

Soziale Medien als *accidental archives*

Ein Gespräch mit Maria Mingo (Mnemonic)

Anna Polze

Gewalt kann in den sozialen Medien vielfältige Formen annehmen. Plattformen sind nicht nur Schauplätze von Hassrede und Shitstorms, sondern auch Orte, an denen affizierende Bilder oder illegale Postings zirkulieren, die grausame Handlungen und Gewalt an Kriegsschauplätzen zeigen. Oft werden sie von automatisierter Content-Moderation entfernt. In einigen Fällen könnte dieses Material jedoch zur Aufklärung von Menschenrechtsverletzungen beitragen. Angesichts dieses ambivalenten Umgangs mit gefährdendem Bild- und Videomaterial hat sich die Organisation Mnemonic etabliert, um solche Daten fallbezogen zu archivieren, zu verifizieren und für Rechtsverfahren zur Verfügung zu stellen. Mnemonic hat ein Protokoll zum Aufbau von archivischen Datenbanken entwickelt, um Bilddaten zu sichern, bevor sie von den Plattformen entfernt werden. Die besondere Zeitökonomie der kurzen Form, welche die affektgeladene Interaktion in den sozialen Medien, für die sich dieser Sammelband interessiert, auszeichnet, erhält aus der Perspektive der Datensicherung weitere Charakteristiken: Nicht nur die Textmenge der Postings selbst ist als *kurze* zu bezeichnen, sondern auch die Zeitlichkeit von geteiltem Content unterliegt einer konstitutiven Knappheit. Ebenso wie Plattformen die maximale Zeichenanzahl eines Postings regulieren, so wirken Content-Moderation-Verfahren auf die Zeitspanne ein, in der Material überhaupt sichtbar ist. Die Bilddaten, für die sich Mnemonic interessiert, sind vor dem Hintergrund ihres ephemeren Charakters zu betrachten und erfordern eine *rapid response*. Diese *response* liegt neben dem instantanen Archivieren im De-Eskalieren des Materials, um zugleich auf die Erweiterung von dessen Wirkungsradius abzielen. Das Gewaltpotenzial expliziter Postings wird durch eine Übersetzung in Tabellen, welche die Metadaten der Bilder, Informationen zu Zeitpunkt und Ort der Veröffentlichung und Hinweise zur Überprüfung der Echtheit des Bildes enthalten, gemindert.

Nicht allein die visuellen Inhalte von Bildern, sondern diese zugehörigen Daten ermöglichen juristische Verhandlungsweisen solcher potenziellen Beweisschnipsel, die sich substanziell von der Zirkulation und Interaktion in den sozialen Medien unterscheiden: Vor Gericht unterliegen Bilder anderen Kriterien der Evidenzerzeugung. Sind es auf Plattformen gesteigerte Sichtbarkeit dank Reichweitenvergrößerung durch möglichst viele Kommentare, Likes oder Re-Posts, die sozialmediale Evidenz ermöglichen, so müssen Bilder vor Gericht durch eine möglichst genaue bildforensische Rekonstruktion von Entstehungs- und Bearbeitungskontext verifiziert werden. Die Arbeit, die Mnemonic leistet, ist darum eine der fallbezogenen Intervention infrastrukturell vorgelagerte, die auf die grundsätzliche Erschließung von Gerichtsräumen abzielt. Das Gespräch thematisiert deshalb die kontinuierliche Lobby-Arbeit, an der sich Mnemonic beteiligt, um öffentlich zugängliches Bildmaterial als Beweismaterial in der internationalen Rechtsprechung zu etablieren. Solche politischen Einsätze zielen jedoch nicht zwangsläufig auf eine Parteinahme oder ein aktivistisches Mandat für bestimmte Positionen ab. Gerade im aufstrebenden Bereich der *open source intelligence* (OSINT) lassen sich aktivistischer Einsatz, parteiliche Fürsprache und die Weiterentwicklung und Veröffentlichung forensischer Methoden oft kaum voneinander trennen. Doch gerade in dieser Frage lohnt sich ein differenzierter Blick, um nicht nur die vermeintliche Parteilichkeit von zivilgesellschaftlichen Initiativen, sondern auch die den Plattformen und ihren Akteur*innen innewohnenden Politiken in ihren Auswirkungen sowohl auf digitale als auch auf konkrete Schauplätze von Gewalt zu problematisieren.

Maria Mingo ist auf die Dokumentation von Gräueltaten durch Medientechnologie spezialisiert. Bei der Organisation Mnemonic ist sie für den Bereich Content-Moderation zuständig und arbeitet mit Social-Media-Plattformen, NGO-Koalitionen und anderen Akteur*innen zu Fragen nach Löschung und Bewahrung von Content.

AP: Wie wurde Mnemonic gegründet und was waren die Hintergründe, die zur Entstehung der Organisation führten?

MM: Mnemonic wurde 2014 als *Syrian Archive* gegründet, das von Syrer*innen als Reaktion auf den Krieg dort betrieben wurde. Das Ziel des *Syrian Archive* war, wichtige Menschenrechtseinhalte, die in den sozialen Medien veröffentlicht worden waren, zu sammeln und zu archivieren, um sie einerseits für strafrechtliche Prozessführungen und andererseits zur Erinnerung an die Ge-

schehnisse zu nutzen. Mnemonic wurde erst später, 2018, als Dachorganisation gegründet. Wir folgen dem Verständnis, dass wir das, was wir für Syrien getan haben, auch für andere Kontexte replizieren können. Natürlich geschieht dies in dem Wissen, dass jeder Konflikt anders ist. Neben dem *Syrian Archive* beherbergt Mnemonic nun auch das *Yemeni Archive*, das *Sudanese Archive* und das *Ukrainian Archive*. Außerdem haben wir ein *rapid-response*-Programm und ein Programm zur Content-Moderation.

Im Jahr 2014 begannen wir mit unserer Tätigkeit, weil mutige Aktivist*innen vor Ort in Syrien die Situation dokumentierten, indem sie Fotos und Videos von den Ereignissen machten. Sie betrachteten die sozialen Medien als Plattformen, auf denen man über die Geschehnisse berichten konnte. Dies geschah oft, bevor Journalist*innen oder Ermittler*innen die Schauplätze erreichten. Es waren also diese von Aktivist*innen und Bürger*innen aufgenommenen Informationen und Schnipsel, die in Echtzeit aus dem Land kamen. Die sozialen Medien wurden dabei einerseits für die Berichterstattung genutzt, andererseits aber auch oft als Backup eingesetzt. Wenn Menschen fliehen oder Kontrollpunkte passieren mussten, luden sie ihr Material auf soziale Medien hoch, um sicherzustellen, dass sie es aus Sicherheitsgründen von ihren Telefonen löschen konnten. Mit dem Upload in die sozialen Medien gäbe es immer noch eine Sicherungskopie dieser Inhalte, auf die die Menschen online zurückgreifen könnten.

AP: Aber die Plattformen der sozialen Medien waren nie als Archive angelegt.

MM: Das stimmt. Sie sind zu *accidental archives* geworden, doch das war nie ihr eigentlicher Zweck. Das *Syrian Archive* ist im Wesentlichen als Archiv entstanden, um die Informationen, die in den sozialen Medien gepostet wurden, sicherer zu bewahren. Vor allem wenn es um Ermittlungen und juristische Prozessführung geht, stellen wir sicher, dass die Inhalte dafür zur Verfügung stehen. Die Personen vor Ort luden ihr Material hoch, kamen zu uns und fragten, ob wir die Inhalte sehen, aufbewahren und archivieren könnten. Manchmal konnten wir das, aber manchmal waren die Inhalte bereits von der Plattform entfernt worden. Informationen sind flüchtig, vor allem, wenn wir über Social-Media-Plattformen sprechen, die keine nachhaltigen Sammelorte sind, eben weil sie nicht als Archive konzipiert sind. Vor diesem Hintergrund hat Mnemonic seine Arbeit aufgenommen.

AP: Erinnerungswerkzeuge oder Mnemotechniken haben natürlich eine lange Geschichte. Eure Arbeit hingegen ist eng mit der digitalen Kultur und den sozialen Medien verbunden. Wie denkt ihr über die Funktion und Rolle der sozialen Medien und Plattformen für politische und juristische Anliegen? Zum Beispiel gab es zu einem gewissen Zeitpunkt hoffnungsvolle Stimmen, die unter Stichworten wie *Facebook-Revolution* auf Möglichkeiten zur Vernetzung, zur Planung von Veranstaltungen und zur Herausbildung von Protesten hingewiesen haben. Auf der anderen Seite sind soziale Medien nicht per se offene Versammlungsorte. Sie folgen einer Aufmerksamkeitsökonomie und ihr Handeln ist von finanziellen Interessen getrieben. Wie würdest du diese Ambiguität beschreiben?

MM: Was den meisten Menschen in den Sinn kommt, wenn sie an soziale Medien in Krisenzeiten denken und an die anfängliche Hoffnung, die du ansprichst, ist der Arabische Frühling. Es geht um die Mobilisierung von Menschen und darum, dass es diese Plattformen gibt, auf denen sich die Menschen organisieren, koordinieren und schließlich massenhaft Aktionen durchführen können. Das Problem, das wir beobachten, ist der Spagat, der entsteht, wenn Plattformen einerseits als Werkzeug zur Organisation und zur Echtzeitberichterstattung genutzt werden und andererseits *safe spaces* darstellen sollen. Mit *safe spaces* meine ich Plattformen, auf denen die von den Nutzer*innen geposteten Inhalte niemandem schaden, also weder den Leuten, die posten, noch denen, die die Beiträge sehen.

Hier ist die Moderation des Contents sehr wichtig. Wir wollen und brauchen eine Moderation der Inhalte. Bestimmter Content sollte nicht online bleiben. Das ist wichtig, vor allem wenn wir über illegale Postings sprechen, die gegen die Richtlinien der Plattformgemeinschaft verstoßen (z.B. besonders gewaltsames Material).

Als Reaktion auf die Menge an Inhalten, die online gepostet wurden, und die zunehmende Nutzung sozialer Medien haben die Regierungen, insbesondere seit den Terroranschlägen in Christchurch im März 2019, einen erheblichen Druck ausgeübt, um schädliche Inhalte sowohl schnell als auch in großem Umfang zu entfernen. Es gibt jedoch grundsätzliche Fragen zu den Maßnahmen der Content-Moderation und deren Steuerung durch die Plattformen: Wie werden die Inhalte entfernt, durch menschliche Moderator*innen oder durch Algorithmen? Und nach welchen Kriterien geschieht dies? Sind es die richtigen Inhalte, die entfernt werden? Was geschieht mit Inhalten, die zu Recht aus dem öffentlichen Raum entfernt werden, die

aber wichtig sind, um beispielsweise Menschen für schwere Verbrechen und Menschenrechtsverletzungen zur Verantwortung zu ziehen?

AP: Welche Folgen hatte dieser Vorstoß der Regierungen? Inwieweit waren sie negativ?

MM: Letztlich lag es an den Plattformen, die Auflagen der Regierungen umzusetzen. Wie wir beobachteten, kam es zu einer überzogenen Reaktion, um Geldstrafen und Risiken zu vermeiden. Die Plattformen schienen eher mehr Inhalte zu entfernen als zu wenig und nutzten oft maschinelle Lernalgorithmen statt menschlicher Prüfer*innen. Natürlich spielen dabei eine Reihe an Faktoren eine Rolle, nicht zuletzt die psychische Gesundheit der Moderator*innen. Ständig mit dieser Art von Material konfrontiert zu sein ist schwierig und herausfordernd. Hoffentlich handeln die Plattformen aus der Motivation heraus, ihre Mitarbeiter*innen zu schützen, aber wahrscheinlich sind es auch Effektivität und Kosteneinsparungen, die sie motivieren. Die Funktionsweise dieser Algorithmen ist jedoch kaum transparent.

Was wir heute beobachten, ist, dass entweder zu viel oder zu wenig schädliche Inhalte entfernt werden, vor allem, wenn es sich um bestimmte Sprachen handelt, die bei den Content-Moderator*innen oder den antrainierten Sprachmodellen unterrepräsentiert sind. Ich will damit nicht sagen, dass es weniger Prüfung und Moderation geben sollte, sondern vielmehr eine bessere, menschliche Moderation, insbesondere bei Bildmaterial aus Krisengebieten. Ich denke da zum Beispiel an Äthiopien, wo Amnesty International gerade eine Klage gegen Meta eingereicht hat. In der Klage wird argumentiert, dass die Algorithmen von Facebook die Verbreitung schädlicher Rhetorik, die sich gegen die tigrayische Gemeinschaft richtet, förderten, weil die Moderationssysteme des Unternehmens solche Inhalte nicht erkannten und nicht angemessen darauf reagierten. Diese Versäumnisse trugen mutmaßlich zur Ermordung des tigrayischen Universitätsprofessors Meareg Amare bei. Berichten zufolge wurde er von einer Gruppe von Männern ermordet, nachdem im Jahr 2021 auf Facebook gegen ihn gerichtete Beiträge veröffentlicht worden waren. Amnesty behauptet, dass sich Metas Herangehensweise an Krisensituationen in Afrika von der in anderen Teilen der Welt, insbesondere in Nordamerika, unterscheidet. Die Algorithmen sind nicht in der Lage, sprachliche Nuancen zu erkennen, und Meta hat deutlich zu wenig Ressourcen in die Moderation von Content anderer Sprachen als Englisch investiert.

AP: Anhand deiner Ausführungen verstehe ich jetzt besser, wie politische Ereignisse und die Struktur sozialer Medien aufeinander einwirken, denn Plattformen sind nicht stabil, sondern entwickeln sich weiter, je nachdem, was mit ihnen gemacht wird. Gleichzeitig spielen Entscheidungen in der Plattform-Gestaltung wie die Einführung des kuratierten Feeds eine Rolle, der sich von persönlichen Nachrichten, Gruppen oder anderen Funktionen unterscheidet. Im von Algorithmen kuratierten Feed kann man über Material stolpern, das man eigentlich nicht zu sehen bereit ist.

MM: Genau, und das ist der Grund, warum Mnemonic ins Leben gerufen wurde. Um dieses Material zeitnah zu erhalten, damit es für juristische Zwecke und andere Erinnerungspraktiken zur Verfügung steht. Je mehr Algorithmen verwendet werden, desto schneller wird Material gelöscht, bevor es überhaupt jemand gesehen hat. Daher existiert eine Menge von Daten und Informationen, die wir nicht archivieren können. Die Frage ist, was mit diesen Inhalten, diesen Bildern und Videos, die entfernt wurden, geschieht. Wir sprechen hier über die Richtlinien der Plattformen zur Aufbewahrung von Content. Sie bewahren den Content auf, weil sie ihn sonst nicht wiederherstellen können. Sie geben an, mit den Strafverfolgungsbehörden zusammenzuarbeiten, doch das könnten sie nicht, wenn sie das gepostete Material nicht speicherten. Allerdings gibt es keine Transparenz in diesem Bereich. Soweit wir wissen, sichern die Plattformen den Content für ein bis drei Jahre. Diese Zeitspanne ist für einen Strafprozess jedoch wahrscheinlich zu kurz.

Wir trugen beispielsweise zu einer Klage bei, die das *Syrian Archive* zusammen mit anderen Partner*innen bei den französischen Untersuchungsrichter*innen zu den Sarin-Gas-Angriffen in Syrien eingereicht hat. Unsere Bemühungen führten zu einem Haftbefehl gegen Präsident al-Assad und drei weitere hohe Beamte. Ich erwähne das aus dem Grund, weil diese Recherche über zehn Jahre Arbeit gekostet hat. Wir steuerten die Nachrichten der sozialen Medien und Open-Source-Content bei, während andere Partner*innen Zeug*innenaussagen oder andere wichtige Informationen einbrachten. Es hat sehr lange gedauert, das Material zusammenzutragen, zu verifizieren und den Fall aufzubauen. Wenn uns zugetragen wird, dass Plattformen Inhalte ein bis drei Jahre lang aufbewahren, sind wir besorgt, weil Ermittlungs- und Rechenschaftsprozesse viel länger dauern. Deshalb fordern wir mehr Transparenz in Bezug auf die Aufbewahrungsrichtlinien für Social-Media-Content und die Auswirkungen, die sie auf Ermittlungen haben können.

AP: Die Frage nach der stabilen Datensicherung über einen langen Zeitraum bringt mich zu der Frage nach eurer Infrastruktur. Welche Art von Informationen müssen aufbewahrt werden und wie müssen diese aufbereitet sein, um sicherzustellen, dass die Inhalte auch in zehn Jahren noch genutzt werden können, selbst wenn sich etwa Bildformate in der Zwischenzeit geändert haben? Welche Zeiträume setzt ihr für eure Arbeit an?

MM: Wir verfügen über ein fantastisches Technik-Team, das unglaubliche Arbeit leistet. Der Aufbau von Infrastruktur ist ein wichtiger Teil unserer Arbeit, eben weil wir Archive mit Open-Source-Informationen entwickeln und pflegen. Es ist eine Menge Arbeit, relevante Inhalte zu sichten und zu identifizieren. Vieles davon geschieht manuell, obwohl wir den Prozess durch Technologie so weit wie möglich automatisieren. Anschließend sichern wir diese Informationen, überprüfen sie aber auch in Bezug auf konkrete Investigationen. Wir bewahren den Content auf unbestimmte Zeit auf, was sehr kostspielig ist. Daher haben wir einen Nachhaltigkeits- und Sicherungsplan für alle unsere Archive entwickelt, um sicherzustellen, dass sie je nach Kontext und Bedarf an relevante Akteur*innen zur Dokumentation oder für Prozessführungen weitergegeben werden können.

Wir archivieren dafür nicht nur das potenzielle Beweismaterial selbst, sondern auch die relevanten Metadaten. Da es sich um Informationen handelt, die für eine mögliche Gerichtsbarkeit gespeichert werden müssen, ist die Art und Weise, wie diese Daten gehandhabt werden, zentral, weil wir nachweisen müssen, dass sie nicht nur relevant, sondern auch stichhaltig sind. Bei Social-Media-Inhalten kann das ziemlich schwierig sein, denn wenn es sich um Inhalte handelt, die man online findet, weiß man in der Regel nicht, was mit diesen Inhalten seit ihrer Aufnahme und bis zu ihrer Veröffentlichung passiert ist. Es gibt Lücken in der Beweismittelkette. Wir halten die Beweismittelkette von dem Moment an aufrecht, in dem wir sie erfassen. Zudem arbeiten wir mit einer standardisierten Liste für Metadaten, die wir sammeln, und einer begründeten Liste der Quellen. Sie enthält Konten, Nutzer*innen und die Art von Informationen, von denen wir aufgrund unserer Beobachtungen und Erfahrungen wissen, dass sie zuverlässig sind. Wir stellen sicher, dass alles so gut wie möglich gehandhabt wird. Unser Ziel ist es, dass die Ermittler*innen wissen, dass sie sich auf diese Informationen verlassen können, wenn sie sie in die Hände bekommen – was ihnen letztlich die Arbeit erleichtert.

AP: Mnemonic ist also eine Schnittstelle zwischen den Bilddaten in den sozialen Medien und der Ermittlungsarbeit, weil ihr über Bewahrungsstandards und -mittel verfügt und die Daten habt, die die Ermittler*innen tatsächlich brauchen könnten.

MM: Absolut, insbesondere mit dem vermehrten Aufkommen von KI. Obwohl es schon lange ein Problem ist, bleibt die grundsätzliche Tatsache, dass Fotos verändert werden können, unser Hauptaugenmerk. Dass es immer neue Werkzeuge zur Veränderung von Bilddaten gibt, bedeutet, dass selbst wenn man die Quelle kennt und weiß, dass es sich um ein Original handelt, die urteilenden Richter*innen dies nicht wissen oder nicht ohne Weiteres nachvollziehen können. Da das Material verändert worden sein könnte, sind die Anforderungen an die Zulässigkeit höher. Das kann zu großen Schwierigkeiten führen. Wenn es eine Lücke in der Beweismittelkette gibt, versuchen wir zum Beispiel zu überprüfen, ob das Bild- oder Videodokument zu dem Zeitpunkt und an dem Ort aufgenommen wurde, an dem es angeblich entstand. Wir müssen also eine Open-Source-Untersuchung durchführen, was sehr zeitaufwändig ist und spezielle Fähigkeiten erfordert. Es bedeutet auch, dass man Tausende von Datensätzen durchforsten muss und vielleicht zu keinem Ergebnis kommt. Ein Foto kann entweder viel oder gar nichts aussagen, je nachdem, wie es aufgenommen wurde. Man arbeitet mit den Informationen, die einem zur Verfügung stehen, und die Überprüfung dieser Informationen ist enorm wichtig. Allerdings befinden sich über 15 Millionen Fotos und Videos in unseren verschiedenen Archiven. Wir können nicht jedes einzelne Objekt verifizieren, das wäre unmöglich. Es geht also auch darum, innerhalb dieser Informationen Prioritäten zu setzen, bestimmte Vorfälle zu überprüfen und darüber hinaus zu ermitteln, damit die von uns bereitgestellten Informationen möglichst nützlich und zuverlässig sind.

AP: Angesichts all dieser Faktoren – die schiere Anzahl verfügbarer Bilder, ihr ephemerer Status in den sozialen Medien, die Rechtsnormen und womöglich die Konventionen und Strukturen der Gerichte selbst – erscheint es so unwahrscheinlich, dass ein Bild zu einem Beweismittel werden kann. All dies spricht gegen Bilder als Beweismaterial. Dennoch ist es das meistgenutzte Werkzeug, um Ereignisse zu dokumentieren. Man greift zum Handy und macht ein Foto oder Video, teilt es in den sozialen Medien.

MM: Der Bereich der Open-Source-Ermittlungen hat sich im letzten Jahrzehnt stark weiterentwickelt. Dies zeigt sich in der größeren Akzeptanz von Open-Source-Informationen, wenn sie in Begleitung einer strengen Analyse ihrer Herkunft, ihrer Handhabung und des Kontextes, in dem sie entstanden sind, vorgelegt werden. Die Relevanz dieser Materialien wird mehr und mehr deutlich. Um auf die Haftbefehle gegen die syrischen Beamten zurückzukommen: Die Tatsache, dass die französischen Ermittlungsrichter*innen die Stichhaltigkeit der Klage und der zur Verfügung gestellten Informationen, einschließlich der Open-Source-Inhalte in den sozialen Medien, erkannt haben, bedeutet sehr viel. Sie hätten keinen Haftbefehl von so großer Tragweite ausgestellt, ohne zumindest den Eindruck zu haben, dass sie über eine ausreichende Grundlage verfügen, so zu handeln. Das zeigt, dass diese Art von Informationen mehr und mehr akzeptiert wird. Trotzdem gibt es noch eine Menge zu tun. Jede Rechtsprechung ist anders. Es gibt einige Grundsätze und Richtlinien, die allgemein gültig sind, wie zum Beispiel das Diktum von Relevanz und Zuverlässigkeit der Informationen. Andere Faktoren variieren von Rechtsprechung zu Rechtsprechung. Der Einsatz von Open-Source-Materialien aus den sozialen Medien ist ein Vorstoß, der in den letzten zehn Jahren an Fahrt aufgenommen hat, und die Notwendigkeit nach Verifizierung wird dabei nie verschwinden.

AP: Auf der Website von Mnemonic steht, dass die NGO an *transitional justice*-Prozessen mitwirkt, und ich habe mich auch gefragt, wie ihr zu Ansätzen strategischer Prozessführung steht. In welchem Verhältnis stehen Content-Verifizierung und Rechtsprechung? Denn es scheint euch um Fälle zu gehen, die als Vorbild für künftige Prozessführungen dienen können. Folgt euer Ansatz, eine Archivinfrastruktur zu entwickeln, einem aktivistischen Interesse?

MM: Gerechtigkeit bedeutet für verschiedene Gemeinschaften unterschiedliche Dinge. Es gibt überall verschiedene Formen. Wir befassen uns unter anderem mit nationaler, regionaler und internationaler Justiz. Einige nationale Behörden zur Aufklärung von Kriegsverbrechen nehmen Fälle an, die anderswo von Ausländer*innen begangen wurden. Die einzige Verbindung besteht vielleicht darin, dass sich die Täter*innen kurzzeitig in diesem Land aufgehalten haben. Für uns heißt das, dass wir eine Bestandsaufnahme machen, um zu verstehen, wer in welchen Situationen ermittelt, um mit den verschiedenen Mechanismen von Gerichtsbarkeiten zu arbeiten. Wir wenden uns entweder selbst mit Informationen an Ermittlungsverfahren, oder Ermittler*innen

kommen zu uns, um zu erfahren, ob wir über bestimmte Informationen verfügen, die für ihre Untersuchungen relevant sein könnten. Letztlich hilft die Kombination aus dem Vorhandensein von Informationen und der Fähigkeit, diese zu analysieren, Rechtsstreitigkeiten voranzutreiben.

AP: Forschungsagenturen wie Forensic Architecture formulieren einen *counter-forensischen* Ansatz. Sie thematisieren und kritisieren, wie staatliche Ermittlungsbehörden arbeiten. Wie steht Mnemonic zu dieser Position?

MM: Einer der größten Unterschiede ist, dass unser Hauptziel darin besteht, potenzielle Beweise für Gerichtsverfahren zu sichern, bevor sie verschwinden. Unser Endziel ist zwar der Gerichtsprozess, aber in erster Linie geht es uns um die Archivierung. Dieser erste Schritt unterscheidet sich stark von anderen Gruppen, die Open-Source-Untersuchungen durchführen. Aus dem Blickwinkel der Datensicherung und Archivierung bedeutet dies, dass wir den Umgang mit und die Auswirkungen unserer Aktivitäten auf die Informationen, die wir besitzen, im Auge behalten müssen. Wenn wir beispielsweise Lobbyarbeit betreiben, müssen wir sehr darauf achten, dass wir keine bestimmten gerichtlichen Ergebnisse anstreben, dass unsere Untersuchungen unparteiisch bleiben und dass wir alle Seiten des Konflikts betrachten. Auch dass wir nichts fordern, was zu Voreingenommenheit führen oder die Informationen, die wir besitzen, in irgendeiner Weise gefährden könnte, ist zentral. Wir können zum Beispiel für die Schaffung von Rechenschaftsmechanismen oder die Anerkennung von Open-Source-Informationen als potenzielle Beweise plädieren. Wir setzen uns auch stark für Content-Moderation ein, müssen aber darauf achten, dass wir keine Grenzen überschreiten, die die von uns gesammelten Informationen in Frage stellen könnten. Oft arbeiten wir mit verschiedenen Gruppen zusammen, um Informationen auszutauschen, um Rechenschaftspflicht zu fördern und doppelte Arbeit zu vermeiden.

AP: Bis jetzt haben wir meistens über die Archivierung von Bildern gesprochen, also über Videos oder Fotos. Aber Bilder stehen in den sozialen Medien nicht für sich allein. Sie werden von Titeln und Hashtags umrahmt und oft folgen ihnen schriftliche Kommentare. Sie erhalten Reaktionen, zum Beispiel Likes. Welche Rolle spielen die verschiedenen Formen von Content, von Text und Bildern in eurem Archivierungsansatz?

MM: Wir leben in einem Zeitalter, in dem manchmal sogar die Täter*innen selbst die entscheidenden Informationen posten, so dass die Beiträge, die den (Bild)Inhalt begleiten und sich womöglich selbst mit ihm verlinken, aus ermittlungstechnischer Sicht wirklich hilfreich sind. Wir versuchen, so viele Daten wie möglich auf standardisierte Art und Weise zu archivieren, und wir versuchen, immer dieselbe Art von Daten zu sammeln. Natürlich hängt es davon ab, was verfügbar ist, aber wir versuchen, so viel wie möglich zu sichern, um Ermittlungen zu unterstützen und gleichzeitig das Prinzip der Datenminimierung zu beachten. So erfassen wir beispielsweise das Foto oder Video, den Beitrag und alle von der Nutzer*in bereitgestellten Informationen sowie alle Metadaten, die uns dabei helfen können, festzustellen, wo und wann und von wem der Content mutmaßlich aufgenommen wurde, um unsere Überprüfung zu erleichtern. Wir haben eine standardisierte Liste der Informationen, die wir sammeln.

AP: Wenn man über andere Formen gewalttätiger Inhalte in den sozialen Medien nachdenkt, wie Hassreden, Verschwörungstheorien und rechtsextreme Propaganda, die zum Beispiel in Kommentarspalten auftauchen, dann sind das zum Teil suggestive Postings, die nicht zum Raster algorithmisierter Content-Moderation passen. Versucht ihr auch, diese interaktive Gewalt, die auf Plattformen selbst stattfindet, zu dokumentieren?

MM: Das hängt sehr stark vom Kontext selbst und der Situation ab. In erster Linie suchen wir nach konkreten Vorfällen, die sich vor Ort – also außerhalb der Plattform – ereignet haben, und natürlich nach Informationen, die damit in Zusammenhang stehen. Wenn wir auch damit zusammenhängende Beiträge finden, die irreführend sein könnten, dokumentieren wir sie nach dem Motto »erst bewahren, dann analysieren«. Wir versuchen dann, mit so vielen Daten wie möglich das Gesamtbild zu betrachten. Unser Hauptaugenmerk liegt nach wie vor auf den Geschehnissen vor Ort, aber Hasspostings haben auch Auswirkungen auf die Situation vor Ort, und es ist sehr wichtig zu beobachten, wie die Plattformen damit umgehen.

AP: Wenn ich als tägliche Nutzerin sozialer Medien einen Post archivieren wollte, würde ich einfach einen Screenshot machen. Dies wäre die gängigste und einfachste Art, etwas festzuhalten, das online passiert ist. Sind Screenshots Teil eurer Archivierungsstrategie?

MM: Ja, das sind sie. In unserer Archivierungspipeline erkennen wir die gesprochene Sprache und standardisieren das Datenformat (wobei wir das alte Format beibehalten). Wir machen Screenshots und laden die Webseite herunter, die den Inhalt enthält. Wir sammeln auch die spezifischen URLs, die im Anschluss in unsere Datenbank eingespeist werden. Die Datenbank speichert eine Offline-Kopie des geposteten Inhalts und sammelt automatisch alle erforderlichen Metadaten.

AP: Ich habe über Screenshots nachgedacht, weil Screenshots selbst Bilder sind und eine bestimmte Selbstevidenz haben. Sie werden oft in videobasierenden Investigationen oder Nachrichtenberichten verwendet, weil man auch einige Elemente der Social-Media-Plattform, die Ästhetik, die Icons und das Framing sehen kann. Das macht intuitiv Sinn. All diese Informationen sind auch in eurem Archiv vorhanden, aber in einer anderen Form. Sie werden kategorisiert, standardisiert, durchsuchbar und auffindbar gemacht.

MM: Eines der Probleme, die man mit einem Screenshot hat, ist die Tatsache, dass man ihn mit zusätzlichen potenziellen Beweisen koppeln oder dass man beweisen muss, dass er nicht nur relevant, sondern auch zuverlässig ist, weil man einen Screenshot leicht hätte verändern können. Man sieht Screenshots oft in Nachrichtenberichten, aber wenn man sie zur Rechenschaftslegung verwenden will, muss man zusätzliche Informationen hinzufügen, um die Stichhaltigkeit zu erhöhen. Es gibt dafür auch Tools wie Hunchly,¹ die helfen können, den Screenshot-Prozess zu dokumentieren, indem sie unauffällig im Browser mitlaufen und jede Seite erfassen, die im Rechercheprozess besucht wurde. Entscheidend ist jedoch eine Datenbank, die es nicht nur ermöglicht, große Mengen an Inhalten zu speichern, sondern diese auch zu analysieren und die Daten in einer durchsuchbaren Weise zu gruppieren.

AP: Wie wir eingangs angesprochen haben, entwickeln sich Plattformen ständig weiter. Ein Beispiel, das mir in den Sinn kommt, ist natürlich Twitter und wie sich die Plattform verändert hat, nachdem Elon Musk sie übernahm. Es würde mich sehr interessieren, wie sich das auf eure Arbeit ausgewirkt hat.

MM: Die Übernahme durch Elon Musk im Oktober 2022 brachte viele Rückschritte in Bezug auf den Menschenrechtsschutz auf dieser Plattform mit

¹ Vgl. hunch.ly. Online unter: <https://www.hunch.ly/> (letzter Zugriff: 20.03.2024).

sich. Wie wir in einem kürzlich erschienenen Artikel, den ich gemeinsam mit dem *Syrian Archive* und *7amleh-The Arab Center for Social Media Advancement* verfasst habe, erwähnen, hat Twitter seitdem ein klares Signal an die Welt gesendet, dass die Menschenrechte und Transparenzstandards auf der Plattform zurückgehen.² Dies zeigte sich unter anderem darin, dass Twitter seine Menschenrechts- und Kommunikationsteams sowie den Trust & Safety Council abbaute, das Verifizierungssystem der Plattform umstellte, die Zahl der Inhaltsmoderator*innen reduzierte und Journalist*innen suspendierte, die selbst über Musk berichteten – um nur einige Änderungen zu nennen.

Einer unserer Schwerpunkte in Bezug auf Content-Moderation ist die Bewertung von Menschenrechtsrisiken innerhalb der Plattformrichtlinien. Letztes Jahr hat Twitter beispielsweise die Anzahl der Beiträge begrenzt, die Personen sehen konnten, je nachdem, ob sie verifizierte Nutzer*innen waren oder nicht, was die Verifizierung ebenfalls kostspielig machte. Für Forscher*innen kann dies sowohl ein erhöhtes Sicherheitsrisiko als auch finanzielle Belastung bedeuten. Bei Ermittlungen erschweren derartige Beschränkungen die Untersuchung eines Falls und die Aufdeckung relevanter Daten, da man nur einen Ausschnitt zu sehen bekommt. Man weiß nicht unbedingt, welche Informationen angezeigt werden und welche nicht. Nach öffentlichen Reaktionen schien Twitter diese Änderungen heimlich zurückzunehmen, aber das zeigt, wie wichtig Risikobewertungen zu Menschenrechten sind, insbesondere wie sich politische Maßnahmen auf Open-Source-Untersuchungen auswirken können. Die Verfügbarkeit von APIs, die Forscher*innen nutzen können, ist für diese Art Arbeit ebenfalls entscheidend.

Ein weiterer Schwerpunkt unserer Lobbyarbeit liegt auf den übermäßigen Löschungen und der Transparenz darüber, wie Algorithmen funktionieren, um sicherzustellen, dass schädliche Inhalte von den Plattformen entfernt werden, aber wichtige Inhalte, insbesondere solche, die unter berichtenswerte Ausnahme bilden, online bleiben. Hier ist der EU Digital Services Act (DSA) eine wichtige Möglichkeit, die notwendigen Informationen von den Plattformen zu erhalten. Der DSA ist eine EU-Verordnung, die das Potenzial hat, weltweit Standards zu setzen. Sie formuliert Verpflichtungen zu Transparenz,

2 Vgl. Maria Mingo/Ahmed Zidan/Itxaso Domínguez de Olazábal (2023): »X's policy changes stifle human rights work and investigations in Syria, Palestine, elsewhere«, in: *mnemonic.org* (21.11.2023). Online unter: <https://mnemonic.org/en/content-moderation/x%E2%80%99s-policy-changes-stifle-human-rights-work-and-investigations-in-syria-palestine-elsewhere-1> (letzter Zugriff: 20.03.2024).

Risikobewertungen und Bestimmungen über den Zugang zu Daten. Die DSA ist am 17. Februar 2024 für alle Plattformen in Kraft getreten, und wir sehen jetzt, wie sie umgesetzt wird und was sie tatsächlich bedeutet.

Doch was passiert mit Inhalten, die entfernt werden müssen? Deshalb konzentrieren wir uns auch auf die Vorratsdatenspeicherung auf den Plattformen. Wir brauchen mehr Klarheit, insbesondere darüber, welche Akteur*innen (abgesehen von den Strafverfolgungsbehörden) einen Antrag auf Vorratsdatenspeicherung stellen können, welche Art von Daten aufbewahrt und wie mit Situationen umgegangen wird, in denen Inhalte algorithmisch entfernt wurden, bevor sie irgendjemand sehen konnte, und wie die Kriterien für die Aufbewahrung konkret aussehen. Wir arbeiten mit den Plattformen zusammen, wo immer es möglich ist, da wir bei vielen Plattformen als vertrauenswürdige Ansprechpartner*innen gelten, und wir werden in Krisenzeiten konsultiert. Allerdings ist oft viel Druck nötig, um eine Veränderung bei den Plattformen zu erreichen. Daher kann die Arbeit in Koalitionen wie der *DSA Human Rights Alliance* hilfreich sein. Unsere Lobbyarbeit zielt darauf ab, Nicht-EU-Stimmen in diese Diskussionen einzubringen und die Auswirkungen der Plattformpolitik auf Open-Source-Untersuchungen hervorzuheben.

AP: Was wäre ein weiteres Beispiel für so eine Arbeit in Koalitionen, um Plattformen zu mehr Transparenz zu zwingen?

MM: Im Dezember 2023 haben wir zum Beispiel zusammen mit 18 anderen Organisationen die Initiative *Stop Silencing Palestine* gestartet, die sich an Meta richtet.³ Die Kampagne macht deutlich, dass Meta die Auswirkungen seiner Content-Moderations-Richtlinien und Praktiken auf Palästinenser*innen berücksichtigen und ernsthafte Maßnahmen ergreifen muss. Es ist notwendig, das Risiko zu vermindern, dass Metas Moderationspraktiken zu schweren Verbrechen und Menschenrechtsverletzungen im Konflikt beitragen. Zu den zentralen Empfehlungen, um Menschenrechtsforschung, Open-Source-Untersuchungen und Strafverfolgung für schwere Verbrechen zu ermöglichen, gehört eine verstärkte menschenrechtliche Sorgfaltspflicht in Bezug auf die Auswirkungen der Moderations- und Kuratierungsmaßnahmen während dieser Krise. Außerdem wird Transparenz darüber gefordert, wo und wie Automatisierung und maschinelle Lernalgorithmen zur Moderation oder Übersetzung von

3 Vgl. stopsilencingpalestine.com. Online unter: <https://stopsilencingpalestine.com/> (letzter Zugriff: 20.03.2024).

Kommentaren und Posts eingesetzt werden. Die Initiative verlangt eine öffentliche Klärung von Metas Politik zur Aufbewahrung von Content in Krisensituationen und bewaffneten Konflikten sowie den Umgang mit Situationen, in denen Content algorithmisch entfernt wurde. Sie plädiert dafür, dass Meta Forschung im Kontext von Menschenrechtskrisen und -konflikten als eine Angelegenheit von öffentlichem Interesse betrachtet und zivilgesellschaftlichen und akademischen Forscher*innen den Zugang zur Content Library und zur API ermöglicht.

AP: Auch in dieser Initiative, aus der politisches Engagement spricht, geht es immer noch um die Sicherung von potenziell wichtigen Daten.

MM: Ganz genau. Und eine der größten Chancen, die wir derzeit haben, ist die EU-DSGVO, die sehr große Online-Plattformen, zu denen die meisten Social-Media-Unternehmen gehören, dazu verpflichtet, die Nutzer*innenrechte in der EU zu gewährleisten. Es ist eine Menge Kontrolle und Beobachtung erforderlich, und die Europäische Union erlässt immer noch delegierte Rechtsakte.

AP: Um abschließend auf die Frage zur Infrastruktur zurückzukommen: Ich habe mich auch gefragt, was hinter dieser Art von potenziellen Beweisen in den sozialen Medien steckt. Manchmal haben die Menschen nicht die Mittel, den Zugang oder sie haben gerade kein WLAN, um sich online zu vernetzen. In manchen Situationen – wie etwa bei Pushbacks an den EU-Außengrenzen – werden Smartphones von der Polizei konfisziert oder vernichtet, um das Entstehen und Veröffentlichen von Informationen zu verhindern. Smartphones sind die wichtigsten Werkzeuge, um sich an der Sicherung und Archivierung von möglichem Beweismaterial zu beteiligen.

MM: Fotos und Videos sind ein Teil des Puzzles. Normalerweise ergänzen sie Zeug*innenaussagen, dokumentarische Beweise und tatsächliche Dokumente. Das Ganze wird auch durch verschiedene Arten der Forensik ergänzt. Es gibt viele Teile desselben Puzzles, aus denen sich letztlich das Gesamtbild zusammensetzt. Wir konzentrieren uns auf Daten aus sozialen Medien, weil wir erkannt haben, dass es notwendig ist, diese zeitkritischen Informationen zu bewahren. Unsere Technologie ist quelloffen, und wir können Partner*innen Schulungen zur Archivierung dieser Inhalte anbieten. Idealerweise wird diese Art von Inhalten mit anderen Informationen kombiniert, um einen soliden,

belastbaren Fall zu erstellen. Wir haben in Syrien gesehen, wie Social-Media-Plattformen als Backup genutzt wurden, gerade für den Fall, dass Telefone beschlagnahmt wurden. Und während wir uns bislang ausschließlich auf Open-Source-Informationen konzentrieren, können Fotos und Videos auch Closed-Source-Informationen sein, z.B. Fotos und Videos, die von Aktivist*innen oder Bürger*innen aufgenommen wurden, aber nicht an die Öffentlichkeit gelangt sind.

Bei der Speicherung dieser Art von Inhalten gibt es viel mehr Sicherheitsaspekte und andere Empfindlichkeiten. Die Dokumentations- und Archivierungsmethodik für Aktivist*innen in der von dir beschriebenen Situation hängt vom Hauptzweck der Informationen ab. Wenn der Endzweck ein Gerichtsprozess ist, könnten Dokumentationswerkzeuge, die dabei helfen, die Daten auf dem Telefon zu verstecken und eine zuverlässige Sicherungskopie zu erstellen, bis die Inhalte hochgeladen werden können, sinnvoll sein. Ein Beispiel für ein solches Tool ist die *eyeWitness to Atrocities*-App, die von der International Bar Association initiiert wurde, bei der ich früher gearbeitet habe.⁴ Die Hauptschwierigkeit bei der Dokumentation von Closed-Source-Inhalten zum Zwecke der Rechenschaftslegung besteht in der Aufrechterhaltung der Beweismittelkette ab dem Zeitpunkt, an dem das Foto aufgenommen wurde. Die Initiative WITNESS hat zum Beispiel zahlreiche Ressourcen für Foto- und Videotechniken entwickelt, um Aktivist*innen dabei zu helfen, Inhalte so zu dokumentieren, dass sie die Chancen erhöhen, diese Zulässigkeitsstandards zu erfüllen.⁵ Dies erfordert jedoch die Anwendung dieser Techniken zum Zeitpunkt der Aufnahme eines Fotos oder Videos, also eine Auseinandersetzung mit den Anforderungen an Beweismaterial, oft lange bevor es überhaupt in den sozialen Medien landet.

4 Vgl. *eyewitness.global*. Online unter: <https://www.eyewitness.global/> (letzter Zugriff: 20.03.2024).

5 Vgl. *witness.org*. Online unter: <https://www.witness.org/> (letzter Zugriff: 20.03.2024).