Modul 9

KI-Manipulation in Social Media

"Zu argumentieren, dass dir das **Recht auf Privatsphäre egal** *ist, weil du nichts zu verbergen hast, ist nichts anderes, als zu sagen, dass dir* **Meinungsfreiheit egal** *ist, weil du nichts zu sagen hast." - Edward Snowden*

Über das Modul

Dieses Modul besteht aus zwei Teilen und behandelt in beiden das Thema Manipulation in Social Media mit unterschiedlichen Schwerpunkten. Die beiden Teile können separat, aber auch aufbauend voneinander verwendet werden. Im ersten Teil geht es um manipulative Algorithmen auf Social Media.

Im ersten Teil des Moduls wird auf den Umgang mit privaten Daten auf Social Media Plattformen eingegangen. Die Schüler*innen setzen sich dabei mit ihrem eigenen Verhalten auseinander und reflektieren die Preisgabe ihrer Daten. In einem weiteren Schritt wird darauf eingegangen, inwiefern diese Daten von Unternehmen genutzt werden, um maßgeschneiderte Werbungen zu generieren. Zuletzt wird im ersten Teil noch auf die Schwierigkeit des Löschens von Daten im Internet eingegangen und Möglichkeiten besprochen.

Im zweiten Teil wird als Ausschnitt zum Thema Verbreitung von Desinformationen im Netz das Phänomen "Deepfake" behandelt. Hierbei soll geklärt werden, worum es sich bei Deepfakes handelt, wie diese entstehen und welche Chancen und Risiken diese mit sich bringen. Es besteht darüber hinaus die Möglichkeit mit Apps wie bspw. Wombo selbst Deepfakes zu erstellen und über die Ergebnisse im Plenum zu diskutieren.

Lernziele

Die Schüler*innen können ...

- ... von ihnen genutzte Social Media Kanäle analysieren und bewerten
- ... Risiken und Gefahren in digitalen Umgebungen kennen, reflektieren und berücksichtigen
- ... Strategien zum Schutz entwickeln und anwenden
- ... ein Verständnis für den Wert von Daten entwickeln und verstehen, dass sie vermeintlich kostenfreie Angebote im Internet mit ihren Daten bezahlen
- ... ihre eigenen Datenspuren analysieren
- ... die Absichten und Strategien hinter Deepfakes erklären
- ... die Gefahren von Fake News in Form von Deepfakes für Demokratie und Gesellschaft einschätzen
- ... Deepfakes als besondere Form der ***KI***-basierten Desinformation erkennen und beschreiben

Agenda - Teil 1: Manipulation durch Algorithmen

Zeit	Inhalt
15 min	Einstiegsfragen in Gruppen und im Plenum besprechen
15 min	Gedankenexperiment + Video
5 min	Social Media Daten Quiz
5 min	Theorie - Algorithmen und die DSGVO
40 min	Rechercheaufgabe - Zeig mir deinen Account und ich sag dir, wer du bist
15 min	Theorie - Das Internet vergisst nie
15 min	Brainstorming - Kann man Bilder wieder endgültig aus dem Netz löschen?

Agenda - Teil 2: Deep Fakes

Zeit	Inhalt
15 min	Quiz - Deepfake oder real?
10 min	Brainstorming - Was sind Deepfakes?
20 min	Recherche - Funktionsweise von Deepfakes
10 min	Wie kann man manipulative Deepfakes entlarven?
20 min	Aufgabe - Eigene Deepfakes erstellen

Einleitung

Manipulation durch Daten auf Social Media

Dieses Modul beginnt mit dem bekannten Zitat des ehemaligen US-Geheimdienstmitarbeiters Edward Snowden, welcher durch das Veröffentlichen von Überwachungsstrategien der USA und Großbritannien in kürzester Zeit große Aufmerksamkeit auf sich zog. Dieses Zitat bringt den Kern der Lernziele dieses Moduls auf den Punkt: im Mittelpunkt steht das Recht auf Privatsphäre im digitalen Raum und wie achtlos wir mit unseren privaten Daten in diesem öffentlichen Raum umgehen. Im Internet werden all unsere Aktivitäten akribisch dokumentiert und ergeben ein detailliertes Profil über unsere Interessen, unser Konsumverhalten und die Kontakte, die wir pflegen.

Als Start in das Unterrichtsmodul können die Schüler*innen einen Fragenkatalog in Gruppen besprechen und die Ergebnisse festhalten, um sie im Plenum zu präsentieren (bspw. durch ein Plakat oder durch eine Padlet-Seite).

Folgende Fragen eignen sich dabei als Input:

- Welche Apps und Websiten benutzt du regelmäßig?
- Wissen deine SocialMedia-Anbieter (z.B. Instagram) deinen richtigen Namen?
- Können bei deinen SocialMedia-Profilen alle Menschen deine Bilder sehen oder nur deine Kontakte?
- Warum bist du bei SocialMedia-Plattformen aktiv?
- Hast du schon mal eine Datenschutzerklärung gelesen? Interessiert es dich, welche Daten von dir gesammelt werden?
- Hast du dich schon mal selbst gegoogelt? Falls ja, hat dich etwas überrascht? Warum?

Da es sich hierbei um persönliche Fragen handelt, sollten die Schüler*innen selbstständig wählen können, wie viel sie in diesen Gruppengesprächen über sich selbst Preis geben wollen.

Oft läuft das Sammeln der Daten nicht offensichtlich ab, wodurch uns die Konsequenzen unseres Verhaltens im digitalen Raum gar nicht oder nur weniger bewusst sind.

Als Beispiel kann hier von der Lehrperson folgendes Szenario beschrieben werden: Die Schüler*innen sollen sich vorstellen, dass alle Leute, denen sie auf der Straße begegnen sofort wissen, wo sie kurz zuvor waren, wo sie ihre Kleidung gekauft haben, was sie gerade gegessen haben oder was sie zuletzt gegoogelt haben.

- Wie würden sich die Schüler*innen dabei fühlen?
- Wie bewusst ist ihnen, welche Informationen sie im Netz veröffentlichen und wer darauf Zugriff hat?

Als visuelles Beispiel kann in dieser Unterrichtsphase auch das **Video** von Tomatolix hergezeigt werden. In diesem Video werden einerseits Möglichkeiten gezeigt, wie man selbst herausfinden kann, welche Daten Internetplattformen speichern, andererseits versucht der gezeigte Reporter nur mithilfe von Social Media Posts eine Person im realen Leben ausfindig zu machen (und schafft das auch innerhalb kürzester Zeit). Die entsprechende Sequenz beginnt im Video bei Minute 7:00 und dauert zirka 8 Minuten. Auch hier können im Anschluss die Schüler*innen gefragt werden, wie sie sich dabei fühlen und ob sie glauben, dass es mit ihren Profilen auch möglich wäre, sie einfach aufzuspüren. Interessant ist unter anderem das Feature "Snap Map" von Snapchat und die damit einhergehenden Möglichkeiten und Risiken anzusprechen, da hierbei der Livestandort der Nutzer*innen angezeigt wird. Einerseits können dadurch entsprechende Filter vorgeschlagen werden (bspw. ein Wien-Filter, wenn man sich in Wien befindet) und die Standort-Informationen für Freund*innen interessant sein, andererseits bietet diese Funktion dadurch einen Nährboden für Stalking.

Nachdem die Schüler*innen sich oberflächlich im Gespräch mit ihrem Nutzungsverhalten von Social Media Plattformen ausgetauscht haben, kann ein kurzes Quiz im Plenum gespielt werden. Die Schüler*innen sollen dabei erraten, welche Daten Plattformen wie Instagram, Snapchat und TikTok sammeln. Das Quiz kann entweder mithilfe des Arbeitsblattes bearbeitet werden oder man überträgt das Quiz auf eine Plattform wie kahoot oder bespricht die Vermutungen der Schüler*innen direkt im Plenum.

die Frage in den Raum gestellt werden, warum sie all diese Daten sammeln? Welchen Nutzen haben sie davon? Social Media Plattformen verwenden unsere Daten, um sie Firmen anzubieten, damit sie zielgruppengerechte Werbung schalten können. Dadurch, dass Instagram oder TikTok genau wissen, dass du weiblich und zwischen 15 und 25 Jahre alt bist, dich für Sport und Nachhaltigkeit

interessierst, werden dir dann Werbeanzeigen wie bspw. Bambus-Yogamatten oder biologische Sportnahrung ohne Zusätze angezeigt.

Social Media Plattformen wollen also nicht primär Menschen miteinander vernetzen, sondern sehen das lediglich als Mittel zum Zweck. Vielmehr geht es darum persönliche Daten zu nutzen, damit andere Unternehmen effiziente Werbung schalten können. Auch wenn diese Websites also am ersten Blick kostenlos sind, zahlen wir passiv. Wir zahlen mit unserer Zeit und Aufmerksamkeit, die die Websites nutzen, um Daten über uns zu sammeln und diese für Werbung von Firmen weiterzuverkaufen.

Die Schüler*innen sollen nun kurz reflektieren, aber auch auf ihren Apps nachsehen, welche Daten sie öffentlich preisgeben (Geburtsdatum, Wohnort, Geschlecht, Beziehungsstatus...) und wie wohl sie sich damit fühlen, dass diese Informationen weitergeben und teils auch öffentlich sind. Infolgedessen, kann darauf hingewiesen werden, dass durch das Anpassen der Privatsphäre-Einstellungen einiges verbessert werden kann. Die Schüler*innen können dafür selbstständig recherchieren, welche Daten sich auf Privat stellen können und wollen. Im Plenum kann besprochen werden, welche Möglichkeiten man hat, um seine persönlichen Daten zu schützen.

Die DSGVO und was sie bedeutet

Was Datenschutz grundsätzlich bedeutet, wird von Land zu Land unterschiedlich definiert. Das kann zu Problemen führen, da es im Internet wiederum keine Landesgrenzen gibt und persönliche Daten häufig durch mehrere Länder wandern, bis sie ihr Ziel erreicht haben. Aus diesem Grund gibt es aber internationale Abkommen wie beispielsweise das "EU-US Privacy Shield". Dieses Abkommen regelt den Datenaustausch zwischen den USA und Europa. Innerhalb Europas gibt es seit dem 25. Mai 2018 die Datenschutzgrundverordnung (kurz "DSGVO"), welche die Verarbeitung personenbezogener Daten durch Behörden und Unternehmen regelt. Durch die DSGVO werden Firmen verpflichtet, den Nutzer*innen Auskunft über ihre gespeicherten Daten zu geben. Sollte dies nicht umgesetzt werden, können hohe Bußgelder verhängt werden.

Weitere Informationen und spannende Erklärvideos zur DSGVO findet man auf der Website: **deinedatendeinerechte.de**

Algorithmen wissen, was du brauchst

Für diese scheinbar maßgeschneiderten Werbungen werden Algorithmen verwendet, die durch den Browserverlauf und den Eckdaten der Nutzer*innen genau angepasst werden. Werbungen können zielgruppengenau erstellt werden und passen somit genau zu den Interessen und zuvor betrachteten Inhalten der Nutzer*innen. Diese Empfehlungs-Algorithmen können jedoch eine komplett einsichtige Weltsicht bei den Nutzer*innen erzeugen und bis zu einer Radikalisierung der Nutzer*innen führen.

Es ist darum wichtig, dass wir als Nutzer*innen Verzerrungen durch einseitiges Angebot erkennen und diesen auch misstrauen. Ein weiterer wichtiger Schritt ist das Transparent-machen von Algorithmen, ethische Richtlinien und eine Analyse von bisher eingesetzten Algorithmen. Eine NGO, die diese Forderungen vorantreibt, wäre z.B. AlgorithmWatch.

Material

- Social Media - Introduction.pdf

Referenzen

1. Video: Das weiß das Internet über dich! - Selbstexperiment
2. Deine Daten Deine Rechte

Zeig mir deinen Account und ich sag dir, wer du bist

Die Schüler*innen haben in ersten Schritten bereits reflektiert, welche Daten sie öffentlich preisgeben und wie wohl sie sich damit fühlen. Nun sollen im Weiteren in Kleingruppen die Empfehlungen, Vorschläge und Werbungen der Profile analysiert werden und nur anhand dessen eine Charakterisierung des*der User*in erstellt werden. Wichtig ist hierbei, dass niemand zu etwas gezwungen werden soll und nur die Profile analysiert werden, die sich dabei wohl fühlen.

In einem weiteren Schritt können die Schüler*innen nun selbstständig die gespeicherten Daten ihrer Social Media Accounts anfragen und diese analysieren. Für die häufiger genutzten Plattformen Instagram, YouTube und TikTok findet sich im Anhang eine Anleitung, wie man diese Daten jeweils anfordern kann. Grundsätzlich findet man aber auch im Internet und in den FAQs der Plattformen detaillierte Anleitungen. Das Anfragen der Daten kann zwischen einigen Stunden bis hin zu einigen Tagen dauern. Es wäre also empfehlenswert, die Daten bereits in der Stunde zuvor anzufordern.

Was weiß Instagram, TikTok und Co. über mich? Analyse der eigenen Daten in Kleingruppen (2-4 SuS)

Man hat über Instagram, Facebook und Co. aufgrund der DSGVO die Möglichkeit alle von sich gespeicherten Daten anzufordern. Doch weißt du, welche Daten diese Seiten tatsächlich haben?

Teil 1 - Analysiert die Werbung, Empfehlungen, etc. eurer Profile

1. Wählt in eurer Kleingruppe eine oder mehrere Plattformen aus, auf denen ihr angemeldet seid. Ihr könnt ein Profil gemeinsam analysieren oder mehrere Profile hintereinander. Sprecht euch aber immer in der Gruppe ab und geht respektvoll miteinander um! Niemand muss seine Profile herzeigen, wenn eure Mitschüler*in das nicht möchte!
2. Im Folgenden versucht ihr die Besitzer*innen der Profile anhand der Inhalte, die euch präsentiert werden, zu charakterisieren. Achtet dabei weniger, auf

die Dinge, die der Person selbst geliked hat, sondern auf jene, die ihr vorgeschlagen werden.
3. Klickt euch dafür durch den Feed der App. Welche Werbungen werden euch angezeigt? Welche Videos oder Profile werden euch vorgeschlagen?
4. Notiert euch mögliche Eigenschaften, Interessen, Hobbys usw., die die Besitzer*innen der Profile haben könnten.
5. Wie sehr ähneln die Charakterisierungen den tatsächlichen Profilbesitzer*innen? Was hat die Profilbesitzer*innen überrascht? Überlegt euch, warum manche Werbungen oder Vorschläge eurer Meinung nach nicht zur Person passen, aber trotzdem vorgeschlagen werden.

Teil 2 - Fordert eure Daten an

Anleitung Instagram

1. Gehe auf dein eigenes Profil.
2. Tippe auf das Rädchensymbol, um zu den Einstellungen zu gelangen. Wähle dort Privatsphäre und Sicherheit aus.
3. Wenn du runterscrollst, taucht die Rubrik Daten-Download auf. Tippe auf "Download anfordern".
4. Gib die E-Mail-Adresse, die du auch für deinen Account verwendest, ein.
5. Gib dein Instagram-Passwort ein.
6. In den nächsten 48 Stunden solltest du deinen Bericht erhalten.
7. Lade die Informationen aus der Mail runter und entpacke den Ordner.
8. Wenn du auf die Datei "index.html" klickst, wirst du auf die Instagram-Website weitergeleitet, in der du dich durch alle vom Unternehmen gespeicherten Informationen durchklicken kannst.

Anleitung YouTube

1. Öffne dein Profil rechts oben und wähle "Meine Daten auf YouTube" aus.
2. Auf YouTube kannst du deine Ergebnisse direkt auf der Website einsehen.

Anleitung TikTok

1. Tippe in der TikTok-App am unteren Rand auf Profil.
2. Tippe am oberen Rand auf die Schaltfläche Menü.
3. Tippe auf Einstellungen und Datenschutz.

4. Tippe auf Konto verwalten und dann auf Deine Daten herunterladen.
5. In den nächsten 3 Tagen solltest du deinen Bericht erhalten.

Teil 3 - Analysiert eure Berichte

Schaut durch die Berichte.

Was überrascht euch? Wovon habt ihr gewusst, dass diese App, diese Informationen speichert?

Durch diese Selbstversuche wird deutlich, wie viele Daten kontinuierlich produziert werden. Bis vor einem Jahrzehnt geschah das hauptsächlich durch PCs. Mittlerweile werden nicht nur durch zusätzlich durch Smartphones und deren Apps stetig Daten generiert, auch neue Technologien im Bereich der Internet of Things (IoT) wie Wearables (bspw. zum Messen der Pulsfrequenz un1d von Bewegungsprofilen) oder kleine Umwelt-Messstationen (bspw. zur Messung der Luftqualität) produzieren eine unfassbare Menge an Daten. Im Jahr 2020 werden pro Nutzer*in 1 GB an Daten pro Tag produziert! [1]

Auf Instagram werden beispielsweise mehr als 1000 Fotos pro Sekunde hochgeladen, das sind ungefähr 100 Millionen Bilder pro Tag! [2]

Material

- Social Media - Introduction.pdf
- Social Media - Worksheet Research.pdf

Referenzen

1. Wirtschaftsforum.de: Datenverbrauch
2. Instagram Statistiken

Das Internet vergisst nie!

Diese Daten verschwinden auch nicht einfach aus dem Internet. Zusätzlich sind viele dieser Inhalte unverschlüsselt zugänglich. Um zu zeigen, dass Daten im Netz, egal ob verschlüsselt oder unverschlüsselt sich für immer außerhalb der eigenen Kontrolle befinden, kann mit den Schüler*innen mithilfe der **Wayback-Machine** eine kleine Zeitreise unternommen werden. Die Schüler*innen können hierbei die eigene Schulhomepage, Vereinswebsites oder Blogs erkunden. Im Anschluss sollte darüber diskutiert werden, welche Vor- und Nachteile solche Internetarchive mit sich bringen und was man als Nutzer*in trotzdem versuchen kann, wenn unangenehme Inhalte ins Netz gelangen.

Sollten private Inhalte wie bspw. Nacktbilder ins Internet gelangen, muss das niemand einfach so hinnehmen. Denn auch wenn eine endgültige Löschung äußerst schwierig ist, lohnt es sich trotzdem dagegen vorzugehen! Auf Social Media Plattformen können die Urheber*innen kontaktiert werden. Sollten diese die Inhalte nicht entfernen, können sie, falls sie gegen die Richtlinien verstoßen (Nacktheit, Gewalt etc.), direkt gemeldet werden. In anderen Fällen können entsprechende Moderator*innen und die Seitenbetreiber*innen kontaktiert werden. Auch bei anderen Websites und Blogs sollte man zuerst die Inhaber*innen der Seite kontaktieren und um ein Löschen der Inhalte bitten. Sollte der Aufforderung nicht nachgekommen werden, kann bei der Polizei eine Anzeige erstattet werden. Dabei kann sich unter anderem auf das "Recht am eigenen Bild" (§ 78 UrhG) oder auf "Pornografische Darstellung Minderjähriger" bei Nacktbildern von unter 18-jährigen bezogen werden. [3]

Für weitere Informationen eignet sich die Website: **www.ombudsstelle.at**

Material

- Social Media - Introduction.pdf
- Social Media - Worksheet.pdf

Referenzen

1. saferinternet.at

Teil 2: Deep Fakes

In diesem Teil des Moduls werden aus Zeitgründen hauptsächlich Video-Deepfakes behandelt, obwohl diese nur einen kleinen Teil im Bereich der digitalen Desinformation und Fake News ausmachen. Weitere spannende, vertiefende Aspekte zum Thema Manipulation in Social Media wären zum Beispiel die Funktion von Social Media Trolls oder Social Bots.

Was sind Deepfakes?

Deepfakes sind manipulierte oder künstlich erzeugte Ton- oder Bildmedien, welche jedoch echt wirken. Sie zeigen Menschen, die scheinbar etwas sagen oder tun, was sie jedoch noch nie gesagt oder gesagt haben. Deepfakes werden mit künstlicher Intelligenz, wie beispielsweise dem maschinellen Lernen und Deep Learning erzeugt.

Durch neue technologische Entwicklungen im Bereich der Bildbearbeitung und -manipulation wirken auch Deepfakes immer authentischer. So wurden einerseits in der Computer-Vision Algorithmen entwickelt und verbessert, welche Gesichtsstrukturen automatisch erkennt und mappt (bspw. die Position von Augenbrauen und Nase), wodurch neuartige Technologien in der Gesichtserkennung entstanden. Andererseits entstehen durch den Siegeszug des Internets - und insbesondere durch Plattformen, auf denen Bilder und Videos geteilt werden - ein unfassbar großer Datenpool mit audiovisuellem Datenmaterial, der dafür verwendet werden kann.

Zwei spezifische ***KI***-Ansätze finden sich häufig in Deepfake-Programmen: Generative Adversarial Networks (GANs) und Autoencoder. GANs sind maschinelle Lernalgorithmen, die eine Reihe von Bildern analysieren und dadurch neue Bilder mit vergleichbarer Qualität erstellen können. Autoencoder können hingegen Informationen über Gesichtsstrukturen aus Bildern extrahieren und diese Informationen verwenden, um einen neuen Gesichtsausdruck zu modellieren.

Da durch diese Techniken sowohl Mimik und Bewegungsarten einer Person realistisch nachgestellt werden können, ist es mittlerweile sehr schwierig zu erkennen, ob man einen Deepfake oder das Original vor sich hat. Jedoch kann nicht nur die Mimik eines vorhandenen Gesichts verändert werden: Gesichter können ausgetauscht komplett neu entstehen.

In Film und Kino gibt es bereits seit vielen Jahren überzeugende Computer Generated Imagery (CGI) Technologien. Der seltsame Fall des Benjamin Button (Originaltitel: The Curious Case of Benjamin Button) wurde beispielsweise 2009 bei den Oscars für Best Visual Effect ausgezeichnet. Dabei wurde Brad Pitt als Protagonist des Films mithilfe von computerbasierten Manipulationen umgekehrt gealtert.

Dabei sind Manipulation der Medien und Bildbearbeitung keineswegs neue Phänomene. Deepfakes sind sozusagen nur eine technologische Weiterentwicklung eines viel älteren Phänomens. Durch das Aufkommen von Social-Media-Plattformen und dem regen Austausch und Teilen von Inhalten (und somit auch falschen Inhalten, bspw. Fake News) wurde die Medienlandschaft maßgeblich verändert. Zusätzlich bieten Apps wie Snapchat, Instagram und TikTok bereits niederschwellige Filter innerhalb der Anwendungen an, mit denen Gesichter verändert und Videos bearbeitet werden können.

Darüber hinaus ist der Anstieg visueller Medien, besonders Videos, als Kommunikationsmittel ebenfalls von Bedeutung. Visuelle Medien gelten als besonders effizienter Weg, um Informationen zu verbreiten. Bisher war durchaus bekannt, dass Fehlinformationen in Texten platziert werden oder Fotos manipuliert werden, Video hingegen galten für viele noch als harte Belege, welche nur schwer zu fälschen sind.

Durch Deepfakes können Falschinformationen verbreitet werden, User*innen können teilweise nicht mehr zwischen Wahrheit und Fiktion unterscheiden. Viele dieser Meldungen werden bewusst erstellt, um in irgendeiner Form Schaden anzurichten. Durch die Verbreitung von Deepfakes kommt es zu einer Unsicherheit unter Internetnutzer*innen: Was entspricht der Wahrheit und ist ein Fakt? Welchen Medien kann man in diesem Fall noch vertrauen und wer manipuliert seine Inhalte? Durch die alleinige Existenz von Deepfakes sind sich viele User*innen nicht mehr sicher, welchen Inhalten sich noch vertrauen können. [4]

Deepfake-Technologien können für eine Vielzahl von Zwecken verwendet werden mit sowohl positiven als auch negativen Auswirkungen. Deepfakes können so z.B. im Bereich von audiovisuellen Medienproduktionen hilfreich sein (z.B., wenn eine Schauspieler*in ausfallen sollte), Mensch-Maschine-Interaktionen können besser ablaufen, aber auch in Bereichen wie Videokonferenzen, Satire- und Kunstprojekten oder der chirurgischen Gesichtsrekonstruktion einen Platz finden. Jedoch gibt es auch eine Vielzahl an negativen Aspekten, wie bspw. Erpressung, Diffamierung, Mobbing, Identitätsdiebstahl, Rufschädigung, Manipulation von

Nachrichtenmedien, Vertrauensverlust in Wissenschaft, Wirtschaft und Politik, Manipulation von Wahlen, Schaden für internationale Beziehungen und für die nationale Sicherheit.

Das Entlarven von Deepfakes kann unter Umständen sehr langwierig sein, wodurch scheinbar kleine Videos zu großen Problemen führen können. Im Laufe der Unterrichtseinheit sollen die Schüler*innen selbstständig Beispiele finden, wem diese Technologien nützen und wem sie schaden können. Durch das Durchspielen von Beispielen kann erahnt werden, welche Ausmaße ein manipulativer Deepfake annehmen kann.

Ein mögliches, fiktives Beispiel über die Auswirkungen von Deepfakes

Auf Instagram und auf Twitter wurde ein scheinbar echtes Video von einem Politiker hochgeladen, wie er vor laufender Kamera gestand, Millionen von Euro hinterzogen zu haben. Mit diesem Video wird nicht nur der Ruf des Politikers geschädigt und psychologischer Schaden zugefügt. Es könnte beispielsweise der Politiker oder die Partei erpresst werden, indem mit weiteren gefälschten Geständnissen gedroht wird. Die Wähler*innen verlieren Vertrauen in die Partei und werden sie bei der nächsten Wahl nicht mehr wählen. Dieses Misstrauen kann so weit gehen, dass generell dem System nicht mehr vertraut wird.

> Als weiteren Input über das Erkennen von Deepfakes bietet sich ein Video von funk.net (ein Angebot von ARD und ZDF) an: **www.funk.net - Erkennst du die Fälschung?**

Im Plenum kann schließlich besprochen werden, wie man gefährliche Deepfakes verhindert werden können. Mögliche Ansätze wären zum Beispiel: keine Videos der eigenen Person ins Netz stellen, Sprachnachrichten vermeiden, sich nicht ungewollte aufnehmen lassen und auf das Löschen der Fotos/Videos bestehen, strenge Gesetze bzg. Deepfakes, strengere Kontrollen (besonders auf Social Media Plattformen), um die Verbreitung einzudämmen.

Aus gesetzlicher Perspektive gibt es bisher keine konkreten Maßnahmen oder Gesetze. Es werden jedoch bereits Strategien entwickelt, wie bspw. Aktionsplan Deepfake der österreichischen Bundesministerien. Dabei werden jedoch keine konkreten Änderungen in der Gesetzeslage gefordert, sondern eine

Sensibilisierung der Bevölkerung und der Einsatz von Softwaretools, welche Deepfakes erkennen sollen und Fact-Checker-Plattformen. [5]

Material

- Social Media - Deepfakes.pdf
- Social Media - Worksheet Deepfakes.pdf
- Social Media - Worksheet Research Deepfakes.pdf

Referenzen

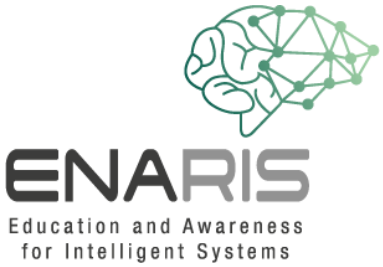
1. saferinternet.at
2. Aktionsplan Deepfake
3. Wie funktionieren Deepfake-Videos? | alpha Lernen erklärt Medienkompetenz
4. Fake videos of real people - and how to spot them (Englisch)
5. ganz konkret: Deepfakes gegen Fakten? | Zeit für Politik
6. Deepfakes: Is This Video Even Real? | NYT Opinion (Englisch)
7. Selbstversuch: Wie macht man ein Deepfake-Video? | Selbstversuch | BR24
8. Täuschung mit Deepfakes | Odysso - Wissen im SWR

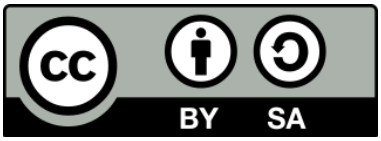
Wir erstellen eigene Deepfakes

Im letzten Schritt dürfen die Schüler*innen selbstständig mithilfe von Apps versuchen, überzeugende Deepfakes zu erstellen. Eine hilfreiche App ist dafür die Bildmanipulations-App Wombo, mit deren Hilfe man Selfies zum Singen bringen kann. Eine weitere mögliche Apps dafür wäre Reface (Achtung: der kostenpflichtige Pro-Modus muss zu Beginn oben auf dem X weggeklickt werden)

Material

- Social Media - Deepfakes.pdf
- Social Media - Worksheet Deepfakes.pdf


ENARIS
Education and Awareness
for Intelligent Systems


BY SA


EUROPEAN UNION


Interreg
Austria-Hungary 2014-2020
European Union – European Regional Development Fund


TU
Graz


mobilis
interaktív kiállítási központ


OESTERREICHISCHE
COMPUTER GESELLSCHAFT
AUSTRIAN
COMPUTER SOCIETY