentdeckt bleiben könnten (und dies auch schon tun).

Plattformen wären im Stand, eine Fragmentierung in Kombination mit anderen Hindernissen aufzuweisen. So können beispielsweise Textinhalte offen zugänglich sein und in Kommentar-Threads unterhalb der Videos verbreitet werden. Ohne die Möglichkeit, audiovisuelle Inhalte systematisch zu durchsuchen oder live gestreamte Inhalte zu erfassen und aufzubewahren, vermitteln die

Kommentare möglicherweise jedoch ein unvollständiges Bild der betreffenden Aktivität; dies ist eine Kombination aus Fragmentierung und technologischen Hindernissen. Alternativ mischen Plattformen private und öffentliche Kanäle eventuell so, dass es unklar ist, ob die Gesamtheit der relevanten Aktivitäten durch eine umfassende Analyse der öffentlichen Kanäle allein verstanden werden kann. Möglich ist auch, dass der Zugriff auf private Chats erforderlich ist, um die Art der Aktivitäten vollständig zu verstehen. Dies könnte ein unannehmbares Maß an Täuschung oder Beteiligung erfordern, um Zugang zu erhalten, was eine Kombination aus Fragmentierung mit ethischen und rechtlichen Hindernissen darstellt.

Zusätzliche Herausforderungen durch Fragmentierung bei der Bekämpfung von schädlichen Inhalten und Verhaltensweisen

Selbst wenn schädliche Inhalte und Verhaltensweisen auf einer Online-Plattform aufgedeckt und bekämpft werden, gibt es eine Vielzahl anderer Plattformen, auf denen sie sich weiter ausbreiten können, wenn die Nutzer:innen innerhalb des Online-Ökosystems weiterziehen. Das ist ein seit langem bestehendes Problem bei der Bekämpfung schädlicher Online-Aktivitäten, zu dessen Lösung bestimmte Maßnahmen entwickelt worden sind. Bei der Entfernung von Inhalten im Zusammenhang mit Kindesmissbrauch und Terrorismus ist beispielsweise das sogenannte „Hashing" eingesetzt worden, bei dem Bilder und Videos mit einer eindeutigen Kennung versehen werden, um Kopien eines bereits gesperrten Bildes leichter aufspüren und erneut zu entfernen zu können. Die Hashing-Technologie wird von Organisationen wie dem Global Internet Forum to Counter Terrorism[21] und der Internet Watch Foundation[22] eingesetzt.

Doch selbst mit solchen Instrumenten bleibt die vollständige Entfernung schädlicherInhalte aus dem Internet eine große Herausforderung.Wenn die genaue Form des Inhalts variiert oder sich weiterentwickelt anstatt direkt kopiert zu werden, kann das Aufspüren und Entfernen ähnlicher oder verwandter Inhalte sogar noch schwieriger werden. So wäre ein bestimmtes kopiertes Bild oder Video wie etwa. das ursprüngliche, per Livestream übertragene Videomaterial des Terroranschlags in Christchurch oder das virale „Plandemic"-Video[23] leichter zu identifizieren, als bearbeitete Versionen davon oder Inhalte, die ein ähnliches Narrativ propagieren wie z. B. zusätzliche Originalinhalte, die den Anschlag in Christchurch verherrlichen oder die Desinformationen der Impfgegner unterstützen. Die durch die Fragmentierung bedingten Probleme bei der Identifizierung relevanter Inhalte könnten sich noch verschärfen, wenn bearbeitete oder ähnliche Inhalte in großem Umfang über eine Reihe verschiedener Plattformen verbreitet werden, die nicht schnell und systematisch durchsucht werden können.

Kapitel 3: Methoden und Tools

Nachdem wir die potenziellen Hindernisse dargelegt haben, betrachten wir nun, wie die von Forscher:innen in Online-Räumen üblicherweise verwendeten Methoden darauf reagieren könnten. Wir beginnen mit der Vorstellung von drei Methoden unter Bezugnahme auf die bestehende Forschung und Literatur. Anschließend stellen wir diese Methoden unseren drei Arten von Hindernissen gegenüber, um die Stärken und Schwächen der einzelnen Methoden bei der Bewältigung der jeweiligen Hindernisse herauszuarbeiten. Es gibt nur sehr wenige Fälle, in denen eines der im vorigen Kapitel genannten Hindernisse die Forschung auf einer bestimmten Plattform vollständig verhindert; sie können jedoch das Spektrum der möglichen Methoden und Tools stark einschränken. Neben der Überprüfung bestehender methodischer Ansätze haben wir auch ein Scoping durchgeführt, um bestehende Tools zum Auffinden und Sammeln von Inhalten auf Alt-Tech-Plattformen zu ermitteln. Im letzten Teil dieses Kapitels stellen wir die Ergebnisse unseres Scoping-Verfahrens vor und zeigen die Möglichkeiten und Grenzen der ermittelten Analysetools auf.

Methode 1: Systematische Suche

Bei dieser Methode werden technologiegestützt große Mengen von Daten und Metadaten direkt aus Online-Plattformen extrahiert. Digitale Technologien haben den Umfang und den einfachen Zugriff auf Kommunikationsdaten erheblich erweitert. Verschiedene seit langem bestehende Methoden – von Copy-Paste bis Web-Scraping – haben es Forscher:innen ermöglicht, Online-Daten in leicht zu analysierende Formen umzuwandeln. Zu den Daten gehören beispielsweise die Inhalte von Online-Texten, Verbindungen zwischen Online-Konten sowie Metadaten wie z. B. die Zeit oder der geografische Standort von Postings.

Die zunehmende Dominanz von Web-2.0-Plattformen (die darauf ausgelegt sind, nutzergenerierte Inhalte und die Beteiligung von Nutzer:innen zu fördern), einschließlich Social-Media-Plattformen, hat die Menge dieser Daten beträchtlich erweitert. In den 2000er Jahren konnten Forscher:innen persönliche Beziehungen nachverfolgen, indem sie etwa beobachteten, wie oft verschiedene Mitglieder eines Online-Forums einander antworteten. In den 2010er Jahren konnten Forscher:innen auf Plattformen wie Facebook umfangreichere „Freundschaftsbeziehungen" zwischen deutlich größeren Gruppen beobachten. Viele Social-Media-Plattformen erleichterten auch den Zugriff auf die Daten, indem sie APIs bereitstellten, die es Forscher:innen ermöglichten, direkt auf verschiedene Arten von Daten von Plattformen zuzugreifen, ohne dass sie ihren eigenen Code von Grund auf neu erstellen mussten.[xiii] Die Entwicklung von KI-basierten Ansätzen hat immer ausgefeiltere Analysemethoden ermöglicht. So wird beispielsweise die Natural Language Processing (NLP) zunehmend eingesetzt, um Trends, Meinungen und Namen in großen Mengen von Online-Texten zu erkennen.

Ein großer Teil der modernen Forschung über Online-Plattformen nutzt technologische Ansätze, um Daten zu finden und zu sammeln. Zu den beliebtesten Tools gehören Google Search, die Twitter-API oder CrowdTangle für Facebook und Instagram. Weitere Technologien wurden von externen Forscher:innen konzipiert. In der von CASM entwickelten Method52 Software werden beispielsweise Daten von mehreren Online-Plattformen[xiv] gesammelt und integriert, Beziehungen zwischen Konten und Inhalten abgebildet und „Klassifikatoren" trainiert, um von Forscher:innen definierte Themen im Text zu unterscheiden. Die Digital Methods Initiative (DMI) bietet ebenfalls ein Portfolio von Tools, die von Wissenschaftler:innenn entwickelt wurden.[24]

Die wichtigsten Vorteile der systematischen Suchwerkzeuge:

- **Geschwindigkeit und Umfang**: Forscher:innen können in Sekundenschnelle Milliarden von Datenpunkten finden, sammeln und abfragen.
- **Systematizität**: Zwar bietet kein Tool einen unvoreingenommenen Einblick in 100 % der Online-Daten, aber die kontrollierbare und quantitative Natur dieser Technologien ermöglicht es, Daten systematisch zu erfassen und zu vergleichen und möglicherweise zu reproduzieren.
- **Präzision**: Forscher:innen, die Abfragetechniken (z. B. boolesche Operatoren) beherrschen, können eine Suche auf genau definierte Inhalte fokussieren;

xiii APIs haben den Plattformen auch ein größeres Maß an Kontrolle über die von ihnen bereitgestellten Daten verschafft, was Bedenken hinsichtlich der Transparenz und der Stabilität von API-gestützten Tools aufkommen lässt.

xiv Derzeit acht Plattformen sowie externe Datensätze, Formate und Quellen (z. B. Media Cloud, Mastodon, RSS Feeds und Google Sheets).

KI-basierte Technologien steigern diese Fähigkeit noch weiter. Dies ist angesichts der Menge an Online-Daten, mit denen Forscher:innen häufig umgehen müssen, äußerst wertvoll.

Die Nachteile:

- **Datenverfügbarkeit**: Die Forschung kann von den verfügbaren Daten geleitet werden, anstatt von einem Forschungsproblem ausgehend nach den am besten geeigneten Daten zu suchen. Vor allem Twitter hat im Verhältnis zur Größe und Vielfalt seiner Nutzerbasis einen übergroßen Anteil an der Forschungsaufmerksamkeit erhalten, was wohl auf die Bandbreite der Daten zurückzuführen ist, die es Forscher:innen im Vergleich zu großen Plattformen wie Facebook, Instagram und insbesondere TikTok zur Verfügung stellt.
- **Genauigkeit**: Forschung, die sich auf offizielle APIs stützt, ist davon abhängig, dass die Plattformen kontinuierlich Zugriff auf genaue Daten gewähren. Plattformen haben möglicherweise jedoch keinen Anreiz, vollständige und genaue Daten bereitzustellen, zudem ist es oft schwer, unabhängig zu überprüfen, ob sie dies auch tun. Dasselbe Problem gilt auch für Datensätze wie Social Science One, die in Zusammenarbeit mit Technologieunternehmen erstellt wurden, um externen Forscher:innen den Zugriff zu ermöglichen. Diese sind aber mit einer Reihe von Problemen behaftet, darunter die Genauigkeit der bereitgestellten Daten und der US-Fokus der Forscher:innen, denen der Zugriff gewährt wird.[25] Die Abhängigkeit von Unternehmen bei der Zugriffsgewährung für legitime Forschungsarbeiten von öffentlichem Interesse kann Forscher:innen auch davon abhalten, Unternehmen öffentlich zu kritisieren, wenn ihre Ergebnisse Mängel in den Praktiken dieser Unternehmen aufzeigen. Schließlich ist es für Außenstehende manchmal möglich, Alternativen zu APIs zu schaffen, die jedoch eventuell gegen die Geschäftsbedingungen der Plattformen verstoßen und die Forscher:innen daher potenziellen rechtlichen Risiken aussetzen.[xv]
- **Rechtliche Risiken**: Alternativen von Drittanbietern zu APIs verstoßen im Zweifel gegen die Geschäftsbedingungen der Plattformen und setzen Forscher:innen damit potenziellen rechtlichen Risiken aus.
- **Technologisches Wettrüsten**: Mit der zunehmenden Diversifizierung der Online-Plattformen, die immer komplexere Strukturen, Metriken und Medientypen umfassen, könnte es schwieriger werden, Tools zu entwickeln, die auf die gesamte Bandbreite potenziell relevanter Daten zugreifen und diese plattformübergreifend vergleichen können. Forscher:innen, die über die notwendigen finanziellen Mittel und technologischen Fähigkeiten verfügen, könnten andere Forscher:innen, denen eines oder beides fehlt, überflügeln, was zu Ungleichheiten im Forschungsbereich und Ungleichgewichten in der Evidenzbasis führt.

Methode 2: Ethnografie

Die Ethnografie ist eine etablierte Schule von Forschungsmethoden, die sich durch eine tiefe und anhaltende Beteiligung an einer Gemeinschaft auszeichnet. Anstatt sich auf Technologien zur Datenerhebung zu verlassen, wählen Forscher:innen einen menschlicheren Ansatz, indem sie sich in Online-Räumen als Formen der Gemeinschaft einbringen, daran teilnehmen und sie beobachten.

Die Ethnografie war ein gebräuchlicher Ansatz in der früheren Forschung über Online-Plattformen, einschließlich vieler klassischer empirischer Arbeiten, z. B. von Nancy Baym oder Henry Jenkins.[26] Damit ging eine Zunahme der Literatur und der Forschungsprogramme zur „digitalen Anthropologie" und „digitalen Ethnografie" einher. Auch wenn die Ethnografie heute weniger im Vordergrund steht als systematische Suchansätze, so ist sie doch nach wie vor ein blühendes Forschungsfeld.

Die wichtigsten Vorteile der ethnografischen Forschungsmethoden:

- **Kontextabhängigkeit**: Die Ethnografie kann ein umfassendes, kontextspezifisches Verständnis der Online-Aktivitäten vermitteln.
- **Begrenzte Datenmenge**: Die Methode eignet sich für die Untersuchung von Nischensubkulturen, die ein Eintauchen erfordern und nicht die größeren Mengen an relevanten Daten liefern, die für

xv Siehe Anhang: Rechtliche Risiken.

- **Alternative Inhaltsformen**: Bei der ethnografischen Forschung können audiovisuelle Inhalte untersucht werden, die mit den üblichen technologischen Hilfsmitteln nicht leicht zu analysieren sind.
- **Anfälligkeit**: Sie ist weniger anfällig, wenn Plattformen sich dafür entscheiden, die Forschungstools einzuschränken z. B. durch die Einschränkung der über APIs verfügbaren Daten.

Die Nachteile:

- **Schwer zu skalieren**: Die intensive Auseinandersetzung mit einer Gemeinschaft eignet sich nicht für die Untersuchung mehrerer Plattformen.Eine Person kann zudem nicht so viele Daten analysieren, wie technologische Tools es können.
- **Weniger systematisch**: Die Ethnografie kann zwar ein vertieftes Verständnis spezifischer Gemeinschaften vermitteln, bietet aber keinen systematischen Überblick über breitere Online-Aktivitäten.
- **Ethische Bedenken**: Ethnografische Forschung in geschlossenen Räumen kann ein gewisses Maß an Täuschung oder Nachahmung erfordern, insbesondere bei der Erforschung von verschlossenen Gemeinschaften wie gewalttätigen extremistischen Gruppen. Außerdem können die Forscher:innen direkt mit schädlichem Material oder potenziellen Sicherheitsrisiken konfrontiert werden.

Methode 3: Crowdsourcing und Umfragen

Zwei weniger verbreitete, aber potenziell wertvolle Methoden zur Erforschung schädlicher Inhalte und Verhaltensweisen sind Crowdsourcing und Umfragen. Bei den Crowdsourcing-Methoden melden die Nutzer:innen von Online-Plattformen den Forscher:innen freiwillig bestimmte Arten von Inhalten. Solche Meldemechanismen können verschiedene Formen annehmen, z. B. Plug-ins[27] oder Meldeformulare für Nutzer:innen, die entweder von Dritten oder von den Online-Diensten selbst angeboten werden. Ein aktuelles Beispiel ist der Einsatz von „Tiplines" zur Meldung von Des- bzw. Misinformationen in WhatsApp-Chats während der indischen Wahlen 2019.[28]

Derzeit sind Crowdsourcing-Methoden noch relativ neu, aber ihre Verbreitung auf Plattformen wie WhatsApp könnte weitere Aufmerksamkeit erregen. Freiwillig von Nutzer:innen gemeldete schädliche Inhalte können auch zur Erstellung von Datenbanken verwendet werden, die bei der Erforschung oder Verhinderung schädlicher Online-Aktivitäten helfen. GIFCT ist eine plattformübergreifende Initiative, die eine Hashing-Datenbank mit Fingerabdrücken von bekanntem Propagandamaterial von Organisationen unterhält, die von den Vereinten Nationen als terroristisch eingestuft wurden.[29] Datenbanken mit gewalttätigen Inhalten können auch dazu verwendet werden, Beweise für potenzielle Kriegsverbrechen zu sichern, selbst wenn diese Inhalte von Plattformen entfernt worden sind, weil sie gegen deren Richtlinien verstoßen. Solche Archive zur Sammlung und Untersuchung gibt es für die Kriege in Syrien[30] und im Jemen.[31]

Eine verwandte Methode für die freiwillige Meldung schädlicher Inhalte ist die Befragung von Internetnutzer:innen zu ihren Erfahrungen. Dieser Ansatz wurde von nationalen Kommunikationsregulierungsbehörden wie dem Office of Communications (Ofcom) im Vereinigten Königreich angewandt.[32] Bei Remote-Usability-Studien zur Benutzerfreundlichkeit gewähren die Nutzer:innen den Forscher:innen Zugang zu ihren Geräten, um ihr digitales Verhalten zu beobachten. Solche Tests können moderiert sein, d. h. alle Teilnehmer:innen nutzen den analysierten Dienst gleichzeitig und können mit den Forscher:innen kommunizieren, oder unmoderiert, d. h. die Nutzer:innen können ihre Sitzungen jederzeit aufzeichnen und die Aufnahmen später einsenden.[33]

Wissenschaftler:innen und Forschungsinstitute haben ähnliche Erhebungen durchgeführt, um die Auswirkungen der Internetnutzung auf die Einstellungen und Verhaltensweisen der Nutzer:innen zu untersuchen. Gil de Zúñiga und Goyanes haben zum Beispiel Daten aus einer Zwei-Wellen-Panel-Umfrage in den USA verwendet, um (vielleicht kontraintuitiv) zu argumentieren, dass Personen, die ihre Nachrichten verstärkt über WhatsApp konsumieren, dazu neigen, weniger über Politik zu wissen und sich eher an illegalen politischen Protestaktionen zu beteiligen.[34] In einer Umfrage des Center für Monitoring, Analyse und Strategie (CeMAS), einer unabhängigen Forschungseinrichtung in Deutschland, haben Forscher:innen einen Zusammenhang zwischen der Häufigkeit der Nutzung der verschlüsselten und bei Anhänger:innen von Verschwörungsideologien sehr

Protesten gegen COVID-19-Beschränkungen festgestellt.[35] Diese Art der Umfrageforschung ist zwar in erster Linie darauf ausgerichtet, die Auswirkungen der Internetnutzung zu messen, kann aber auch verwendet werden, um mehr über die Reichweite schädlicher Inhalte und Narrativen herauszufinden. Im Jahr 2020 unterstützte das ISD beispielsweise eine von der Tufts University durchgeführte Umfrage über die Verbreitung von QAnon-bezogenen Überzeugungen in der US-amerikanischen Bevölkerung.[xvi]

Die wichtigsten Vorteile von Crowdsourcing und Umfragen als Forschungsmethoden:

- **Sie vereinen die Vorteile der systematischen Suche und der Ethnografie**: Crowdsourcing und Umfragen finden Daten durch menschliche Beteiligung, nicht durch plattformspezifische Abfragen und sind daher weniger anfällig für z. B. API-Beschränkungen. Aber sie finden auch Daten über eine größere Bandbreite an Material als ethnografische Methoden.
- **Personalisierung**: Diese Forschungsmethoden geben Einblicke in die personalisierten Erfahrungen der Nutzer:innen sozialer Medien. Da algorithmische Systeme auf Grundlage des bisherigen Verhaltens der Nutzer:innen unterschiedliche Ergebnisse erzeugen, ermöglicht dieser Ansatz den Forscher:innen, Einblicke in ein breiteres Spektrum von Nutzererfahrungen zu gewinnen.
- **Einfluss**: Indem sie den Forscher:innen die Möglichkeit geben, über die beschreibende Verfolgung der Online-Dynamik hinauszugehen, können diese Methoden die Auswirkungen schädlicher Inhalte und Verhaltensweisen im Internet auf breitere politische Einstellungen und Verhaltensweisen messen. Insbesondere können Umfragen Erkenntnisse über das Publikum und nicht nur über die Produzent:innen von Inhalten liefern.

Die Nachteile:

- **Genauigkeit**: Da die Daten von einer Vielzahl von Akteur:innen stammen, die sich in Bezug auf Sorgfalt, Kenntnisse oder Aktivitätsniveau unterscheiden können, ist es schwierig, die Systematik, Zuverlässigkeit und Genauigkeit der Eingaben zu gewährleisten.
- **Weitergabe**: Diese Forschungsmethoden sind darauf angewiesen, dass die Gruppenteilnehmer:innen Informationen außerhalb der Gruppe weitergeben. Dies kann ethische Bedenken aufwerfen, und die Rekrutierung von Teilnehmer:innen kann in bestimmten Gruppen schwieriger sein (z. B. sind Mitglieder rechtsextremer Gruppen möglicherweise nicht bereit, mit Forscher:innen zusammenzuarbeiten, die sich kritisch über Rechtsextremismus geäußert haben).
- **Begrenzte Plattformgröße**: Es wird wahrscheinlich schwierig sein, die Nutzer:innen kleinerer, nischenorientierter Plattformen systematisch zu befragen. Da auch die Nutzerbasis kleiner ist, könnte es problematisch werden, die Nutzer:innen dieser Plattformen zu identifizieren. Möglicherweise sind sie auch nicht bereit sind, an der Forschung teilzunehmen.
- **Rechtliche Risiken**: Bestimmte Crowdsourcing-Methoden sind mit rechtlichen Risiken verbunden. So könnte die Verwendung von Technologien Dritter (z. B. Internetbrowser-Erweiterungen oder Plug-ins) gegen die Nutzungsbedingungen der Plattformen verstoßen.[36]
- **Technologische Bedenken**: Es kann ein größeres technisches Fachwissen und höhere Kosten erfordern, technische Tools zu entwickeln und zu betreiben bzw. professionelle Meinungsforschungsinstitute zu beauftragen.

xvi QAnon ist eine weitreichende Verschwörungsideologie, die behauptet, dass eine elitäre Gruppe von Kinderhändler:innen und Pädophilen die Welt seit Jahrzehnten beherrscht. Siehe „Survey on QAnon and Conspiracy Beliefs", Tufts University and Institute for Strategic Dialogue, September 2020, https://www.isdglobal.org/wp-content/uploads/2020/10/qanon-and-conspiracy-beliefs-full_toplines.pdf.

Methoden vs. Hindernisse

Die folgende Gegenüberstellung gibt einen Überblick über die Anwendbarkeit der einzelnen Methoden in Bezug auf die einzelnen Hindernisse sowie über alle weiteren Fragen.

Hindernis Forschungs-methode	Technologie	Ethik und Recht	Fragmentierung
Systematische Suche	Durch eine umfassende und kontinuierliche Datensammlung können frühzeitig Beispiele für neu entstehende Plattformen und Technologien entdeckt werden. Die Technologien selbst könnten Hindernisse für einen groß angelegten systematischen Datenzugriff darstellen (siehe Diskussion in der Spalte „Fragmentierung").	Datenschutz- und rechtliche Bedenken schränken die Nutzung groß angelegter Datenerhebungen zunehmend ein, ohne dass die Nutzungsbedingungen der Plattformen verletzt werden.[xvii] Es gibt Möglichkeiten, einen umfangreichen Datenzugriff umzusetzen und gleichzeitig den Datenschutz zu wahren, z. B. durch den „differentiellen Datenschutz", bei dem Rauschen in die Daten eingebracht wird, um die wahre Identität zu verschleiern. Viele Forscher:innen sind besorgt, dass die derzeitigen Methoden keine genauen Ergebnisse liefern, insbesondere bei der Erforschung bestimmter Inhalte (z. B. schädlicher Inhalte). Diese Methoden sind jedoch relativ neu, und es gibt noch Raum für weitere Entwicklungen.[37]	Die systematische Suche war bisher die Methode zur Überwindung von Fragmentierungshindernissen. Ob dies auch in Zukunft der Fall sein wird, hängt von der genauen Form der künftigen Plattformen und der Such- und Monitoringtechnologien ab. Eine zunehmende Fragmentierung über Nischenplattformen und/oder der Verlust von systematischen API-Endpunkten wird den Nutzen der systematischen Suchtechnologie einschränken. Neue Entwicklungen in der KI-gestützten Suche könnten es ermöglichen, die systematische Suche an diese Veränderungen anzupassen. Dennoch könnte es weiterhin ethische Probleme geben in Bezug auf die Frage, ob Plattformen diese Art des Datenzugriffs zulassen.
Ethnografie	Potenziell eine wirksame Methode zur Überwindung technologischer Hindernisse; einer Gemeinschaft anzugehören, ermöglicht den Forscher:innen, sich gemeinsam mit anderen Teilnehmer:nnen an neue Technologien anzupassen. Diese Methode kann den Forscher:innen auch Frühwarnungen und Einblicke in die Entwicklung neuer Technologien geben.	Eine intensive und langfristige Einbindung in eine Gemeinschaft kann dazu beitragen, potenzielle ethische Bedenken zu zerstreuen (z. B. fühlen sich die Teilnehmer:innen wohler, wenn sie das Gefühl haben, dass die Forscher:innen auch Mitglieder der Gemeinschaft sind). Umgekehrt kann eine intensive, langfristige Einbindung auch ethische Probleme verschärfen, wenn z. B. ein Abschlussbericht den Erwartungen der Gemeinschaft zuwiderläuft, Forscher:innen detaillierte und persönliche Informationen weitergeben oder die Forschung auf einem Vertrauensverhältnis beruhte. Bei Forschungsarbeiten zu schädlichen Inhalten oder Verhaltensweisen ist dieses negative Szenario wahrscheinlicher.	Die Ethnografie ist für die Überwindung dieses Hindernisses ungeeignet; die Methode ist schwer zu skalieren und eignet sich im Allgemeinen nicht für die direkte Durchsuchung großer Mengen von Material. Dies ist ein Nachteil gegenüber dem tiefen und kontextbezogenen Verständnis, das der Methode eigen ist.
Crowdsourcing	Wie die ethnografischen Forschungsmethoden zeigen, können sich menschliche Teilnehmer:innen an neue Technologien anpassen. Sie könnten die Forscher:innen auch zu frühen Beispielen für neue Technologien und Plattformen führen. Wenn möglich, sollten die Teilnehmer:innen geschult werden, um ihr Verständnis für die relevanten Plattformen und technologischen Entwicklungen zu verbessern.	Wenn Crowdsourcing auf bestehende Teilnehmer:innen von Online-Gemeinschaften zurückgreift, gibt es potenzielle ethische Grauzonen, wenn es darum geht, die informierte Zustimmung anderer Teilnehmer:innen einzuholen, die nicht an der Forschung beteiligt oder darüber informiert sind; solange jedoch keine sensiblen persönlichen Daten weitergegeben werden, kann Crowdsourcing ethisch vertretbar sein. Das möglicherweise unzureichende Verständnis der Teilnehmer:innen für Fragen des Datenschutzes könnte zu einer übermäßigen Weitergabe von Daten führen, was ethische und ggf. rechtliche Probleme nach sich ziehen könnte. Bei der Verwendung „eingeschleuster" Teilnehmer:innen ergeben sich ähnliche Probleme wie bei der Ethnografie.	Crowdsourcing in großem Maßstab ermöglicht das Monitoring einer Vielzahl von Plattformen durch eine Vielzahl von menschlichen Beobachtern und kann daher gut geeignet sein, um Probleme der Fragmentierung zu lösen. Bei einem solchen Crowdsourcing stellen sich Fragen der Systematizität, Zuverlässigkeit und Skalierung.

xvii Es ist anzumerken, dass Plattformen andere, eher eigennützige Anreize für die Einschränkung des Datenzugriffs haben können. Die Einschränkung des Datenzugriffs für Forscher:innen und Journalist:innen verringert die Transparenz und damit das Risiko, die Versäumnisse der Plattformen beim Schutz ihrer Nutzer:innen und der breiteren Gesellschaft vor Online-Schäden aufzudecken, sowie die Rolle, die ihre Produkte und Geschäftsmodelle bei der Verschlimmerung oder Verstärkung dieser Schäden spielen können.

Tools

Im folgenden Kapitel werden die Ergebnisse aus unserem Scoping-Verfahren vorgestellt, das darauf abzielt, Analysetools für Alt-Tech-Plattformen zu identifizieren. Zwar gibt es nur wenige Analysetools für diese Plattformen, doch wurden im Laufe der Jahre einige Tools entwickelt, die einen systematischen Zugriff auf Inhalte und/oder die Ermittlung breiter Metriken (z. B. Follower:innen, Aufrufe und Veränderungen im Laufe der Zeit) ermöglichen oder die zur Unterstützung manueller Forschungsarbeiten verwendet werden können.

Tools zur Analyse sozialer Medien ermöglichen den Zugriff auf Daten sowie das Monitoring und die Analyse von Online-Diskursen, Trends und Verhaltensweisen. Diese Tools werden von Marketingfachleuten, politischen Parteien, Sicherheitsdiensten, Regierungsbehörden und Forscher:innen für eine Vielzahl von Zwecken eingesetzt.

Einige Tools sind zwar frei verfügbar, aber es gibt erhebliche Unterschiede im Hinblick auf den Datenzugriff und die Transparenz der Methoden und Technologien, die zur Sammlung, Analyse und Präsentation der Erkenntnisse verwendet werden. Die meisten der weit verbreiteten Tools sammeln öffentlich zugängliche Daten von großen Social-Media-Plattformen wie z. B. Twitter (Brandwatch), Reddit (Pushshift) oder Facebook und Instagram (CrowdTangle). Doch während einige Tools eine Schlüsselwörter-basierte Suche auf der gesamten Plattform ermöglichen, beschränken andere die Suche auf bestimmte Kanäle, Konten oder Interessengemeinschaften. CrowdTangle, das von Journalist:innen und Forscher:innen häufig genutzt wird, bietet keinen systematischen Zugriff auf Kommentare, sondern nur auf Beiträge von öffentlichen Seiten und Gruppen. Breit angelegte Trenddaten etwa zur Verfolgung von Follower-Zahlen, Likes oder Videoaufrufen sind für die meisten großen Plattformen über Open-Source-Tools wie Social Blade verfügbar. Dazu gehören auch einflussreiche Plattformen wie YouTube oder TikTok, die aufgrund des eingeschränkten Datenzugriffs (TikTok) oder der überwiegend audiovisuellen Inhalte (beide) oft als schwer zu erforschen angesehen werden.

Möglicherweise gibt es aufgrund der geringeren kommerziellen Bedeutung und der größeren technischen Vielfalt der Alt-Tech-Plattformen deutlich weniger Tools zur Verfolgung und Analyse ihrer Inhalte, Trends und Verhaltensweisen. Auf der Grundlage unserer Durchsicht der vorhandenen Literatur, in der die im Rahmen des Plattform-Scopings identifizierten Plattformen erörtert werden, untersuchten das ISD und CASM dreizehn Tools, die über einen gewissen Datenzugriff und in einigen Fällen auch über Analysefunktionen für Alt-Tech-Plattformen verfügen: 4cat, Archived.Moe, Dewey Defend, DISBOARD, Lyzem, Method52, OSINT Combine Alt-Tech Social Search, Social Blade, Telegago, TelegramDB, Tgram.io, TGStat und Unfurl.[38]

Die meisten dieser Tools bieten keinen systematischen Zugriff auf die Daten der jeweiligen Alt-Tech-Plattformen. Nur 4cat, TGStat, Dewey Defend und Method52 ermöglichen einen systematischen Zugriff auf Inhalte und nicht nur auf allgemeine Follower- und View-Metriken oder Account-Profil-Informationen. Außerdem ist von diesen nur 4cat kostenlos und öffentlich zugänglich; es ist ein Open-Source-Analysetool, das vom Open Intelligence Lab (OILab) und dem DMI an der Universität Amsterdam entwickelt wurde. Wie der Name 4cat schon andeutet, ist es darauf spezialisiert, Daten von threadbasierten Plattformen wie 4chan und seit kurzem auch 8kun (früher 8chan) zu sammeln. Es ermöglicht Forscher:innen auch, Datensätze von anderen Plattformen zu erstellen, darunter BitChute (Scraping-Ergebnisse der Videosuchfunktion), Parler, Telegram (basierend auf den Telegram-API-Anmeldeinformationen der Forscher:innen) und Reddit (über die externe Pushshift-Datenbank). Basierend auf der Struktur der von den einzelnen Plattformen gewonnenen Daten stehen in 4cat selbst weitere Analysemodule zur Verfügung, die u.a. die Identifizierung von zusammenhängenden Beiträgen, die sich gegenseitig antworten, die Erkennung von beleidigenden Äußerungen und die Sammlung der beliebtesten Bilder innerhalb eines Datensatzes ermöglichen. 4cat ist relativ einzigartig, da es seit langem Zugriff auf die Daten insbesondere der Chan-Sites hat; je nach Thema können die Daten auf 4chan bis ins Jahr 2012 zurückreichen.

Andere Tools erlauben einen systematischen Datenzugriff nur für Abonnent:innen. Zum Beispiel bietet die öffentliche Version von TGStat hauptsächlich Zugriff auf breite Metriken, die zeigen, wie sich die Abonnentenzahlen und Ansichten für Telegram-Kanäle im Laufe der Zeit verändert haben; die Möglichkeit, nach Beiträgen auf Telegram zu suchen, die Schlüsselwörter enthalten, ist jedoch nur zahlenden Abonnent:innen vorbehalten. Dewey Defend ist ebenfalls nur für lizenzierte Nutzer:innen zugänglich und ermöglicht es ihnen, Inhalte auf einer Vielzahl von

Plattformen zu finden, darunter 4chan, 8kun, BitChute, Gab, Gettr, Kiwi Farms, MeWe, Minds, Parler, Poal und Rumble sowie Telegram-Kanäle, die manuell von Nutzer:innen hinzugefügt wurden.

Neben den wenigen Tools mit systematischem Zugriff auf Inhalte und den Tools, die breite Trenddaten wie Follower-Zahlen, Likes oder Videoaufrufe liefern – wie z. B. TGStat und Social Blade, das neben den großen Plattformen auch Twitch, Odysee und DLive abdeckt –, gibt es eine Reihe von Tools, mit denen Forscher:innen nach bestimmten Inhalten auf verschiedenen Alt-Tech-Plattformen suchen können. TelegramDB und Tgram.io beispielsweise autorisieren Nutzer:innen, Telegram nach Gruppen, Kanälen und Bots zu durchsuchen, während Telegago und Lyzem zusätzlich Beiträge durchsuchen können. Bei einigen dieser Tools ist es schwierig zu sagen, wie die Daten gesammelt werden und wie umfassend sie sind, da die Suchergebnisse unvollständig erscheinen können, wie etwa bei Tgram.io. Es gibt andere Suchtools, die sich auf einzelne Plattformen konzentrieren, z. B. Archived.Moe, mit dem Nutzer:innen alle 4chan-Boards nach Beiträgen durchstöbern (4cat beschränkt sich auf ausgewählte Boards wie /pol/ oder /k/), bzw. DISBOARD, mit dem Nutzer:innen nach Discord-Servern suchen können. OSINT Combine schließlich, ein Unternehmen, das sich auf Schulungen und Software für Open-Source-Intelligence spezialisiert hat, hat ein Tool für die soziale Suche auf Alt-Tech Plattformen entwickelt, mit dem Useer:innen die Möglichkeit haben, nach Beiträgen in Parler, Gab, Minds, BitChute, DLive, Rumble und verschiedenen Boards auf JustPaste.it, WrongThink und 8kun zu forschen. Was andere Alt-Tech-spezifische Open-Source-Intelligence-Tools angeht, so extrahiert Unfurl Informationen aus URLs, einschließlich Zeitstempeln und anderen Domain-Informationen, und verfügt über eine spezielle Funktion zum Parsen von Informationen aus Discord-Links.

Kapitel 4: Auswahl der Plattformen für Phase II der Forschung

In den vorangegangenen Kapiteln haben wir zwischen drei Arten von Hindernissen für die Erforschung von Online-Plattformen unterschieden: technologische, ethische und rechtliche Hindernisse sowie die Fragmentierung:öffentliche und zugängliche Inhalte, die jedoch nicht systematisch durchsuchbar sind. Diese Hürden schließen sich nicht gegenseitig aus, und verschiedene Funktionalitäten innerhalb von Plattformen können zuweilen unterschiedliche Hindernisse für Forscher:innen darstellen. In ähnlicher Weise haben wir drei wichtige methodische Ansätze zur Identifizierung schädlicher Online-Inhalte ermittelt: die systematische Suche, die ethnografische Forschung sowie Crowdsourcing und Umfragen.

Für die Fallstudien in englischer, französischer und deutscher Sprache, die in Phase II dieses Projekts durchgeführt werden, haben wir eine Kombination aus Forschungshindernissen und geeigneten methodischen Ansätzen ausgewählt, um diese Hindernisse anzugehen. Auf der Grundlage des Plattform-Scopings haben wir Plattformen identifiziert, die von Akteur:innen in jedem geografischen Kontext zunehmend genutzt werden (eine Plattform pro Kontext), um Extremismus, Hass oder Desinformationen zu verbreiten. Im Folgenden skizzieren wir einige der Vorteile und wahrscheinlichen Herausforderungen bei der Erforschung dieser Plattformen.

Fragmentierungshindernisse: Discord

Für die Fallstudie, die sich mit dem Vereinigten Königreich befasst, schlagen wir vor, schädliche Inhalte und Verhaltensweisen auf **Discord** zu untersuchen, einer Plattform, die in erster Linie Fragmentierungshindernisse aufweist. Wie viele andere Plattformen, z. B. Reddit und Facebook, bietet Discord eine Reihe von thematischen Gemeinschaften, sogenannte „Server", denen die Nutzer:innen beitreten können, um mit anderen zu chatten. Viele von ihnen chatten im privaten Modus, aber viele tun dies auch öffentlich, obwohl man immer noch einen Benutzernamen benötigt, um beizutreten. Die größten öffentlichen Server können über 100.000 Mitglieder haben[39]. Während viele von ihnen sich mit Spielen oder Anime beschäftigen, gibt es auch einige, die sich sozialen/politischen Diskussionen widmen, von denen einige ausdrücklich Verbindungen zu Gemeinschaften auf 4chan aufweisen.[40]

Auf Reddit, Facebook und anderen Plattformen kann über die API auf Diskussionen in öffentlichen Gruppen zugegriffen werden. Das bedeutet, dass Forscher:innen rasch in der Lage sind, Erwähnungen relevanter Schlüsselwörter (z. B. „Stop the Steal") in einer Reihe von öffentlichen Gruppen zu finden. Eine ähnlich weitreichende Funktionalität steht bei Discord nicht zur Verfügung. Die Möglichkeit, Nachrichten über die API von Discord zu suchen und herunterzuladen, funktioniert nur serverweise.[41] Einige Nutzer:innen haben dies automatisiert, um in großem Umfang zu arbeiten.[42] Es scheint jedoch, dass die Forscher:innen im Voraus wissen müssen, in welchen Kanälen sie suchen möchten. In Anbetracht der großen Anzahl von Kanälen auf Discord sowie der Tatsache, dass Kanäle, die zweifelhafte Inhalte enthalten, manchmal gelöscht und/oder umbenannt werden, kann dies eine systematische Suche erheblich erschweren.[xviii] Das Problem ist nicht, dass die Informationen versteckt sind; sie wären leicht zu finden, wenn die Forscher:innen bereits wüssten, wo sie suchen müssen.

Der API-Client von Discord ermöglicht es Forscher:innen, sich mit einem Server zu verbinden und Kanalnachrichten live zu erfassen sowie historische Nachrichten zu sammeln. Um sich mit einem Server zu verbinden, gibt es zwei Möglichkeiten, wie sich Forscher:innen identifizieren können; beide stellen technische und ethische Herausforderungen dar.

- **Bot-Konto**: Gemäß den Discord-Regeln muss jede Automatisierung über einen Bot-Account ausgeführt werden, um Spamming, Phishing und anderes schädliches Verhalten zu verhindern.[43] Bot-Konten können nicht ohne weiteres Servern beitreten; sie müssen manuell von einem Server-Administrator bzw. einer Server-Administratorin zu einem Server hinzugefügt werden: z. B. von der Person, die den Server erstellt hat, oder von einer anderen Person, die über diese Rechte verfügt. Diese Person hat die Möglichkeit, einen solchen Zugang zu verweigern. Zudem wird der Bot in der Benutzerliste eindeutig als solcher gekennzeichnet, was insbesondere in Gemeinschaften, die sensible Themen diskutieren, Verdacht erregen könnte.

xviii Der Discord-Server „Slippy" (siehe Levin, Nancy, „10 Largest Discord Servers", Largest.org, 18. August 2019, https://largest.org/technology/discord-servers/) scheint beispielsweise durch den Server „Dream World" (https://discord.com/invite/dreamworld) ersetzt worden zu sein, obwohl dies unklar ist.

- **Benutzerkonto**: Es ist möglich, einen Bot hinter einem normalen Benutzerkonto laufen zu lassen – ein sogenannter „Self-Bot". In diesem Fall treten die Forscher:innen den Servern als normale Nutzer:innen bei, etwa mit einem Einladungslink, und der Bot gibt sich anschließend als dieser Nutzer aus. Dieses irreführende Verhalten verstößt gegen die allgemeinen Geschäftsbedingungen von Discord, was eine zusätzliche ethische Herausforderung darstellt. Discord kann diese Konten sperren, wenn sie entdeckt werden. Es ist unklar, ob Discord aktiv Verbindungen überwacht, um Konten zu entdecken, die eine solche Täuschung betreiben, oder ob sich die Plattform darauf verlässt, dass sie von anderen Nutzer:innen wegen verdächtigen Verhaltens gemeldet werden.

Darüber hinaus ist eine der Kernfunktionen von Discord die Kombination von Text- und Sprachkommunikation, etwa damit Spieler:innen gemeinsam spielen und währenddessen chatten können. Ohne den Kontext des Gesprächs ist ein Großteil der Textkommunikation möglicherweise ohne Informationsgehalt.

Ethische Hindernisse: Telegram

Für die Fallstudie, die sich mit Deutschland befasst, schlagen wir vor, schädliche Inhalte und Verhaltensweisen auf **Telegram** zu untersuchen, einer Plattform, die in erster Linie ethische Hindernisse aufweist. Telegram ist eine Messenger-App mit plattformähnlichen Qualitäten, die zu einem wichtigen Online-Raum für Extremist:innen, Verschwörungsideolog:innen und Desinformationsakteur:innen geworden ist. Insbesondere in Deutschland hat sich Telegram zu einem zentralen Knotenpunkt für COVID-19-bezogene Verschwörungsideologien, Desinformation und extremistische Mobilisierung entwickelt.

Telegram ermöglicht mehrere Kommunikationsmodi, darunter One-to-One-Messaging, Gruppenchats, private und öffentliche Kanäle. Die ethischen und bis zu einem gewissen Grad auch technologischen Hindernisse für Forscher:innen variieren daher je nach Art des Kommunikationsmodus, der auf Telegram verwendet wird.

Öffentliche Kanäle können eine unbegrenzte Anzahl von Abonnent:innen haben. Kanaladministrator:innen haben zwar die Möglichkeit, Kommentarbereiche zu aktivieren, aber sie können Telegram auch ausschließlich für die Kommunikation von einer Person zu vielen verwenden. Daher stellen größere öffentliche Kanäle kein besonderes ethisches Problem dar, da es kaum eine begründete Erwartung an den Schutz der Privatsphäre geben kann. Es sollte jedoch beachtet werden, dass die Größe von öffentlich sichtbaren Telegram-Kanälen drastisch variieren kann, was zu Erwartungen an den Schutz der Privatsphäre in kleinen Kanälen führt. Ähnliche Überlegungen gelten für Telegram-Gruppen, die auf 200.000 Mitglieder beschränkt sind. In öffentlichen Gruppen befinden sich möglicherweise Nutzer:innen mit berechtigten Erwartungen an die Privatsphäre, insbesondere wenn es sich um kleinere Gruppen handelt.

Trotz seines Rufs als datenschutzfreundliche Plattform bietet die API von Telegram in der Tat Datenzugriff für alle Kanäle und Gruppen, in denen ein Nutzer bzw. eine Nutzerin registriert ist. Dazu gehören auch historische Daten, die bis zum Zeitpunkt der Einrichtung eines Kanals oder einer Gruppe zurückreichen. Daten, die von Gruppen erfasst werden, enthalten auch einige Informationen über einzelne Gruppenmitglieder.

Der Zugriff auf Inhalte und die „Mitgliedschaft" in Gruppen hängt von der Art der Kommunikation ab. Inhalte in öffentlichen Kanälen und Gruppen sind sichtbar, ohne dass man beitreten muss. Der Zugriff auf die systematischen, historischen Daten aus öffentlichen Kanälen und Gruppen ist jedoch nur für Gruppen- bzw. Kanalmitglieder möglich. Für den Beitritt genügt es oft, einfach auf „Beitreten" zu klicken. Es können aber auch weitere Fragen gestellt werden, um die Aufnahme zu prüfen, was unter Umständen eine Täuschung seitens der Forscher:innen erfordert. Telegram begrenzt die Anzahl der öffentlichen und/oder privaten Gruppen und Kanäle, denen ein Nutzer bzw. eine Nutzerin über eine Telefonnummer beitreten kann, auf 500, was einige praktische Herausforderungen bei der Erforschung der Plattform mit sich bringt.

One-to-One-Messaging und private Gruppen entsprechen dagegen eher der Beschreibung von Telegram als einer Messaging-App vergleichbar mit WhatsApp oder Signal. Diese Formen der Kommunikation werden auch von den extremsten und potenziell gewalttätigen Gruppen als Kommunikations- und Mobilisierungsmittel genutzt. Der Zugriff auf diese Chats wäre wahrscheinlich nicht ohne ein gewisses Maß an Täuschung möglich.

Es ist denkbar, unterschiedliche Methoden zuverwenden oder sogar zu kombinieren, um Teilbereiche von Telegram zu untersuchen. Eine **systematische Suche** nach in öffentlichen Kanälen und Gruppen geposteten Links könnte genutzt werden, um potenziell relevante geschlossene Chats zu identifizieren. Zu berücksichtigen ist, dass Telegram zwar eine ID bereitstellt für die Kanäle/Gruppen, aus denen Inhalte weitergeleitet wurden, diese aber nicht angewendet werden können, um automatisch und systematisch die Namen der relevanten Kanäle zu identifizieren. **Ethnografische Methoden** könnten wiederum genutzt werden, um die Machbarkeit des Zugangs zu diesen geschlossenen und wahrscheinlich risikoreichen Bereichen innerhalb von Telegram zu testen.

Technologische Hindernisse: Odysee

Für die Fallstudie, die sich mit Frankreich befasst, schlagen wir vor, schädliche Inhalte und Verhaltensweisen auf **Odysee** zu untersuchen – einer Plattform, die in erster Linie technologische Hindernisse aufweist. Einige Plattformen, die für schädliche Akteur:innen von Bedeutung sind, weisen technologische Barrieren auf, dergestalt, dass sie den Zugriff auf Daten einschränken oder aufgrund technischer Eigenschaften die Erforschung erschweren. Dezentralisierte und/oder blockchainbasierte Plattformen wie Odysee weisen technologische Hemmnisse auf, die es wert sind, untersucht zu werden, besonders im Zusammenhang mit der zunehmenden Präsenz von Extremist:innen und Verschwörungsideolog:innen auf der Plattform, insbesondere in Frankreich.

Odysee ist eine Video-Hosting-Plattform, die teilweise auf LBRY läuft, einem dezentralen, blockchainbasierten Filesharing-Netzwerk. Die Plattform ist eine der am häufigsten verlinkten Plattformen in unseren Datensätzen französischer und deutscher Extremist:innen und scheint eine zunehmend beliebte, relativ libertäre Alternative zu Video-Hosting-Plattformen mit strengeren Richtlinien zu sein. Die Dezentralisierung macht es schwierig, schädliche Inhalte auf Odysee zu bekämpfen, da die technischen Möglichkeiten der Administrator:innen begrenzt sind, Inhalte oder Aufzeichnungen von Inhalten vollständig zu entfernen und Nutzer:innen zu sperren.

Odysee ist nicht nur dezentralisiert, sondern auch blockchainbasiert und ermöglicht es Urheber:innen, ihre Inhalte zu monetarisieren. Odysee bietet die Möglichkeit der Monetarisierung von Aufrufen, abhängig unter anderem von der durchschnittlichen Betrachtungsdauer, der durchschnittlichen Anzahl der Aufrufe, der Art des Inhalts und dem Engagement, als auch von direkten Spenden und Werbeaktionen für Websites und Apps, für die die Urheber:innen LBRY-Credits erhalten.[44] Nach dem Durchlaufen einer Kryptowährungsbörse können diese in nicht-digitale Währungen umgewandelt werden

Da es sich hierbei um ein relativ unerforschtes Terrain handelt, wäre zu prüfen, ob auf Daten von dezentralen und/oder blockchainbasierten Plattformen systematisch zugegriffen werden kann (**systematische Suche**), welche Daten verfügbar werden und ob bzw. welche zusätzlichen Hindernisse dabei entstehen. Da Odysee über keine öffentliche APIs verfügt, bleibt unklar, ob ein direkter Zugriff auf die Daten der Plattform realisierbar ist. Forscher:innen müssten an dem LBRY-Netzwerk arbeiten, auf dem Odysee aufbaut. Dies könnte den Zugriff auf die Videobibliothek von Odysee gewähren, auch wenn noch unklar ist, ob Kommentare und andere Metadaten enthalten sind. Es könnte sich herausstellen, dass der Zugriff auf nützliche Daten von der Plattform unmöglich ist. Es ist vorstellbar, dass nützliche Daten prinzipiell von Odysee abgerufen werden können, dies aber große Ressourcen benötigt oder Forschungsmethoden erfordert, die aus ethischer Sicht umstritten sind. Daher zielt diese Arbeit zum Teil einfach darauf ab, Probleme für Forscher:innen und Nutzer:innen aufzuzeigen, die dringlicher werden könnten, wenn Odysees Popularität weiter wächst, insbesondere unter Extremist:innen und Verschwörungsideolog:innen.

Kapitel 5: Mögliche Zukunftsszenarien

In den vorangegangenen Kapiteln wurde dargelegt, dass technologische Entwicklungen, ethische Erwägungen und Fragen der Fragmentierung die Erforschung des breiten Ökosystems der Online-Plattformen zunehmend behindern könnten. Um zu veranschaulichen, wie diese Trends zusammenlaufen, stellen wir zwei mögliche Zukunftsszenarien vor: ein pessimistisches und ein optimistisches. Es sei darauf hingewiesen, dass die beiden hier skizzierten Szenarien die Extreme eines Spektrums denkbarer Ergebnisse darstellen; das tatsächliche, künftige Online-Ökosystem und das regulatorische Umfeld könnten durchaus irgendwo dazwischenliegen. Die Ergebnisse werden von Plattform zu Plattform variieren; diese bieten bereits jetzt ein breites Spektrum an unterschiedlichen Funktionen, Affordanzen, Fähigkeiten und Unternehmensphilosophien.[xix] Auf der Grundlage der Ergebnisse dieses Berichts geben wir eine Reihe von ersten Empfehlungen für politische Entscheidungsträger:innen, Regulierungsbehörden, Forscher:innen und Plattformen. Diese Empfehlungen werden in den nächsten Phasen des Projekts überprüft und aktualisiert.

Pessimistisches Szenario

Es entwickelt sich eine Reihe von Plattformen, die entweder aufgrund ihrer ideologischen Haltung, ihres Geschäftsmodells und/oder ihrer technischen Konzeption schädliche Inhalte und Verhaltensweisen fördern. Sie begünstigen nicht nur das Wachstum neuer Narrative, sondern auch neue technologische Entwicklungen, wie z. B. das Erproben davon, wie AR/VR genutzt werden kann, um hochgradig ansprechende Radikalisierungsinhalte zu erstellen oder um emotional intensivere Formen von Online-Missbrauch und -Belästigung zu erleichtern, die sich insbesondere gegen Frauen, Minderheiten und Jugendliche richten.[45]

Diese Räume sind Forscher:innen nur zugänglich, wenn sie vorgeben, Mitglieder extremer Gemeinschaften zu sein. Es werden immer mehr Screening-Technologien eingesetzt, um Identitäten zu überprüfen; oder Forscher:innen müssen bestimmte schädliche Verhaltensweisen nachweisen, bevor sie Zutritt zu einem Online-Raum erhalten. Viele Forscher:innen und, was noch wichtiger ist, Ethikgremien sind nicht bereit, den nun notwendigen Grad an Täuschung oder Beteiligung zu unterstützen, der für eine Mitgliedschaft in solchen Gemeinschaften erforderlich ist. Das Verhältnis von schädlichen Aktivitäten zu verfügbaren Forscher:innen nimmt rapide zu.

Durch die Organisation und/oder die Integration mehrerer Plattformen können schädliche Inhalte aus diesen spezialisierten Bereichen schnell auf Mainstream-Plattformen gelangen, wodurch neue Zielgruppen erreicht und schädliche Auswirkungen weiter verstärkt werden. Die blockchainbasierte Monetarisierung von Inhalten fördert die weitere Verbreitung der ansprechendsten, radikalsten oder schädlichsten Inhalte. Aufgrund des weit verbreiteten Einsatzes von KI und Blockchain-Technologie können Inhalte, sobald sie „in freier Wildbahn" sind, leicht mutieren und sich nicht ohne weiteres zentral kontrollieren oder wirksam moderieren lassen. Während Anti-Hass Aktivist:innen versuchen, vergleichbare Taktiken und Technologien wie die spezialisierten Hass-Akteure zu nutzen, müssen sie feststellen, dass sie ständig versuchen, aufzuholen, und ihre Botschaften nur ein begrenztes Publikum erreichen.

Außerdem gehen die Plattformen diese Probleme weder wirksam an, noch arbeiten sie mit Forscher:innen bzw. mit Strafverfolgungs- oder Regulierungsbehörden zusammen. Vorschriften zur Verbesserung der Online-Sicherheit, zur Erhöhung der Transparenz und zum Zugriff von Regulierungsbehörden und Forscher:innen auf Daten werden von bestimmten Plattformen ignoriert oder abgelehnt, insbesondere von solchen, die in Ländern mit schwächerer Regulierung, Aufsicht oder Durchsetzung angesiedelt sind.[46] Kleinere, aber hochgradig toxische Plattformen, die schädliche Inhalte hosten oder schädliche Verhaltensweisen fördern, entgehen den Gesetzen, die in erster Linie zur Regulierung der größten und marktbeherrschenden Technologieplattformen konzipiert wurden.

Optimistisches Szenario

Die Verbreitung von Plattformen, die vermeintlich der „freien Meinungsäußerung" gewidmet sind, führt zu einer fragmentierten Landschaft von Räumen für schädliche Inhalte, Verhaltensweisen und Gemeinschaften. Der zunehmende Nischencharakter dieser Räume (verschiedene Plattformen für verschiedene Arten von Hass,

xix „Affordanzen" (vom Englischen „affordance", auch „Angebotscharakter", „Aufforderungscharakter", „Anbietung") beschreibt die technologischen Möglichkeiten, die den Nutzer:innen durch das Design und die Funktionalitäten der Plattform geboten werden.

Extremismus und Desinformation) ermöglicht es spezialisierten Forscher:innen, schädliche Inhalte und Verhaltensweisen leicht zu lokalisieren und zu identifizieren. Einige dieser Plattformen stellen zwar Hindernisse für den Beitritt auf, diese sind jedoch nicht zu schwerwiegend, um sicherzustellen, dass neue Mitglieder problemlos beitreten können. Die kontinuierliche Vermarktung neuer Räume bedeutet, dass einschlägige Plattformen durch systematisches Monitoring leicht gefunden werden können. Innergemeinschaftliche Konflikte zwischen Gruppen lassen sich verwenden, um das Durchsickern von Informationen aus privaten Räumen zu fördern, die von Hass-, Extremismus- oder Desinformationsakteur:innen frequentiert werden.

Die derzeitige Situation hält an, in der sich Narrative in spezialisierten Hass-, Extremismus- und Desinformationsräumen entwickeln, bevor sie sich auf Mainstream-Plattformen ausbreiten. Allerdings sind Forscher:innen aus den oben genannten Gründen in der Lage, Gegenmaßnahmen gegen viele Online-Risiken vorzubereiten, bevor sie in Mainstream-Räumen ankommen und dort verstärkt werden. Wirksame Online-Regelungen, die klare Transparenzanforderungen und Mechanismen für den Datenzugriff zu Forschungszwecken festlegen, werden eingeführt und aktiv durchgesetzt. Die Plattformen sind bereit, mit Forscher:innen und Regulierungsbehörden zusammenzuarbeiten. Darüber hinaus führt die Entwicklung der Datenschutz- und Online-Sicherheitsgesetze zu klaren Leitlinien und Anforderungen, wie die Bereitstellung von Daten mit den Belangen des Schutzes der Privatsphäre in Einklang gebracht werden kann. Die Entwicklungen im Bereich des differentiellen Datenschutzes ermöglichen es Forscher:innen, auf umfangreiche Datensätze zuzugreifen, ohne die Privatsphäre zu gefährden. Der Einsatz von Crowdsourcing-Methoden wie z. B. Tiplines nimmt ebenfalls zu, unterstützt durch soziale Medien und Nachrichtenplattformen, die zunehmend reibungslose und ansprechende Techniken entwickeln, um ein solches Verhalten zu fördern.

Forscher:innen und Behörden sind in der Lage, eine Reihe von Narrativen zu verfolgen, wie sie sich insbesondere durch Fortschritte in der KI entwickeln:

- Zunehmend leistungsfähige NLP, insbesondere für audiovisuelle und Live-Inhalte.
- Datenerfassungstechnologien, die sich selbst trainieren und aktualisieren können, um auf die verschiedenen Plattformstrukturen zuzugreifen, bzw. zu agieren, wenn sich diese Strukturen ändern.

Die Blockchain-Technologie entwickelt sich in einer Art und Weise, die standardmäßig Transparenz und Rechenschaftspflicht in den Vordergrund stellt; dadurch kann die Quelle schädlicher Narrative leichter ermittelt werden.

Empfehlungen

Für politische Entscheidungsträger:innen und Regulierungsbehörden

- **Bei der Entscheidung, welche Plattformen in den Anwendungsbereich von Gesetzen fallen sollten, sollten die politischen Entscheidungsträger:innen die Risiken, die von den Plattformen ausgehen, sowie ihre Größe, ihre Funktionen und ihre Nutzerzahlen berücksichtigen**. Wenn ein höheres Risiko dies rechtfertigt, sollten Regierungen angemessene und verhältnismäßige rechtliche Verpflichtungen für kleinere Plattformen mit hohem Risiko einführen, um sicherzustellen, dass sie nicht zu undurchsichtigen Online-Räumen werden, die von schädlichen Aktivitäten dominiert werden, die sich dem Zugriff von Regulierungsbehörden und Forschern entziehen.
- **Die politischen Entscheidungsträger:innen sollten dafür sorgen, dass die anstehenden und künftigen Rechtsvorschriften ausreichende Bestimmungen zur Transparenz der Plattformen und zum Datenzugriff für Regulierungsbehörden und zugelassene externe Forscher:innen enthalten**. Um technologische Hindernisse und die Fragmentierung zu überwinden, sollten Plattformen ermutigt werden, angemessene Schritte zu unternehmen, um einen strukturierten und systematischen Datenzugriff zu ermöglichen. Wenn Plattformen nicht in den Anwendungsbereich von Vorschriften fallen, die sie dazu verpflichten, Forscher:innen Zugriff auf Daten zu gewähren, sollten die politischen Entscheidungsträger:innen gesetzliche Ausnahmen und/oder Schutzmaßnahmen für datenschutzkonforme Forschung im öffentlichen Interesse einführen, um ein besseres Verständnis der Risiken auf diesen Plattformen zu erreichen.
- **Die politischen Entscheidungsträger:innen sollten überlegen, wie die Regulierung von Social-Media-Plattformen und anderen Online-Diensten zukunftssicher gestaltet werden kann**, um die potenziellen Risiken zu berücksichtigen, die von einer Reihe neuer Technologien ausgehen. Die Regulierungen sollten so flexibel gestaltet werden, dass die Regulierungsbehörden sich an neue Formen schädlicher oder illegaler Online-Aktivitäten anpassen können, um sicherzustellen, dass die Regulierung des Online-Ökosystems und ihre Durchsetzung die Risiken abschwächt und nicht einfach verlagert.
- **Die politischen Entscheidungsträger:innen sollten sicherstellen, dass die Regulierung Anreize für „Safety-by-Design"-Ansätze und ethische Designprinzipien im gesamten Technologiesektor schafft und diese fördert**, damit Online-Risiken und potenzielle Schäden bei der Entwicklung neuer Dienste, Plattformen oder Funktionen berücksichtigt werden. Viele der in diesem Bericht hervorgehobenen Plattformen wurden nicht dafür konzipiert, Schäden zu begünstigen, aber es gibt Fälle, in denen Designänderungen dazu beitragen könnten, diese Risiken zu mindern. Es ist wahrscheinlich einfacher, diese Risiken während des Prozesses der Entwicklung und Einführung einer neuen Plattform, eines neuen Dienstes oder einer neuen Funktion zu berücksichtigen, als nachträglich Abhilfemaßnahmen einzuführen, um grundsätzlich unsichere Designentscheidungen auszugleichen.
- **Regierungen und Regulierungsbehörden sollten mit ihren Kollegen auf internationaler Ebene zusammenarbeiten**, um einen uneinheitlichen Flickenteppich von Online-Regulierungen so weit wie möglich zu vermeiden. Ein international uneinheitliches Regulierungsumfeld würde nicht nur den offenen, freien und interoperablen Charakter des globalen Internets untergraben, sondern könnte auch die Versuche untergraben, das Internet sicherer zu machen, indem es Unternehmen und Plattformen ermöglicht wird, sich in Ländern mit der schwächsten oder gar keiner Regulierung niederzulassen. Außerdem sollten sich Regierungen und Regulierungsbehörden abstimmen, um einheitliche Anforderungen für den Datenzugriff zu gewährleisten. Dadurch würde vermieden, dass Unternehmen übermäßig belastet und gezwungen werden, mehrere, voneinander abweichende Prozesse und Systeme einzurichten.

Für Forscher:innen und die Zivilgesellschaft

- **Die Zivilgesellschaft sollte sich weiterhin für digitale Regulierungen einsetzen, die die Online-Menschenrechte schützen und fördern**. Diese Vorschriften sollten ein ausgewogenes Gleichgewicht zwischen den verschiedenen Rechten herstellen,

von der freien Meinungsäußerung über die Privatsphäre bis hin zum Schutz vor Diskriminierung oder Hetze.

- **Zivilgesellschaft, akademische Forscher:innen und Förderer:innen der digitalen Forschung sollten zusammenarbeiten und in die Weiterentwicklung von Forschungsmethoden, Tools und Fachwissen investieren**, um mit der raschen und kontinuierlichen Entwicklung des Online-Ökosystems Schritt halten zu können. Neue Methoden und Tools werden für ein wirksames Monitoring und eine wirksame Erfassung dieser Entwicklung von entscheidender Bedeutung sein, da die Vielfalt und die Anwendungen neuer Technologien weiter zunehmen und damit auch das Spektrum und die Arten von Risiken, die von neuen oder entstehenden Plattformen ausgehen.
- **Die Zivilgesellschaft und akademische Forscher:innen sollten die bestehenden Normen, Grundsätze und Leitlinien für legale, ethische und sichere Online-Forschung weiter überarbeiten und harmonisieren**. Dies ist insbesondere für Online-Räume notwendig, die weder ganz öffentlich noch ganz privat sind, und für neue Technologien wie AR/VR. Die Forscher:innen sollten auch ihre Ressourcen bündeln und Fachwissen gemeinsam nutzen, einschließlich ethischer Leitlinien, um diese zunehmend komplexen rechtlichen, ethischen und sicherheitsrelevanten Herausforderungen zu bewältigen.
- **Die Zivilgesellschaft und akademische Forscher:innen sollten gemeinsame, offene Repositories entwickeln, um potenziell bedenkliche Plattformen und/oder technische Entwicklungen zu erfassen und zu kennzeichnen**. Bestimmten Plattformen wird in der Social-Media-Forschung überdurchschnittlich viel Aufmerksamkeit entgegengebracht; es müssen Crowd-Repositories und Frühwarnsysteme eingerichtet werden, die mehr Plattformen im gesamten Online-Ökosystem umfassen. Dies sollte unter Wahrung der Privatsphäre geschehen, indem beispielsweise keine personenbezogenen Daten auf Inhalts- oder Profilebene gespeichert werden.
- Digitale Regulierungen werden zunehmend in den wichtigsten Ländern eingeführt. **Die Forschungsgemeinschaft und die Zivilgesellschaft sollten eine proaktive Rolle spielen, um Unternehmen und Plattformen dabei zu helfen, ihren Verpflichtungen zur Einhaltung der Vorschriften nachzukommen und Best Practices zu entwickeln** – insbesondere jene Unternehmen und Plattformen, die nur über begrenzte finanzielle bzw. technische Ressourcen oder nur über begrenztes Fachwissen über das breite Spektrum von Online-Risiken und -Schäden verfügen.

Für Plattformen:

- **Die Unternehmen sollten bei der Entwicklung neuer Online-Plattformen und neuer Funktionen für bestehende Plattformen nach dem Prinzip „Safety-by-Design" vorgehen und ethische Designprinzipien anwenden**. Diese Ansätze regen die Entwickler:innen dazu an, während des gesamten Entwicklungsprozesses die potenziellen Risiken und Auswirkungen neuer Arten von Plattformen, Funktionen und neuer Technologien zu berücksichtigen, was letztlich dazu beiträgt, dass Abhilfemaßnahmen eingebaut und nicht nachgerüstet werden. Bei der Entwicklung neuer Plattformen oder Funktionen sollten die Unternehmen so früh und so umfassend wie möglich die Zivilgesellschaft und akademische Expert:innen zu einem breiten Spektrum von Online-Risiken und -Schäden konsultieren. Die Betroffenen sollten auch miteinbezogen werden, insbesondere jene aus unverhältnismäßig stark betroffenen Randgruppen oder Minderheiten.
- **Die Unternehmen sollten Forschung von öffentlichem Interesse in den Nutzungsbedingungen ihrer Plattformen zulassen und proaktiv konstruktive Beziehungen mit der Zivilgesellschaft und den Forschungsgemeinschaften aufbauen**, um potenzielle Risiken und Schäden auf ihren Plattformen zu erkennen, zu verstehen und abzumildern. Die Plattformen sollten auch untereinander zusammenarbeiten, um Best Practices auszutauschen und neue potenzielle Probleme und Lösungen zu identifizieren.
- **Online-Plattformen sollten den Zugriff auf öffentliche Daten über strukturierte APIs und Suchfunktionen ermöglichen und, soweit möglich, den Umfang der verfügbaren öffentlichen Daten erweitern bei gleichzeitiger**

Wahrung der Rechte der Nutzer:innen auf Datenschutz und Sicherheit. Alle Bereiche einer Plattform, die öffentlich sind und/oder von denen die Nutzer:innen eine angemessene Sichtbarkeit erwarten können, sowie alle Formen von textlichen und audiovisuellen Inhalten, die in diesen Online-Räumen gehostet werden, sollten transparent und für datenschutzkonforme Forschung von öffentlichem Interesse zugänglich sein – sowohl sowohl in Bezug auf Echtzeit- als auch auf historische Daten. Der Datenzugriff sollte so weit wie möglich konsistent bleiben, damit langfristige Studien nicht durch Änderungen oder Einschränkungen beim Zugriff negativ beeinträchtigt werden.

Zusammenfassung

Wie im vorangegangenen Kapitel beschrieben, wird in Phase II dieses Projekts angewandte Forschung durchgeführt, um verschiedene methodische Ansätze auf drei Plattformen zu erproben. Ziel ist es, die verschiedenen Hindernisse, die wir in diesem Bericht identifiziert haben, zu überwinden und die Erkenntnisse des Forschungsbereichs darüber zu erweitern, welche Methoden bei den bestehenden Formen des Datenzugriffs anwendbar sind. Die Erprobung neuer Ansätze wird es uns auch ermöglichen, über die drei Arten von Forschungshindernissen nachzudenken, die wir hier identifiziert haben: technologische, ethische und rechtliche Hindernisse sowie Fragmentierung. Sie gilt es bei Bedarf zu aktualisieren oder zu ergänzen. Diese Fallstudien werden zusammen mit dem in dem Bericht vorgenommenen Plattform-Scoping bei der Entwicklung praktischer Lösungen für den Zugang zu und die Analyse des ständig wachsenden und vielfältigen Angebots an Online-Plattformen dienen.

Die aus dieser Forschung gewonnenen Erkenntnisse werden in Phase III des Projekts einfließen, in der es darum geht, praktische, technische und regulatorische Lösungen für den Datenzugriff und die Transparenz für diese Arten von Online-Räumen zu finden, ohne die Rechte der Nutzer:innen zu verletzen. Wir werden unsere Ergebnisse mit einschlägigen Forschungsexpert:innen und Vertreter:innen von Technologieunternehmen austauschen und diskutieren, die sich mit der Bereitstellung von Daten und Transparenz befassen. Wir hoffen darüber hinaus, ein breiteres Gespräch mit anderen Forscher:innen anzuregen, damit sie Empfehlungen und Lehren aus ihren eigenen Erfahrungen mit den von uns ermittelten Hindernissen sowie den damit verbundenen Grenzen und Hürden geben können. In dieser Phase des Projekts treten wir auch mit Regierungen und politischen Entscheidungsträger:innen in Kontakt, um unsere Erkenntnisse zu teilen über das sich entwickelnde Online-Ökosystem und die Herausforderungen, Bedrohungen und Chancen, die diese fortlaufende Entwicklung für den Datenzugriff und die Datentransparenz mit sich bringt. Dabei beschäftigen wir uns auch mit den Auswirkungen auf die Online-Sicherheit und die regulatorischen und nicht-regulatorischen Ansätze der digitalen Politik.

Schließlich werden wir in den kommenden Phasen des Projekts auch die oben skizzierten potenziellen Szenarien und Empfehlungen überprüfen, um die Bandbreite der künftigen Möglichkeiten für das Online-Ökosystem und das regulatorische Umfeld zu analyiseren. Dabei stellt sich auch die Frage, wie Forscher:innen, politische Entscheidungsträger:innen und Plattformen auf diese Veränderungen reagieren sollten. Wir beziehen in diese Szenarien die Ergebnisse und Lehren aus unseren bevorstehenden Forschungsarbeiten, Beiträge anderer Forscher:innen und Experten:innen aus den Bereichen Politik und Datenschutz sowie weitere technologische bzw. regulatorische Entwicklungen mit ein.. In den letzten zehn Jahren hatten Forscher:innen und politische Entscheidungsträger:innen im Bereich der digitalen Technologien allzu oft Mühe, mit den raschen und gewaltigen Veränderungen Schritt zu halten, die wir im Internet beobachten konnten, sowie mit den Auswirkungen, die diese auf unsere Rechte, Gesellschaften und Demokratien hatten. Wir hoffen, dass dieses Projekt einen zukunftsweisenden Beitrag leisten kann, um sicherzustellen, dass wir besser auf das, was auf uns zukommt, vorbereitet sind.

Endnoten

1 https://www.isdglobal.org/isd-publications/researching-the-evolving-online-ecosystem-annex/

2 „Fragen und Antworten: Gesetz über digitale Dienste", Europäische Kommission, 20. Mai 2022, https://ec.europa.eu/commission/presscorner/detail/de/QANDA_20_2348.

3 „The Draft Online Safety Bill and the legal but harmful debate", UK Parliament, 24. Januar 2022, https://publications.parliament.uk/pa/cm5802/cmselect/cmcumeds/1039/report.html.

4 „Gemeinschaftsstandards", Facebook, https://transparency.fb.com/de-de/policies/community-standards/; „WhatsApp verantwortungsvoll nutzen", WhatsApp, https://faq.whatsapp.com/general/security-and-privacy/how-to-use-whatsapp-responsibly/; „Gemeinschaftsrichtlinien", Instagram, https://www.facebook.com/help/instagram/477434105621119; „Kommunikationsrichtlinien", Google, https://about.google/community-guidelines/; „Community-Richtlinien", YouTube, https://www.youtube.com/intl/ALL_de/howyoutubeworks/policies/community-guidelines/; „Regeln", Twitter, https://help.twitter.com/de/rules-and-policies/twitter-rules; „Community-Richtlinien", TikTok, https://www.tiktok.com/community-guidelines; „Code of Conduct", Microsoft, https://answers.microsoft.com/de-de/page/codeofconduct. Für einen Überblick darüber, wie sich diese im Laufe der Zeit auf Facebook, Instagram, Twitter und YouTube entwickelt haben, siehe Katzenbach, Christian et al., The Platform Governance Archive, Alexander von Humboldt Institute for Internet and Society, 2021, https://doi.org/10.17605/OSF.IO/XSBPT.

5 Scrivens, Ryan et al., „Examining Online Indicators of Extremism in Violent Right-Wing Extremist Forums", Studies in Conflict & Terrorism, 2021, https://doi.org/10.1080/1057610X.2021.1913818.

6 Goldman, Eric, „Content Moderation Remedies", 28 Michigan Technology Law Review 1, Santa Clara Univ. Legal Studies Research Paper, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3810580#.

7 Kreißel, Philip et al., „Hass auf Knopfdruck. Rechtsextreme Trollfabriken und das Ökosystem koordinierter Hasskampagnen im Netz", Institute for Strategic Dialogue und ichbinhier, 2018, https://www.isdglobal.org/wp-content/uploads/2018/07/ISD_Ich_Bin_Hier_2.pdf.

8 Guerin, Cécile und Fourel, Zoé, „COVID-19 : aperçu de la défiance anti-vaccinale sur les réseaux sociaux", Institute for Strategic Dialogue, 2021, https://www.isdglobal.org/wp-content/uploads/2021/04/COVID-19-aperçu-de-la-défiance-anti-vaccinale-sur-les-réseaux-sociaux.pdf.

9 O'Connor, Ciarán, „The Conspiracy Consortium: Examining Discussions of COVID-19 Among Right-Wing Extremist Telegram Channels", Institute for Strategic Dialogue, 2021, https://www.isdglobal.org/wp-content/uploads/2021/12/The-Conspiracy-Consortium.pdf.

10 Gerster, Lea et al., „Stützpfeiler Telegram. Wie Rechtsextreme und Verschwörungsideolog:innen auf Telegram ihre Infrastruktur ausbauen", Institute for Strategic Dialogue, 2021, https://www.isdglobal.org/wp-content/uploads/2021/12/ISD-Germany_Telegram.pdf.

11 Für Beispiele von dokumentierter Belästigung und Missbrauch siehe Basu, Tanya, „The Metaverse has a groping problem already", MIT Technology Review, 16. Dezember 2021, https://www.technologyreview.com/2021/12/16/1042516/the-metaverse-has-a-groping-problem/; Bokinni, Yinka, „A barrage of assault, racism and rape jokes: my nightmare trip into the metaverse", The Guardian, 25. April 2022, https://www.theguardian.com/tv-and-radio/2022/apr/25/a-barrage-of-assault-racism-and-jokes-my-nightmare-trip-into-the-metaverse; Robertson, Derek, „Crimefighting in the Metaverse", Politico, 13. April 2022, https://www.politico.com/newsletters/digital-future-daily/2022/04/13/who-will-protect-you-in-the-metaverse-00025070. Für Beispiele von anfänglichen Unternehmensrecherchen und -antworten siehe Blackwell, Lindsay et al., „Harassment in Social Virtual Reality: Challenges for Platform Governance", Proceedings of the ACM on Human-Computer Interaction, 3(100), November 2019, https://dl.acm.org/doi/10.1145/3359202; Gleason, Mike, „Microsoft, Meta tackle harassment in virtual worlds", TechTarget, 17. Februar 2022, https://www.techtarget.com/searchunifiedcommunications/news/252513581/Microsoft-Meta-tackle-harassment-in-virtual-worlds.

12 Jurdak, Raja, Dorri, Ali und Kanhere, Salil S., „Protecting the 'right to be forgotten' in the age of blockchain", The Conversation, 30. Oktober 2018, https://theconversation.com/protecting-the-right-to-be-forgotten-in-the-age-of-blockchain-104847.

13 Shapiro, Elizabeth Hansen et al., „New Approaches to Platform Data Research", Netgain Partnership, Februar 2021, https://drive.google.com/file/d/1bPsMbaBXAROUYVesaN3dCtfaZpXZgl0x/view.

14 Dickson, EJ, „Inside MeWe, Where Anti-Vaxxers and Conspiracy Theorists Thrive", Rolling Stone, Mai 2019, https://www.rollingstone.com/culture/culture-features/mewe-anti-vaxxers-conspiracy-theorists-822746/.

15 Crawford, Angus, „Kik chat app 'involved in 1,100 child abuse cases'", BBC, 21. September 2018, https://www.bbc.co.uk/news/uk-45568276.

16 Goel, Vindu, „Facebook Promises Deeper Review of User Research, but Is Short on the Particulars", New York Times, 2. Oktober 2014, https://www.nytimes.com/2014/10/03/technology/facebook-promises-a-deeper-review-of-its-user-research.html.

17 Perez, Sarah, „Facebook rolls out more API restrictions and shutdowns", TechCrunch, 2. Juli 2018, https://tcrn.ch/2IKza9A.

18 Munn, Luke, „More than a mob: Parler as preparatory media for the U.S. Capitol storming", First Monday, 26(3), Februar 2021, https://doi.org/10.5210/fm.v26i3.11574; Gais, Hannah und Cruz, Freddy, „Far-Right Insurrectionists Organized Capitol Siege on Parler", SPLC, 8. Januar 2021, https://www.splcenter.org/hatewatch/2021/01/08/far-right-insurrectionists-organized-capitol-siege-parler.

19 Shieber, Jonathan, „Parler jumps to No.1 on App Store after Facebook and Twitter ban Trump", TechCrunch, 9. Januar 2021, https://techcrunch.com/2021/01/09/parler-jumps-to-no-1-on-app-store-after-facebook-and-twitter-bans/.

20 Lee, Micah, „Inside Gab, the Online Safe Space for Far-Right Extremists", The Intercept, 15. März 2021, https://theintercept.com/2021/03/15/gab-hack-donald-trump-parler-extremists/.

21 „FAQs / Explainers", Global Internet Forum to Counter Terrorism, https://gifct.org/explainers/.

22 „Image Hash List", Internet Watch Foundation, https://www.iwf.org.uk/our-technology/our-services/image-hash-list.

23 Macklin, Graham, „The Christchurch Attacks: Livestream Terror in the Viral Video Age", Combating Terrorism Center, 12(6), Juli 2019, https://ctc.usma.edu/christchurch-attacks-livestream-terror-viral-video-age/; Frenkel, Sheera, Decker, Ben und Alba, Davey, „How the 'Plandemic' Movie and Its Falsehoods Spread Widely Online", The New York Times, 21. Mai 2020, https://www.nytimes.com/2020/05/20/technology/plandemic-movie-youtube-facebook-coronavirus.html.

24 „Data Critique and Platform Dependencies: How to Study Social Media Data? Digital Methods Winder School and Data Sprint 2022", Digital Methods Initiative, https://wiki.digitalmethods.net/Dmi/WinterSchool2022.

25 Timberg, Craig, „Facebook made big mistake in data it provided to researchers, undermining academic work", The Washington Post, 10. September 2021, https://www.washingtonpost.com/technology/2021/09/10/facebook-error-data-social-scientists/.

26 Siehe insbesondere Baym, Nancy K., Tune In, Log On: Soaps Fandom, and Online Community, SAGE Publications, Inc., 2000; Jenkins, Henry, Convergence Culture, NYU Press, 2006.

27 „How it works", Ad Observer, https://adobserver.org.

28 Kazemi, Ashkan et al., „Tiplines to Combat Misinformation on Encrypted Platforms: A Case Study of the 2019 Indian Election on WhatsApp", arXiv:2106.04726, Juli 2021, https://doi.org/10.48550/arXiv.2106.04726.

29 „Homepage", Global Internet Forum to Counter Terrorism, https://gifct.org/.

30 „Homepage", Syrian Archive, https://syrianarchive.org.

31 „Homepage", Yemeni Archive, https://yemeniarchive.org.

32 „User Experience of Potential Online Harms within Video Sharing Platforms", OFCOM (UK Government), 1. Februar 2020, https://www.gov.uk/find-digital-market-research/user-experience-of-potential-online-harms-within-video-sharing-platforms-ofcom.

33 Schade, Amy, „Remote Usability Tests: Moderated and Unmoderated", Nielsen Norman Group, 12. Oktober 2013, https://www.nngroup.com/articles/remote-usability-tests/.

34 Gil de Zúñiga, Homero und Goyanes, Manuel, „Fueling civil disobedience in democracy: WhatsApp news use, political knowledge, and illegal political protest", New Media & Society, Oktober 2021, https://doi.org/10.1177%2F14614448211047850.

35 Lamberty, Pia, Holnburger, Josef und Goedeke Tort, Maheba, „CeMAS-Studie: Das Protestpotential während der COVID-19-Pandemie", CeMAS, 17. Februar 2022, https://cemas.io/blog/protestpotential/.

36 Siehe z. B. Bond, Shannon, „NYU Researchers Were Studying Disinformation on Facebook. The Company Cut Them Off", NPR, 4. August 2021, https://www.npr.org/2021/08/04/1024791053/facebook-boots-nyu-disinformation-researchers-off-its-platform-and-critics-cry-f; Clark, Mike, „Research Cannot Be the Justification for Compromising People's Privacy", Meta, 3. August 2021, https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/; Edelson, Laura und McCoy, Damon, „We Research Misinformation on Facebook. It Just Disabled Our Accounts", The New York Times, 10. August 2021, https://www.nytimes.com/2021/08/10/opinion/facebook-misinformation.html.

37 Shapiro et al., op. cit.

38 4cat, Archived.Moe, Dewey Defend, DISBOARD, Lyzem, Method52, OSINT Combine Alt-Tech Social Search, Social Blade, Telegago, TelegramDB, Tgram.io, TGStat und Unfurl.

39 „Top 100 Biggest Discord Servers", Discord, https://discords.com/servers/top-100.

40 „Discord servers tagged with 4chan", DISBOARD, https://disboard.org/servers/tag/4chan.

41 Discord API, https://discord.com/developers/docs/resources/channel#get-channel-messages

42 „Discord-Scraper", GitHub, https://github.com/Dracovian/Discord-Scraper#readme.

43 „Discord Developer Portal — Documentation — OAuth2", Discord, https://discord.com/developers/docs/topics/oauth2#bot-vs-user-accounts.

44 Leidig, Eviane, „Odysee: The New YouTube for the Far-Right", Global Network on Extremism & Technology, 17. Februar 2021, https://gnet-research.org/2021/02/17/odysee-the-new-youtube-for-the-far-right/.

45 Bokinni, op. cit.

46 Meaker, Morgan, „Germany Has Picked a Fight With Telegram", WIRED, 3. Februar 2022, https://www.wired.co.uk/article/germany-telegram-covid.

ISD | Powering solutions to extremism and polarisation

Amman | Berlin | London | Paris | Washington DC

www.isdglobal.org