costaid

Coping Strategies Against Information Disorder

Leitfaden für Frontliner

Kofinanziert von der Europäischen Union

Finanziert von der Europäischen Union. Die geäußerten Ansichten und Meinungen sind jedoch ausschließlich die des Autors/der Autoren und spiegeln nicht unbedingt die der Europäischen Union oder der Europäischen Exekutivagentur für Bildung und Kultur (EACEA) wider. Weder die Europäische Union noch die EACEA können für diese verantwortlich gemacht werden.

Die Autoren:

Eliane Smits van Waesberghe & Tim Paulusse - Verwey-Jonker Instituut (Hauptredaktion)

Leen D'Haenens & Joyce Vissenberg - KU Leuven

Tzvetalina Genova - Internationales Institut für Management

Wolfgang Eisenreich - Wissenschaftsinitiative Niederösterreich

Sonja Bercko Eisenreich - Integra-Institut

Alenka Valjašková - QUALED

Pantelis Balaouras - Connexions

Erklärung zum Urheberrecht:

costaid
Coping Strategies Against Information Disorder

Kapitel 3

Technologie und Werkzeuge

Zielgruppe

Diese Leitlinien richten sich an die so genannten "First-Liner". "First-liners" ist ein übergreifender Begriff für alle Personen, die in direktem Kontakt mit Menschen stehen, die durch Informationsstörungen gefährdet sind, wobei der Schwerpunkt auf Gruppen in der beruflichen Bildung liegt. Beispiele für Personen, die unter diesen Oberbegriff fallen, sind: Erzieher, Lehrer, Ausbilder, Jugendberater und -betreuer, Sozialarbeiter und Jugendarbeiter. Diese Liste ist jedoch nicht erschöpfend. Der Anwendungsbereich dieses Projekts umfasst auch andere Personen, die im Bildungs-, Sozial- oder Gesundheitsbereich tätig sind.

3.1 Einleitung

Das Konzept der Fehlinformation (gefälschte Nachrichten) wurde bereits in früheren Kapiteln erörtert. Nachrichten werden in der Regel von professionellen Nachrichtenanbietern verbreitet, darunter öffentlich-rechtliche Medien, kommerzielle Nachrichtenmedien, unabhängiger professioneller Journalismus oder Amateurnutzer (im Falle von Social-Media-Plattformen). Nachrichten sind zugänglich und in verschiedenen Formaten verfügbar:

- **Digitale Nachrichten**: Nachrichten, die über internetbasierte Kanäle im digitalen Medienformat (Text, Bilder, Audio, Video) verbreitet werden.
- **Printmedien**: Zeitungen und Zeitschriften mit text- und bildbasierten Inhalten.
- **Rundfunk**: Fernsehen und Radio mit video- und audiobasierten Inhalten.

In diesem Abschnitt konzentrieren wir uns hauptsächlich auf das erste Format, die digitalen Nachrichten. Die hier geführten Diskussionen können jedoch auch auf andere Formate, die Video, Audio und Bilder verwenden, angewendet werden.

Der Zugang zu digitalen Nachrichten erfolgt überwiegend über Nachrichtenplattformen, die von professionellen Nachrichtenanbietern betrieben werden. Die Nutzer können auf diese Nachrichtenplattformen entweder kostenlos oder über Abonnementdienste zugreifen. Diese Plattformen bieten Web-Feeds in gängigen Browsern und News-Feeds auf Social-Media-Plattformen an, so dass die Nutzer personalisierte Nachrichten-Updates erhalten können.

Ein Newsfeed ist eine Webseite oder ein Bildschirm, der häufig aktualisiert wird, um die neuesten Nachrichten oder Informationen anzuzeigen. Personalisierte Newsfeeds sind in Webbrowser (Webfeeds) und Social-Media-Plattformen integrierte Dienste, die den Nutzern Nachrichten auf der Grundlage ihrer persönlichen Präferenzen liefern. Diese Präferenzen werden durch verschiedene Faktoren bestimmt, z. B. durch das Abonnieren von Webfeed-Kanälen, den Besuch bestimmter Webseiten und vieles mehr.

Auf sozialen Medienplattformen erhalten die Nutzer auch von anderen Nutzern freigegebene Medien, die auch Fehlinformationen enthalten können.

Personalisierte Nachrichten erscheinen auf dem Browser oder der Social-Media-Plattform eines Nutzers durch Dienste und Algorithmen der sozialen Medien. Die Nutzer sind sich möglicherweise nicht vollständig darüber im Klaren, wie diese Algorithmen funktionieren, da unklar bleibt, ob die Auswahl der Nachrichten ausschließlich auf den Präferenzen der Nutzer oder auf anderen Kriterien beruht. Beispielsweise präsentieren Social-Media-Plattformen Nachrichten auf der Grundlage der Nachrichtenanbieter, denen ein Nutzer folgt, der Lesegewohnheiten seiner Freunde oder der kürzlich angeklickten Artikel, die eine Art von Präferenz darstellen können. Folglich wird eine Liste von Nachrichten gefiltert, d. h. es werden nicht alle Nachrichten angezeigt, sondern nur diejenigen, die der Algorithmus für den Nutzer als besonders interessant erachtet. Viele argumentieren, dass dies zu einer Informationsblase führt, in der die Nachrichten von einem Algorithmus ausgewählt werden und möglicherweise keine Nachrichten enthalten, an denen ein Nutzer interessiert ist, die der Algorithmus aber nicht berücksichtigt. Daher ist es wichtig, sich nicht ausschließlich auf

persönliche Newsfeeds zu verlassen, sondern professionelle, vertrauenswürdige Nachrichtenplattformen zu besuchen. Es ist immer wichtig, die Glaubwürdigkeit und Genauigkeit von Nachrichten kritisch zu bewerten, unabhängig vom Format oder der Übermittlungsmethode.

3.2 Suchmaschinen und Algorithmen

Viele Websites im Internet sind bestrebt, ihre Nutzer zu binden und deren Verweildauer auf der Plattform zu maximieren. Dies gilt insbesondere für Social-Media-Plattformen, die verschiedene Strategien anwenden, um das Nutzererlebnis zu verbessern. Eine dieser Strategien ist die Anzeige von Inhalten, die den Interessen der Nutzer entsprechen. Allerdings können Websites keine Gedanken lesen, um die Vorlieben der Nutzer zu erkennen. Deshalb werden Algorithmen eingesetzt, um Nutzerdaten zu analysieren und personalisierte Inhalte zu liefern.

Algorithmen sind komplexe Formeln, die die Inhaltspräferenzen einer Person beobachten, messen und berechnen. Dazu gehören Faktoren wie die Verweildauer des Nutzers bei bestimmten Arten von Videos, die Zeit, die er mit einem bestimmten Beitrag verbringt, oder Engagement-Aktionen wie das Hinterlassen von Likes oder Kommentaren. Durch die Analyse dieser Daten bestimmen die Algorithmen die Art von Inhalten, die die Nutzer über einen längeren Zeitraum beschäftigen. Diese Informationen werden dann verwendet, um ähnliche Inhalte zu kuratieren und den Nutzern zu empfehlen.

Dieser Ansatz mag zwar logisch und harmlos erscheinen, doch die Verwendung von Algorithmen hat auch ihre Schattenseiten. Die Vorliebe des Algorithmus für Inhalte, die der Nutzer interessant findet, kann zu Filterblasen führen, in denen die Nutzer nur bestimmte Sichtweisen zu sehen bekommen. Filterblasen schränken die Vielfalt der Inhalte ein und können zu Echokammern führen oder bestehende verstärken.

Zusätzlich zu Filterblasen können Algorithmen die Extremität von Inhalten verstärken, indem sie zunehmend Nischen-, Rand- und Extrembeiträge empfehlen. Das Ziel ist es, die Nutzer bei der Stange zu halten, aber dies kann dazu führen, dass die Nutzer in Online-Räume ohne unterschiedliche Standpunkte gelenkt werden, was zu einem sogenannten "Kaninchenbau" führt.

Filterblasen und Kaninchenlöcher setzen die Nutzer radikalen Inhalten und den damit verbun-denen Gemeinschaften aus. Diese Online-Räume bieten einen fruchtbaren Boden für die Entwicklung, das Wachstum, die Verzerrung und die Verbreitung von Fehlinformationen und Desinformationen.

Je weiter die Nutzer in den Kaninchenbau vordringen, desto mehr werden extreme Argumente und falsche Informationen zur Normalität. Diese Normalisierung erleichtert die Verbreitung und Akzeptanz von Fehlinformationen und Desinformationen aus radikalen Quellen.

Wenn wir die Auswirkungen von Algorithmen auf die Nutzererfahrungen verstehen, können wir die mit Filterblasen, Echokammern und Kaninchenlöchern verbundenen Risiken besser einschätzen. Dieses Wissen ist unerlässlich, um sich in der Online-Landschaft zurechtzufinden und die Herausforderungen zu bewältigen, die sich aus der Unordnung der Informationen ergeben.

3.3 Online-Strategien von Desinformations- und extremistischen Organisationen

In den vorangegangenen Abschnitten haben wir untersucht, wie Informationsstörungen in Echokammern und Filterblasen entstehen und aufrechterhalten werden. Aber wie gelangen falsche Informationen zu Menschen außerhalb dieser Räume?

Die gebräuchlichsten Methoden

Fehlinformationen und Desinformationen können über unzählige Kommunikationsformen verbreitet werden. Am häufigsten werden sie jedoch über verschiedene Kanäle verbreitet, z. B. über Social-Media-Plattformen, Websites, E-Mail und Mundpropaganda. Die gebräuchlichsten Methoden zur Verbreitung von Falschinformationen stehen in engem Zusammenhang mit den sieben Kategorien problematischer Inhalte, die in *Kapitel 1: Verständnis von "Fake* News" erörtert werden:

- **Satire oder Parodie**: Einige Fehlinformationen und Desinformationen werden zu Unterhaltungszwecken oder als Satire erstellt, können aber als echte Nachrichten missverstanden werden. Satirische Websites oder Konten in sozialen Medien können humorvolle oder übertriebene Geschichten veröffentlichen, aber Leser, die sich ihres satirischen Charakters nicht bewusst sind, können sie für tatsächliche Informationen halten.
- **Clickbait**: Artikel, die Falsch- oder Desinformationen enthalten, verwenden oft sensationelle oder irreführende Überschriften, um Aufmerksamkeit zu erregen und mehr Klicks oder Aufrufe zu generieren. Sie zielen darauf ab, die Neugierde oder die Emotionen der Menschen auszunutzen, um den Verkehr auf eine Website zu lenken und Einnahmen durch Werbung zu erzielen.
- **Falsche Darstellung**: Hierbei werden tatsächliche Nachrichten verzerrt oder falsch dargestellt, indem Fakten selektiv dargestellt oder wichtige Informationen weggelassen werden. Dazu kann es gehören, dass Aussagen aus dem Zusammenhang gerissen, Bilder oder Videos verändert oder die Bedeutung einer Geschichte so verdreht werden, dass sie in ein bestimmtes Narrativ passt.
- **Nachahmung**: Fehlinformationen und Desinformation können auch darin bestehen, dass sich seriöse Nachrichtenquellen oder Personen des öffentlichen Lebens als solche ausgeben, um falschen Informationen Glaubwürdigkeit zu verleihen. Dies kann durch die Erstellung gefälschter Websites oder Social-Media-Konten geschehen, die legitime Quellen imitieren und den Lesern vorgaukeln, die Informationen seien vertrauenswürdig.
- **Politische Manipulation**: Informationsstörungen werden manchmal in der Absicht geschaffen oder verbreitet, die öffentliche Meinung oder Wahlen zu beeinflussen. Dies kann die Verbreitung falscher Informationen über politische

Kandidaten, die Manipulation der öffentlichen Stimmung oder die Ausnutzung bestehender Vorurteile und Spaltungen in der Gesellschaft beinhalten.

- **Fälschung**: Fehlinformationen und Desinformationen können völlig frei erfunden sein und haben keine Grundlage in der Realität. Dabei werden falsche Geschichten, Zitate oder Ereignisse erfunden, um Leser oder Zuschauer in die Irre zu führen.
- **Verstärkung durch soziale Medien**: Soziale Medienplattformen spielen eine wichtige Rolle bei der Verbreitung von Falsch- und Desinformationen. Falsche Geschichten können sich schnell verbreiten, wenn Nutzer sie teilen und weiterverbreiten, oft ohne die Richtigkeit der Informationen zu überprüfen. Die Algorithmen von Social-Media-Plattformen können ebenfalls zur Verstärkung beitragen, indem sie Inhalte auf der Grundlage von Engagement und nicht von Genauigkeit fördern.

Es gibt auch andere gängige Methoden zur Verbreitung von Falsch- und Desinformationen, die ausführlicher behandelt werden sollten: Trolling, Doxxing und Mainstreaming. Dies geschieht in den nächsten beiden Unterkapiteln.

Trolling und Doxxing

Eine häufig angewandte Strategie ist das Trolling. Trolling ist definiert als der absichtliche Einsatz von Unhöflichkeit, Aggression, Täuschung und Manipulation in der Online-Kommunikation, um Konflikte oder Belustigung zu provozieren. Trolle zetteln Online-Konflikte an, indem sie täuschen, manipulieren oder aggressiv sind. Sie bringen Unterhaltungen zu ihrem eigenen Vergnügen zum Entgleisen und betreiben im Grunde digitales Mobbing.

In kleinem Maßstab mag Trolling relativ harmlos erscheinen und als bloßes Ärgernis erscheinen. Wenn jedoch organisierte Gruppen von Trollen ein bestimmtes Ziel verfolgen, kann sich dieses Ärgernis schnell in eine Desinformationsepidemie verwandeln. Ein Beispiel dafür ist Russlands Einsatz von Trollen in sozialen Medien als "Waffe". Russland setzte ein großes Netzwerk von Trollen ein, um weltweit Desinformationen in mehreren Sprachen zu verbreiten, mit dem Ziel, den Online-Diskurs über Russland zu kontrollieren. Diese Trolle verbreiteten nicht nur falsche Informationen, sondern zielten auch auf Social-Media-Nutzer mit Beiträgen ab, die von dem Narrativ abwichen, das sie verbreiten sollten. Infolgedessen hielten sich viele Nutzer sozialer Medien mit Diskussionen über Russland zurück, was es den Trollen ermöglichte, den Diskurs mit ihren Fehlinformationen zu kontrollieren.

Beim Doxxing, einer weiteren Form des Internet-Mobbings, werden persönliche Informationen oder Identitäten von Personen ohne deren Zustimmung online veröffentlicht. Diese Taktik beinhaltet zwar nicht speziell die Verbreitung von Desinformationen, ist aber eine weitere Strategie, die Internet-Trolle anwenden, um die Berichterstattung über ein bestimmtes Thema zu kontrollieren, ähnlich wie beim Trolling. Doxxing kann eingesetzt werden, um Nutzer sozialer Medien einzuschüchtern und ihre Bereitschaft zu unterdrücken, Inhalte zu posten, die dem vom Troll bevorzugten Narrativ widersprechen.

Das Verständnis der Auswirkungen von Trolling und Doxxing ist von entscheidender Bedeutung, um die verschiedenen Taktiken zu erkennen, die zur Manipulation und Kontrolle von Online-Narrativen eingesetzt werden. Diese Strategien tragen nicht nur zur Verbreitung von Falschinformationen bei, sondern stellen auch eine Herausforderung für die Förderung eines offenen und informierten digitalen Umfelds dar.

Mainstreaming

Eine wichtige Strategie zur Verbreitung von Desinformationen und extremistischen Inhalten ist deren Normalisierung oder "Mainstreaming". Die Exposition spielt in diesem Prozess eine entscheidende Rolle. Die Exposition gegenüber Fehlinformationen und Desinformationen kann dazu führen, dass sich die falschen Vorstellungen der Menschen über die jeweiligen Themen hartnäckig halten und in ihren Köpfen normalisiert werden. Diese Exposition kann in verschiedenen Formen erfolgen.

Einflussnehmer

Eine häufige Form der Exposition ist die Verbreitung von Mensch zu Mensch, bei der Einzelpersonen Informationen mit anderen teilen. Dies kann durch persönliche Interaktionen oder in größerem Umfang mit Meinungsmachern in sozialen Medien geschehen. Influencer, die eine große Reichweite über verschiedene Gruppen hinweg haben, können unwissentlich oder absichtlich falsche Informationen verbreiten und damit eine große Anzahl von Personen beeinflussen. Eine solche weit verbreitete Verbreitung führt zu einer Normalisierung von Fehlinformationen in verschiedenen Zielgruppen.

Ironie, Satire und Meme

Extremistische Personen und Organisationen bedienen sich häufig des Humors, der Satire und der Ironie, um ihre Ideen zu verbreiten.

Satire kann ein wirkungsvolles Mittel sein, um unterdrückerische Ideologien herauszufordern, Erzählungen zu verändern oder Nischenansichten innerhalb des Mainstreams zu normalisieren. Im Bereich der Fehlinformation wird Satire auf einem Spektrum eingesetzt. Parodie-Websites wie The Onion oder De Speld veröffentlichen nicht-faktische Inhalte zum Zwecke der Belustigung, ohne die Absicht, die Öffentlichkeit zu täuschen. Bestimmte Personen und Gruppen nutzen Satire und Ironie jedoch in böswilliger Absicht, um den Mainstream-Journalismus und die Wissenschaft zu diskreditieren oder extremistische Ideen und Desinformationen zu verbreiten. Durch den Einsatz von Satire und Humor werden solche Inhalte im politischen Diskurs zugänglicher und akzeptabler, wodurch sie einem breiteren Publikum zugänglich gemacht werden.

Extremistische Inhalte sprechen junge Menschen oft als eine Form der Unterhaltung oder der Sensationssuche an. Junge Menschen, die auf der Suche nach Sinn sind, neigen dazu,

intensive und neuartige Erfahrungen zu machen, was sie anfälliger für extremistisches Gedankengut und die damit verbundene Desinformation macht.

Memes, d. h. weit verbreitete humorvolle kulturelle Inhalte, sind ein weiteres Mittel zur Verbreitung extremistischer Ideologien. Memes gibt es in verschiedenen Formaten, darunter Bilder, Videos, Audioclips, Emojis und Symbole. Während Meme selbst nicht per se schädlich sind, nutzen Extremisten sie, um ihre Ideen zu normalisieren. Der spielerische Charakter von Memen ermöglicht es Extremisten, die Schädlichkeit ihrer Botschaften zu verschleiern, zu entlarven oder zu leugnen. Diese "ausgefallenen" oder provokativen Inhalte werden akzeptabler, und wenn sie mit Vorwürfen des Sexismus, Rassismus oder der Fremdenfeindlichkeit konfrontiert werden, können die Urheber sie leicht als "nur ein Scherz" abtun. Diese Verwischung der Grenzen zwischen spielerischem Unfug und problematischen Inhalten schafft Unklarheit und macht es schwierig, unschuldige Witze von extremistischen Botschaften zu unterscheiden. Pepe der Frosch, eine Internet-Cartoonfigur, die ursprünglich als harmloser Scherz gedacht war, wurde von weißen Rassisten im Internet vereinnahmt. Dies sorgte bei den Internetnutzern für Verwirrung, da sich extremistische Versionen dieses Memes unter die harmlosen mischten. Die Normalisierung extremistischer Inhalte tritt ein, wenn immer mehr Menschen mit diesen Botschaften konfrontiert werden, wodurch die Grenzen zwischen dem, was akzeptabel ist, und dem, was nicht akzeptabel ist, verwischt werden.

3.4 Manipulierte Inhalte

Aus technischer Sicht sind alle Informationen oder "Nachrichten" in den Medien eine Kombination aus Text, Bild, Audio und Video. Das Problem besteht jedoch darin, festzustellen, ob die Informationen authentisch sind oder nicht. Es ist erwähnenswert, dass Fehlinformationen möglicherweise echte Bilder verwenden, aber die Geschichte manipulieren und die tatsächlichen Fakten verfälschen.

In der Vergangenheit wurde allgemein davon ausgegangen, dass jeder einen Text schreiben kann, während man davon ausging, dass Bilder, Audio- und Videodateien mehr oder weniger authentisch sind und nur von Fachleuten verändert werden können. Mit den jüngsten technologischen Fortschritten können jedoch auch Bilder, Audios und Videos manipuliert werden. Dies kann von Fachleuten oder durch Anwendungen mit künstlicher Intelligenz, wie die Deepfake-Technologie, erreicht werden. Daher muss unterschieden werden, ob ein Audio- oder Videomaterial wirklich von einem Mikrofon oder einer Videokamera aufgenommen wurde oder ob es das Ergebnis einer Bearbeitung durch Experten oder von Systemen der künstlichen Intelligenz (KI) ist (generative KI und synthetische Medien: Stimmenklone, Deepfake-Videos). Außerdem sollte es technisch möglich sein, die ursprüngliche Quelle, den Produzenten oder Herausgeber einer Bild-, Audio- oder Videoressource zu identifizieren. Der Grund dafür ist, dass Ressourcen im World Wide Web und in den sozialen Medien mehrfach geteilt, kopiert oder weiterverbreitet werden können. Daher kann es für normale Nutzer schwierig sein, die ursprüngliche Quelle und den Produzenten zu identifizieren, selbst wenn sie den Verdacht haben, dass es sich bei den Nachrichten um Fehlinformationen handeln könnte.

Um die Nutzer in die Lage zu versetzen, zwischen echten und gefälschten Nachrichten zu unterscheiden, sind mehrere Schritte erforderlich. Weitere Informationen dazu finden Sie in *Kapitel 2: Aktionen und Fähigkeiten*. Es wird empfohlen, dieses Kapitel zu lesen, um die komplizierten Details hinter der Erkennung von Falschinformationen zu erfahren. Hier ist jedoch eine kurze, sehr vereinfachte Zusammenfassung:

- **Schritt 1: Förderung des Bewusstseins der Nutzer:** Die Nutzer müssen sich bewusst sein, dass Nachrichten gefälscht sein können. Die Durchführung von Sensibilisierungsmaßnahmen ist von entscheidender Bedeutung, um die Nutzer darüber zu informieren, was gefälschte Nachrichten sind und wie sie sich vor deren Folgen schützen können.
- **Schritt 2: Überprüfen Sie die Zuverlässigkeit der Herausgeber**: Das zunehmende Bewusstsein für gefälschte Nachrichten veranlasst die Nutzer dazu, die Zuverlässigkeit von Nachrichtenquellen und Herausgebern zu hinterfragen. Es ist wichtig, das Medienformat zu berücksichtigen, ob es sich um einen Fernsehsender, eine Zeitschrift, eine Zeitung (online oder gedruckt) oder eine Social-Media-Plattform handelt. Medienkanäle, die eine einfache gemeinsame Nutzung oder Weiterverbreitung von Nachrichten ermöglichen, sind möglicherweise weniger zuverlässig. Umgekehrt sind Medienkanäle, die die Identifizierung und Überprüfung von Quellen erleichtern, in der Regel zuverlässiger.

Was die über das Internet verbreiteten Nachrichten anbelangt, so sollten Anbieter von Diensten wie News Feeds und Social Media Networks neue Technologien nutzen, um die Zuverlässigkeit der Quellen zu überprüfen und den ursprünglichen Herausgeber und die Quelle zu ermitteln. Die Blockchain-Technologie ist eine solche Technologie, die diese Bemühungen erleichtern kann.

Durch die Befolgung dieser Schritte und den Einsatz von Technologien können die Nutzer in die Lage versetzt werden, sich in der digitalen Landschaft zurechtzufinden, echte Nachrichten von Fehlinformationen zu unterscheiden und fundierte Entscheidungen über die Informationen zu treffen, auf die sie stoßen.

Deepfakes verstehen: Synthetische Medienmanipulation

Deepfakes sind nach der Definition des Cambridge Dictionary "*Video- oder Tonaufnahmen, bei denen das Gesicht oder die Stimme einer Person durch die einer anderen Person ersetzt wird, so dass sie echt erscheint*".

In dem Artikel "Deepfake explained" aus dem Jahr 2020 erwähnt die Autorin Meredith Somers, dass "*ein Deepfake sich auf eine bestimmte Art von synthetischen Medien bezieht, bei denen eine Person in einem Bild oder Video mit dem Abbild einer anderen Person ausgetauscht wird*". Außerdem wird erklärt, dass "*der Begriff 'Deepfake' erstmals Ende 2017 von einem Reddit-Nutzer gleichen Namens geprägt wurde. Dieser Nutzer schuf einen Bereich auf der Online-Nachrichten- und Aggregations-Website, in dem er pornografische Videos teilte, die eine Open-Source-Gesichtstausch-Technologie verwendeten*.

Deepfakes finden in verschiedenen Bereichen Anwendung und werden für unterschiedliche Zwecke eingesetzt. Einige bemerkenswerte Beispiele sind:

- **Erpressung:** Deepfakes können verwendet werden, um falsches, belastendes Material zu erzeugen, was zu Erpressung führen kann. Da es zudem immer schwieriger wird, Deepfakes von echten Inhalten zu unterscheiden, können Opfer von Erpressungen behaupten, dass die Beweise gefälscht sind, was ihnen eine glaubhafte Bestreitbarkeit verschafft.
- **Pornografie:** Deepfake-Pornografie hat im Internet erheblich an Bedeutung gewonnen. Einem Bericht des niederländischen Cybersicherheitsunternehmens Deeptrace zufolge waren etwa 96 % aller Online-Depfakes pornografisch.
- **Politik:** Deepfakes wurden eingesetzt, um bekannte Politiker in Videos falsch darzustellen, Desinformationen zu verbreiten und die öffentliche Wahrnehmung zu manipulieren. Beispiele sind Deepfakes mit Barack Obama, Donald Trump, Volodymyr Zelenskyy und Wladimir Putin.
- **Schauspielerei/Filme:** Es gibt Spekulationen über die Verwendung von Deepfakes zur Schaffung digitaler Schauspieler in zukünftigen Filmen. Während digital konstruierte oder veränderte Menschen bereits in Filmen zu sehen waren, könnten Deepfakes zu neuen Fortschritten in diesem Bereich beitragen.

- **Soziale Medien:** Deepfakes werden von Nutzern auf verschiedenen Social-Media-Plattformen verwendet. Einzelpersonen ersetzen Gesichter in beliebten Film- oder Serienszenen durch ihre eigenen und erstellen so personalisierte Videos. Plattformen wie Facebook haben Maßnahmen ergriffen, um Deepfakes zu erkennen und als Fälschungen zu kennzeichnen, so dass sie in den Feeds der Nutzer weniger wichtig sind.

Erkennen und Abschwächen von Deepfakes: technologische Ansätze

Forscher erforschen aktiv Methoden zur Erkennung und Bekämpfung von gefälschten Audio- und Videodaten. Es werden verschiedene Ansätze verfolgt:

- **Algorithmische Erkennung:** Ein Ansatz besteht in der Entwicklung von Algorithmen, die manipulierte Inhalte erkennen können. Diese Algorithmen analysieren verschiedene visuelle und auditive Anhaltspunkte, um Unstimmigkeiten oder Anomalien zu erkennen, die auf einen Deepfake hinweisen. Durch den Einsatz von Techniken des maschinellen Lernens und der künstlichen Intelligenz können diese Algorithmen ihre Erkennungsfähigkeiten mit der Zeit verbessern.
- **Blockchain-Technologie:** Eine andere Technik schlägt vor, die Blockchain-Technologie zu nutzen, um die Quelle der Medien zu überprüfen. Die Blockchain ist ein digitales Hauptbuch, das Transaktionen in einem Computernetzwerk auf sichere, transparente und fälschungssichere Weise aufzeichnet. Sie nutzt Dezentralisierung und Kryptografie, um Vertrauen zu schaffen, ohne dass eine zentrale Behörde erforderlich ist. In diesem Szenario müssten Videos durch ein Blockchain-Ledger überprüft werden, bevor sie auf Social-Media-Plattformen angezeigt werden. Indem sichergestellt wird, dass nur Videos aus vertrauenswürdigen Quellen zugelassen werden, könnte die Verbreitung potenziell schädlicher Deepfake-Medien reduziert werden.
- **Digitale Signaturen:** Einige schlagen vor, alle Videos und Bilder, die mit Kameras, einschließlich Smartphone-Kameras, aufgenommen wurden, digital zu signieren, um Fälschungen zu bekämpfen. Dies würde bedeuten, dass jedem Medium eine eindeutige digitale Signatur zugewiesen wird, so dass jedes Foto oder Video bis zu seinem ursprünglichen Besitzer zurückverfolgt werden kann. Dieser Ansatz kann zwar dabei helfen, die Herkunft von Inhalten nachzuvollziehen, es bestehen jedoch Bedenken hinsichtlich eines möglichen Missbrauchs, z. B. zur Verfolgung von Dissidenten oder zur Verletzung der Privatsphäre.

3.5 Referenzen

Aro, J. (2016). Der Krieg im Cyberspace: Propaganda und Trolling als Mittel der Kriegsführung. *European View*, *15*(1), 121-132. https://doi.org/10.1007/s12290-016-0395-5

Cambridge English Dictionary: Bedeutungen & Definitionen. (2023). https://dictionary.cambridge.org/dictionary/english

Daniels, J. (2018). The Algorithmic Rise of the "Alt-Right". *Contexts*, *17*(1), 60-65. https://doi.org/10.1177/1536504218766547

Egelhofer, J. L., & Lecheler, S. (2019). Fake News als zweidimensionales Phänomen: ein Rahmen und eine Forschungsagenda. *Annals of the International Communication Association*, *43*(2), 97-116. https://doi.org/10.1080/23808985.2019.1602782

Greene. (2019). "Deplorable" Satire: Alt-Right Memes, White Genocide Tweets, and Redpilling Normies. *Studies in American Humor*, *5*(1), 31-69. https://doi.org/10.5325/studamerhumor.5.1.0031

Hardaker, C. (2013). "Uh. . . ich will nicht pingelig sein,,,,,, aber... .the past tense of drag is dragged, not drug." *Journal of Language Aggression and Conflict*, *1*(1), 58-86. https://doi.org/10.1075/jlac.1.1.04har

IED. (2022, 23. August). *Wie funktionieren Social Media Algorithmen*. Institut für die Entwicklung des Unternehmertums. https://ied.eu/blog/technology-blog/how-do-social-media-algorithms-work/

Johnson, D., & Johnson, A. (2023, Juni 15). Was sind Deepfakes? Wie gefälschte KI-gestützte Audio- und Videodateien unsere Wahrnehmung der Realität verzerren. *Business Insider.* https://www.businessinsider.com/guides/tech/what-is-deepfake?international=true&r=US&IR=T

Levy, G., & Razin, R. (2019). Echo Chambers and Their Effects on Economic and Political Outcomes. *Annual Review of Economics*, *11*, 303-328. https://doi.org/10.1146/annurev-economics-080218-030343

Lewis, B., & Marwick, A. E. (2017). Medienmanipulation und Desinformation im Internet. *Data & Society Research Institute.* https://www.posiel.com/wp-content/uploads/2016/08/Media-Manipulation-and-Disinformation-Online-1.pdf

McNealy, J. (2015). Leser reagieren negativ auf die Offenlegung der Identität des Posters. *Newspaper Research Journal*, *38*(3). https://doi.org/10.1177/0739532917722977

Munn, L. (2019). Alt-right pipeline: Individuelle Wege zum Online-Extremismus. *First Monday.* https://doi.org/10.5210/fm.v24i6.10108

Probe, I. (2020, Januar 13). Was sind Deepfakes - und wie kann man sie erkennen? *The Guardian.* https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them

Schumpe, B. M., Bélanger, J. J., Moyano, M., & Nisa, C. F. (2020). Die Rolle der Sensationssuche bei politischer Gewalt: Eine Erweiterung der Significance Quest Theory. *Journal of Personality and Social Psychology*, *118*(4), 743-761. https://doi.org/10.1037/pspp0000223

Seth, S. (2023, September 11). Die 10 größten Nachrichtenmedienunternehmen der Welt. *Investopedia.* https://www.investopedia.com/stock-analysis/021815/worlds-top-ten-news-companies-nws-gci-trco-nyt.aspx

Somers, M. (2020, Juli 21). Deepfakes, erklärt. *MIT Sloan.* https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

Tandoc, E. C., Lim, Z. W., & Ling, R. (2017). Defining "Fake News": A Typology of Scholarly Definitions. *Digital Journalism*, *6*(2), 137-153. https://doi.org/10.1080/21670811.2017.1360143

Van Puffelen, M. (2021). Rechtsextremisme: Geweld met een rechtsextremist motie. In *DSP-groep*. DSP-groep. https://www.dsp-groep.nl/wp-content/uploads/18MP-Rechtsextremisme-DSP-2021.pdf

Van Wonderen, R. (2023). *Rechts-extremistische Radikalisierung auf Social Media Plattformen*. Verwey-Jonker Instituut.

Van Wonderen, R. (2023). *Richtlijn / onderbouwing Radicalisering*. Verwey-Jonker Instituut.

Van Wonderen, R. & Peeters, M. (2021). *Werken aan weerbaarheid tegen desinformatie en eenzijdige meningsvorming. Evaluatie lesprogramma Under Pressure.* Utrecht: Verwey-Jonker Instituut. https://www.verwey-jonker.nl/wp-content/uploads/2022/04/120550_Werken-aan-weerbaarheid-tegen-desinformatie-eenzijdige-meningsvorming.pdf.

Wasike, B. (2022). Wenn der Influencer sagt: "Spring! Wie Influencer Signaling das Engagement mit COVID-19-Fehlinformationen beeinflusst. *Social Science & Medicine*, *315*, 115497. https://doi.org/10.1016/j.socscimed.2022.115497

Wolfowicz, M., Weisburd, D., & Hasisi, B. (2021). Untersuchung der interaktiven Auswirkungen der Filterblase und der Echokammer auf die Radikalisierung. *Journal of Experimental Criminology*, *19*(1), 119-141. https://doi.org/10.1007/s11292-021-09471-0