

Beigabe: Interaktive Lehreinheit

## Deepfakes in verschiedenen Anwendungskontexten und deren ethische und gesellschaftliche Folgen

Cora Bieß

*Unter Mitarbeit von Maria Pawelec.*

Internationales Zentrum für Ethik in den Wissenschaften (IZEW), Universität Tübingen

### Vorbemerkungen

Die hier vorgestellten Lehrmaterialien sind eine Beigabe zur Monografie *„Deepfakes: Technikfolgen und Regulierungsfragen aus ethischer und sozialwissenschaftlicher Perspektive“* von Maria Pawelec und Cora Bieß, erschienen 2021 im Nomos-Verlag in der Reihe *„Kommunikations- und Medienethik“*. Die Lehrmaterialien entstanden im Rahmen des Abschlussworkshops des Projekts *„TANGO: Technik-Innovation: sozialwissenschaftliche und ethische Analysen zu Governance“*, das von 2019 bis 2020 am Internationalen Zentrum für Ethik in den Wissenschaften (IZEW) der Universität Tübingen durchgeführt wurde (gefördert vom Bundesministerium für Bildung und Forschung). Im Rahmen des Abschlussworkshops, der auf Grund der Corona-Pandemie virtuell stattfand, wurden die vorliegenden Positionspapiere in einer interaktiven Einheit von den Teilnehmenden aus Politik, Wissenschaft, Wirtschaft und Zivilgesellschaft diskutiert.

Es handelt sich somit um didaktische Materialien, die ursprünglich für einen professionellen Kontext entwickelt und von Expert\*innen diskutiert wurden. Der erfolgreiche Einsatz dieser Materialien in universitären Lehrveranstaltungen hat jedoch auch gezeigt, dass die Lehrmaterialien ebenfalls sehr gut von Studierenden angenommen werden. Sie können somit im Rahmen der universitären und Erwachsenenbildung zu einer vertieften Auseinandersetzung mit dem Thema Deepfakes und damit zu einer gestärkten Medienkompetenz beitragen. Die Autorinnen können sich darüber hinaus vorstellen, dass die Materialien – gegebenenfalls in leicht abgewandelter Form – auch im Rahmen der Schulbildung etwa ab Klasse 10 eingesetzt werden könnten.

Erstellt wurden die Lehrmaterialien im Oktober 2020. Aufgrund der rasanten technischen Entwicklung der Deepfake-Technologie (und entsprechender Detektionssysteme) sowie regulatorischen Entwicklungen (etwa im Bereich der Plattformregulierung) muss dies vor der Verwendung der Materialien stets berücksichtigt und diese gegebenenfalls an die veränderten Rahmenbedingungen angepasst werden.

Die Methode wird im vorliegenden Text in zwei Varianten vorgestellt 1) als virtuelle Methode (etwa im Rahmen von Online-Seminaren oder virtuellen Workshops) und 2) als präsenzbasierte Gruppenmethode.

Didaktische Ziele der Methode sind:

- Kennenlernen der vielfältigen Anwendungsbereiche von Deepfakes
- Reflexion (und Abwägung) der jeweils unterschiedlichen ethischen und gesellschaftlichen Implikationen von Deepfakes in unterschiedlichen Anwendungskontexten
- Reflexion über Regulierungsmöglichkeiten von Deepfakes

## 1) Virtueller Gesprächskreis

### **Vorbereitung:**

Vorab wird den Teilnehmenden jeweils eine Rolle zugeschickt, die sie während der Onlinesitzung im interaktiven Gesprächskreis über Deepfakes einnehmen werden. Als Rollenbeschreibung dienen die Positionspapiere, die zentrale Informationen über den jeweiligen Charakter und deren Positionen enthalten. Jeder Charakter repräsentiert einen Anwendungskontext von Deepfakes, den die Teilnehmenden wiederum im Gesprächskreis repräsentieren.

Zu verteilende Rollen:

- Vertreter\*in einer Frauenrechtsbewegung
- CSR-Beauftragte\*r von Facebook
- Cybersecurityexpert\*in der Bundeswehr
- Wissenschaftler\*in, die/der an synthetischer Stimmerzeugung arbeitet
- Politische\*r Künstler\*in und Dokumentarfilmer\*in
- CEO in der Unterhaltungsindustrie
- Mitarbeiter\*in einer Nachrichtenagentur
- Journalist\*in von Postillon
- Ermittler\*in im Bereich Kinderpornographie
- Museumspädagog\*in

Die Rollenbeschreibungen (in der Form von Positionspapieren) befinden sich als Druckvorlage im Anhang.

### **Ablauf:**

Zu Beginn der Onlinesitzung kann den Teilnehmenden optional zehn Minuten Zeit gegeben werden, um sich ihre Positionspapiere erneut durchzulesen und sich mit ihrem Charakter vertraut zu machen.

Währenddessen werden von den Multiplikator\*innen Break Out-Sessions eingerichtet. Wichtig ist hierbei darauf zu achten, dass in jedem Raum jede Rolle nur einmal vergeben wird. Im Idealfall werden zwei Break Out-Sessions mit jeweils fünf Personen erstellt. Einer möglichen Gruppeneinteilung nach sind in Gruppe 1 dann vertreten:

1. CSR-Beauftragte\*r von Facebook
2. Cybersecurityexpert\*in der Bundeswehr
3. Wissenschaftler\*in, die/der an synthetischer Stimmerzeugung arbeitet
4. Politische\*r Künstler\*in und Dokumentarfilmer\*in
5. Journalist\*in von Postillon

In Gruppe 2 sind dann vertreten:

6. Vertreter\*in einer Frauenrechtsbewegung
7. CEO in der Unterhaltungsindustrie
8. Mitarbeiter\*in einer Nachrichtenagentur
9. Ermittler\*in im Bereich Kinderpornographie
10. Museumspädagog\*in

**Tipp:** Die Aufteilung kann von Moderator\*innen je nach Gruppengröße individuell angepasst werden. Wenn Sie weniger Teilnehmende haben, wählen Sie vorab aus, welche Rollen Sie verteilen. Wenn Sie mehr Teilnehmende haben, können Sie die Anzahl der Break Out-Räume erhöhen, sodass mehrere Kleingruppen parallel diskutieren.

- Teilen Sie den Teilnehmenden mit, dass nun zwei Runden eines virtuellen Gesprächskreises folgen werden. Bevor Sie die Teilnehmenden in die Break Out-Räume schicken, schreiben Sie entweder auf ein Whiteboard, auf eine PowerPoint Folie, oder im Chat die zu diskutierende Aussage: *„Deepfakes sind eine große Chance für die Gesellschaft.“*
- Nun werden die Teilnehmenden eingeladen, den Break Out-Räumen beizutreten (Minimum 15 Minuten). Im Break Out-Raum diskutieren die Teilnehmenden in der Gruppe aus der jeweiligen Sicht ihres Charakters/ ihrer Rolle die Aussage: *„Deepfakes sind eine große Chance für die Gesellschaft.“*
- Es folgt die Zusammenführung im Plenum: Nachdem die Break Out-Session beendet wurde, besteht im Plenum Raum für Verständnisfragen zum Ablauf und gegebenenfalls zur Aufgabenstellung. Anschließend wird jede Gruppe gebeten, kurz den Gesprächsverlauf im Plenum nach zu skizzieren.
- Danach wird eine zweite Runde für einen Gesprächskreis eingeleitet. Sie können individuell entscheiden, ob Sie dieselbe Gruppenbesetzung in der Raumaufteilung beibehalten, oder ob Sie die Gruppen neu mischen. Wichtig ist jedoch, dass jeder Charakter/jede Rolle nur einmal pro Raum vertreten ist. Bevor Sie die Teilnehmenden erneut in die Break Out-Räume schicken, schreiben Sie entweder auf ein Whiteboard, auf eine PowerPoint Folie, oder im Chat die zu diskutierende Aussage für die zweite Runde: *„Die Gesellschaft hat schon einen guten Umgang mit Deepfakes gefunden. Es muss nichts mehr unternommen werden.“*
- Im Breakout-Raum diskutieren die Teilnehmenden in der Gruppe aus der jeweiligen Sicht ihres Charakters/ ihrer Rolle dann die Aussagen: *„Die Gesellschaft hat schon einen guten Umgang mit Deepfakes gefunden. Es muss nichts mehr unternommen werden.“*
- Im Plenum wird nun jede Gruppe gebeten, den Gesprächsverlauf in den Break Out-Sessions zusammenzufassen.
- Daran anschließend kann eine allgemeine Diskussion über die Impulse, die aus den Break Out Sessions entstanden sind, eröffnet werden.

**Tipp:** Die Teilnehmenden können während der Diskussion in den Break Out-Räumen gebeten werden, die zentralen Aussagen und Argumente auf einem Whiteboard festzuhalten und dieses abzuspeichern, bevor sie ins Plenum zurückkehren. So kann die Vorstellung des Gesprächskreises im Plenum visualisiert werden. Die Moderation erhält zudem eine Dokumentation des Gesprächskreises in den Kleingruppen.

## 2) Präsenzbasierter Gesprächskreis über die Anwendung von Deepfakes

### **Vorbereitung:**

Vorab wird zehn der Teilnehmenden jeweils eine Rolle ausgehändigt, die sie während des interaktiven Gesprächskreises über Deepfakes einnehmen werden. Als Rollenbeschreibung dienen die Positionspapiere, die zentrale Informationen über den jeweiligen Charakter und dessen Position enthalten. Jeder Charakter repräsentiert einen Anwendungskontext von Deepfakes, den die Teilnehmenden wiederum im Gesprächskreis repräsentieren.

Zu verteilende Rollen:

- Vertreter\*in einer Frauenrechtsbewegung
- CSR-Beauftragte\*r von Facebook
- Cybersecurityexpert\*in der Bundeswehr
- Wissenschaftler\*in, die/der an synthetischer Stimmerzeugung arbeitet
- Politische\*r Künstler\*in und Dokumentarfilmer\*in
- CEO in der Unterhaltungsindustrie
- Mitarbeiter\*in einer Nachrichtenagentur
- Journalist\*in von Postillon
- Ermittler\*in im Bereich Kinderpornographie
- Museumspädagog\*in

Die Rollenbeschreibungen (in der Form von Positionspapieren) befinden sich als Druckvorlage im Anhang.

Sechs Teilnehmende erhalten das Arbeitsblatt ‚Sechs Denkhüte nach Edward De Bono‘

Zu verteilende Hüte:

1. Gelber Hut
2. Weißer Hut
3. Schwarzer Hut
4. Roter Hut
5. Grüner Hut
6. Blauer Hut

Die sechs Hüte können durch gefaltete Papierhüte in den entsprechenden Farben erstellt werden. Alternativ können auch sechs reale Hüte in den entsprechenden Farben für die Methode verwendet werden.

Die Beschreibung der einzelnen Hüte und ihrer Perspektiven befinden sich als Druckvorlage im Anhang.

### **Ablauf:**

- Die Teilnehmenden (im Idealfall 16 Personen) werden in zwei Gruppen unterteilt. Zehn Teilnehmenden wird je eine Rolle, entsprechend einem Deepfake-Anwendungskontext, zugeteilt und das entsprechende Positionspapier wird ausgehändigt. Es wird ein Halbkreis mit zehn Stühlen gebildet, auf denen die Teilnehmenden für den Gesprächskreis Platz nehmen. Optional können Namensschilder vor die Teilnehmenden aufgestellt werden, um sichtbar zu machen, welche Rolle sie vertreten.

- Sechs Teilnehmende repräsentieren jeweils einen Denkhut. Hierfür werden sie gebeten, sich in den Zuschauer\*innenraum zu setzen. Dort setzen sie ihren zugeteilten Hut auf oder platzieren ihn vor sich auf dem Boden. Zudem wird ihnen das Arbeitsblatt ‚Sechs Denkhüte nach Edward De Bono‘ ausgehändigt.
- Beide Gruppen bekommen nun zehn Minuten Vorbereitungszeit, um sich in Einzelarbeit individuell mit der zugewiesenen Rolle vertraut zu machen.
- Nun wird der Gesprächskreis angeleitet. Die (zehn) Teilnehmenden, die fortan einen Deepfake-Anwendungskontext repräsentieren, schlüpfen ausgehend von ihren Positionspapieren in die Rolle ihres Charakters und nehmen im Halbkreis Platz, um nun die Aussage zu diskutieren: *„Deepfakes sind eine große Chance für die Gesellschaft.“*
- Die Teilnehmende der sechs Hüte hören dem Gesprächskreis aufmerksam zu und machen sich währenddessen Notizen aus der entsprechenden Perspektive ihres „Denkhuts“ (weißer, schwarzer, gelber, grüner, blauer Hut).
- Nach ca. 30 Minuten wird der Gesprächskreis beendet.
- Nun wird die Gruppe der Denkhüte aktiv: Die Teilnehmenden stellen nacheinander im Plenum kurz die wichtigsten Argumente und Kernaussagen des Gesprächskreises für ihre eingenommene Perspektive (weißer, schwarzer, gelber, grüner Hut) vor.
- Am Ende ordnet der blaue Hut die Argumente. Der blaue Hut gibt einen abschließenden Überblick über die fünf Perspektiven und die dazugehörigen Argumente (weiß- Fokus auf Daten und Fakten; schwarz- Fokus auf Gefahren und Risiken; gelb-Fokus auf Chancen und Möglichkeiten; grün- Fokus auf Kreativität und innovative Ideen; rot- Fokus auf genannte Gefühle und Emotionen). Damit fasst der blaue Hut am Ende die Kernaussagen und Argumente zusammen.
- Die zehn Teilnehmenden, die zuvor im Gesprächskreis die Rolle eines bestimmten Charakters entsprechend einem Deepfake-Anwendungskontext eingenommen haben, können nun aktiv zuhören und überprüfen, ob ihrer Meinung nach ihre Argumente richtig wiedergegeben wurden, oder ob sie durch die sechs verschiedenen Perspektiven (der sechs Hüte) auf neue Erkenntnisse stoßen.
- Der ganze Ablauf kann wiederholt werden, in dem in einer zweiten Runde ein weiterer Gesprächskreis initiiert wird, mit der Aussage: *„Die Gesellschaft hat schon einen guten Umgang mit Deepfakes gefunden. Es muss nichts mehr unternommen werden.“*
- Am Ende wird die Methode abgeschlossen, in dem die Teilnehmenden wieder aus ihren zugewiesenen Rollen heraustreten. Dabei kann visuell sichtbar „der Hut abgenommen werden“ sowie das Namensschild vor sich beiseitegelegt werden.

**Tipp:** Die Gruppe kann auch geteilt werden, sodass nur jeweils fünf Charaktere auf dem Podium miteinander im Gesprächskreis interagieren. Da zwei Runden vorgesehen sind, könnte dann getauscht werden und fünf weitere Teilnehmende nehmen auf dem Podium Platz. Die anderen fünf verfolgen als Zuschauer\*innen die Dynamik des Gesprächskreises.

### Hintergrund

Die Methode geht auf den Ansatz ‚Sechs denkende Hüte‘ des britischen Kognitionswissenschaftlers Edward de Bono zurück. Diese Technik aus dem Bereich der gewaltfreien Kommunikation dient dazu, kontroverse Themen aus verschiedenen Blickwinkeln zu betrachten. Sie fördert paralleles Denken, also systematisch unterschiedliche Positionen und Denkweisen einzunehmen.

Jeder Hut entspricht einer bestimmten charakteristischen Denkweise oder einem Blickwinkel. Dadurch soll eine effiziente Diskussion erreicht werden und gleichzeitig kein Blickwinkel außer Acht




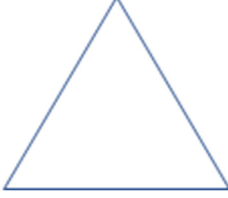
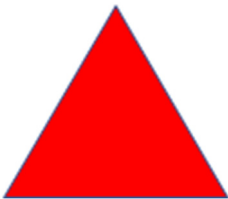
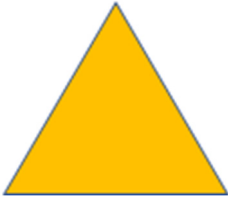
gelassen werden. Die ‚Sechs-Hüte-Methode‘ hilft Einzelpersonen, eine Vielzahl von Perspektiven einzunehmen, und Gruppen, zu ausgewogenen Urteilen zu kommen.

In den meisten Gruppenkontexten fühlen sich Einzelpersonen häufig dazu gezwungen, konsequent eine spezifische Perspektive zu verfolgen. Ziel der Methode ist es nicht, Differenzen zu überwinden, sondern sie als solche zu akzeptieren und ihnen Raum zu geben.

Die ‚Sechs-Hüte-Methode‘ trägt dazu bei, eine gemeinsame Basis für Lösungsansätze zu schaffen, indem ein Bewusstsein dafür entsteht, dass mehrere Perspektiven auf eine Thematik vorhanden sind. Beim Tragen eines bestimmten Denkhuts eröffnet die Methode die Möglichkeit zur kreativen Entscheidungsfindung.

## Anhang – Materialien als Druckvorlage

### Arbeitsblatt „Sechs Denkhüte nach Edward De Bono“

 <p><b>Blauer Hut</b></p>	<p><b>Moderation und Übersicht der Diskussion</b>          Der blaue Hut nimmt eine Sonderstellung ein. Er moderiert die Diskussion und ordnet und strukturiert die genannten Argumente. Dabei sichtet er Pro &amp; Contra, eliminiert Doppeltes und gewichtet die Argumente.</p>
 <p><b>Grüner Hut</b></p>	<p><b>Fokus auf Kreativität</b>          Im Zentrum stehen Konstruktivität und die Formulierung von kreativen Ideen.</p>
 <p><b>Schwarzer Hut</b></p>	<p><b>Fokus auf Kritikpunkte und Skepsis</b>          Nennung von Risiken, Problemen, Gefahren, negativen Auswirkungen.</p>
 <p><b>Weißer Hut</b></p>	<p><b>Fokus auf Daten und Fakten</b>          Benennung von bereits bekannten Fakten und Anforderungen. Der weiße Hut repräsentiert den „Ist-Zustand“ und nennt realistische Einschätzungen. Denkbar ist auch, wissenschaftliche Erkenntnisse, Studien oder Statistiken zu benennen.</p>
 <p><b>Roter Hut</b></p>	<p><b>Fokus auf Intuition und Gefühlen</b>          Formulierung von persönlichen Meinungen, subjektiven Wahrnehmungen oder Einschätzungen, Intuitionen und Gefühlen.</p>
 <p><b>Gelber Hut</b></p>	<p><b>Fokus auf Chancen und positive Auswirkungen</b>          Nennung von Best Cases und Idealvorstellungen, optimistische Sicht.</p>

**Ansatz**

Als Teilnehmende ziehen Sie einen spezifischen „Denkhut“ auf, indem Sie aus einem bestimmten Blickwinkel argumentieren. Diese spezifische Perspektive auf das Thema kann von Ihrer persönlichen Auffassung abweichen.

Die Methode kann Spannungen zwischen Teilnehmenden erzeugen, wenn Sie kontroverse Themen aus verschiedenen Perspektiven diskutieren. Wichtig hierbei ist, sich dessen bewusst zu sein, dass die Hüte Kategorien von Denkmustern sind und nicht der Menschen oder ihrer Einstellungen selbst. Das Tragen eines Hutes bedeutet, bewusst eine Perspektive einzunehmen, die nicht notwendigerweise die Eigene ist.

**Ablauf**

- Sie bekommen einen der sechs Hüte.
- Setzen Sie den Hut auf, oder stellen Sie den Hut vor Ihren Platz und machen Sie sich mit der Perspektive Ihres Hutes vertraut. (Bei Bedarf schauen Sie zur Vorbereitung das Kurzvideo als Zusammenfassung und Erklärung der Sechs Hüte Methode an: <https://www.youtube.com/watch?v=BmEzlfw2r68>)
- Hören Sie dem Gesprächskreis aufmerksam zu und machen Sie sich währenddessen Notizen aus Ihrer entsprechenden Perspektive.
- Präsentieren Sie anschließend im Plenum kurz die für Sie wichtigsten Argumente des Gesprächskreises.
- Sie schließen die Methode ab, indem Sie quasi ‚Ihren Hut abnehmen‘, um physisch aus der Perspektive des jeweiligen Huts zu treten.



## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### Vertreter\*in einer Frauenrechtsorganisation

- Sie sind ein/e Vertreter\*in einer Frauenrechtsorganisation und möchten auf die Genderkomponente von Deepfakes aufmerksam machen.
- 96% aller Deepfakes zirkulieren im pornographischen Bereich, was Sie als Indikator dafür sehen, wie vorherrschend Sexismus und Frauenhass in unserer Gesellschaft ist. Viele dieser Deepfakes betreffen bekannte westliche Schauspielerinnen und Sängerinnen sowie koreanische Pop-Stars. Zunehmend werden jedoch auch unbekannte Frauen Opfer von Deepfakes.
- In Bezug auf Frauenhass und Deepfakes sind Sie zunehmend besorgt über sogenannte Rache pornos, also die Erstellung synthetischen pornografischen Materials als Rache, beispielweise an der Ex-Freundin.
- Sie sind der Meinung, dass Fake Pornos ein sehr hohes Missbrauchspotenzial aufweisen, da dieser Anwendungskontext fast ausschließlich Frauen betrifft und vermehrt auch Privatpersonen.
- Deepfakes im pornografischen Bereich können inzwischen von Privatpersonen ohne große Hindernisse erstellt oder erworben werden. Personen benötigen dazu oft selbst kein technisches Wissen, sondern können auf entsprechenden Foren die Erstellung von Rache pornos als Dienstleistung erwerben. Voraussetzung hierfür ist lediglich genügend Bildmaterial, was zum Teil bereits durch die Links zu Social-Media-Profilen gegeben ist.
- Zudem gibt es auch Apps, die Deepfake Pornos erstellen. Ein Beispiel hierfür war die App „DeepNude“. Sie generierte aus Fotos einer bekleideten Frau ein Nackt-Foto (und funktionierte nur mit Bildern von Frauen). Das Programm wurde binnen kürzester Zeit 100.000fach heruntergeladen, bis schließlich die Entwickler\*innen selbst die App vom Markt nahmen (sie anschließend aber verkauften).
- Fake Pornos können als Grundlage für Kriminalität wie für Erpressungen und Verleumdungen, oder weiteres, strafrechtlich relevantes Verhalten dienen und Betroffenen auf dem Arbeitsmarkt und im Privatleben massiv schaden. Sie sehen nicht nur massiv die Persönlichkeitsrechte Betroffener verletzt, sondern sind auch besorgt, welche tiefgreifenden persönlichen Schäden durch Fake Pornos verursacht werden können.
- Sie wollen zudem auf die Gefahr hinweisen, dass Deepfake-Pornos gezielt in der Politik eingesetzt werden, da vermehrt Politikerinnen Ziel von rufschädigenden Kampagnen durch pornographische Deepfakes werden. Die Verwendung von Fake Pornos gegen politische Gegnerinnen weist eine zentrale Genderkomponente auf.
- Sie setzen sich für den Opferschutz von Porno-Fakes ein. Ihrer Meinung nach erreicht gefälschte und Rachepornografie mit Deepfakes eine neue Tragweite und der rechtliche Umgang aus der Opferperspektive ist unzureichend reguliert: Durch die Erstellung von Deepfakes kam das Opfer, dessen Gesicht oder Körper abgebildet wurde, physisch nicht in Berührung mit einem pornografischen Akt. Ein Fake Porno wird ohne dessen Bewusstsein und ohne dessen Kenntnis erstellt. Die Ersteller\*innen verwenden oft legales Bildmaterial, das sie beispielsweise auf Social Media Plattformen abgreifen. Zudem bleiben sie oft anonym. Pornografische Deepfakes strafrechtlich zu verfolgen, ist somit bis heute schwierig.

## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### CSR-Beauftragte\*r von Facebook

- Sie sind leitendes Mitglied der CSR-Abteilung von Facebook. Als solches nehmen Sie häufig an öffentlichen und medialen Debatten über die Moderation von Inhalten auf sozialen Medien und die Verantwortung großer Plattformbetreiber teil. Dabei werden Sie häufig mit der Kritik und Sorge konfrontiert, dass soziale Medien entscheidend zur Verbreitung von Fake News und Desinformation beitragen. In den letzten Jahren ist zunehmend auch von möglichem Täuschungspotenzial durch Deepfakes die Rede.
- Sie sind der Meinung, dass Facebook es Milliarden Menschen erlaubt, sich kostenfrei miteinander zu vernetzen, in Kontakt zu bleiben und sich über das Tagesgeschehen zu informieren und auszutauschen. Dabei ist die Meinungsfreiheit ein hohes Gut, das respektiert und gewahrt werden muss. Dazu gehören auch Satire und Parodien.
- Dennoch sind Sie der Meinung, dass die Verbreitung eindeutig illegaler Inhalte unterbunden werden muss. Dazu gehören Ihres Erachtens unter anderem Aufrufe zur Gewalt sowie nicht einvernehmliche pornografische Inhalte.
- Sie sind der Meinung, dass solche Inhalte auf Facebook ohnehin nicht verbreitet werden dürfen und dass der jetzige Regulierungsrahmen ausreicht. Immerhin verbot Facebook bereits 2018 in seinen Policies jegliche Form von Pornografie.
- Das Netzwerkdurchsetzungsgesetz von 2017 war zudem eine Reaktion auf vermehrte Hassbotschaften im Netz. Es verpflichtet Facebook dazu, ein Beschwerdemanagement für rechtswidrige Inhalte einzurichten und vierteljährlich darüber zu berichten. Facebook muss illegale Inhalte innerhalb von 24 Stunden löschen. Über zweifelhafte Inhalte muss Facebook innerhalb von 7 Tagen entscheiden und dies dokumentieren.
- Dies bürdet Facebook eine enorme Verantwortung auf, über die Legalität privater und politischer Meinungsäußerung zu entscheiden. Zudem muss Facebook enorme Ressourcen aufbringen, etwa für Faktenchecker\*innen.
- Die großen Bemühungen, die Facebook in der Bekämpfung von Desinformation und Deepfakes im Speziellen unternimmt, werden wenig honoriert und Facebook steht trotz allem weiterhin im Kreuzfeuer der Kritik.
- Sie wünschen sich, dass Politik und Öffentlichkeit die Bemühungen von Facebook stärker anerkennen. Dazu gehört auch die ‚Deepfake Detection Challenge‘ auf der Entwickler\*innenplattform Kaggle im Frühjahr 2020, die mehr als 2000 internationale Entwickler\*innen und Teams dazu motivierte, innovative Machine Learning-Lösungen zu entwickeln, um gefälschte Videos zu erkennen. Facebook stellte hierfür umfangreiches, echtes und gefälschtes Videomaterial als Trainingsdatensatz und insgesamt \$1 Million Preisgeld zur Verfügung.
- Die Verantwortung für die Bekämpfung von Desinformationen darf nicht auf private Unternehmen abgewälzt werden. Stattdessen bedarf es einer echten Zusammenarbeit aus Wirtschaft, Politik, Wissenschaft und Gesellschaft, um dem Problem zu begegnen.
- Sie wünschen sich eine stärkere Förderung technischer und geisteswissenschaftlicher Forschung zur Bekämpfung von Desinformation sowie eine Stärkung der Zivilgesellschaft und der Lehrkraftausbildung zur Erhöhung der Medienkompetenz der Bevölkerung.

## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### Cybersecurityexpert\*in der Bundeswehr

- Sie sind Cybersecurityexpert\*in der Bundeswehr und wollen auf die zunehmende Gefahr von hybriden Bedrohungen aufmerksam machen. Hybride Bedrohungen sind eine Kombination aus klassischen Militäreinsätzen, wirtschaftlichem Druck, Computerangriffen bis hin zu Propaganda in den Medien und sozialen Netzwerken. Ziel der Angreifer\*innen ist es, Gesellschaften zu destabilisieren. Der Cyber-Raum ist hierbei ein bevorzugter Operationsraum. Angriffe aus dem Internet sind leicht zu tarnen, da Täter\*innen häufig anonym operieren. Der Fokus von Cyber-Angriffen liegt meistens auf der Beeinflussung der öffentlichen Meinung: Von der gezielten Steuerung von Diskussionen in sozialen Netzwerken bis hin zur Manipulation von Informationen auf Nachrichtenportalen. Die Bundesregierung schätzt die Gefahr von hybriden Bedrohungen als hoch ein. Sie sehen eine direkte Gefahr von ausländischen Hackerangriffen oder gezielten Desinformationskampagnen, auch mit Hilfe von Deepfakes.
- Durch Cyberspionage könnten feindliche Akteure echte und sensible Dokumente beschaffen und diese zusammen mit Fälschungen an die Öffentlichkeit geben, um die Fälschungen zu verifizieren. Sie sehen die Außenpolitik gefährdet, indem sie durch solche Deepfakes, aber auch Video- und Audio-Deepfakes, getäuscht wird. Entscheidungsträger\*innen könnten als Reaktion auf überzeugende Deepfakes andere Entscheidungen treffen.
- Sie sehen zudem die Gefahr, dass gezielt platzierte Deepfakes zwischenstaatliche Irritationen auslösen, Konflikte anheizen oder sogar provozieren könnten, etwa indem falsche Videos von Angriffen auf bestimmte Bevölkerungsgruppen kursieren, oder Politiker\*innen abfällige Bemerkungen über andere Staaten in den Mund gelegt werden.
- Als Cybersecurityexpert\*in sehen Sie auch weitere Gefahren, die von Deepfakes ausgehen, wie zum Beispiel im Bereich von Identifizierung oder Authentifizierung. Biometrische Merkmale wie etwa das Gesicht, die Stimme oder die Iris oder lassen sich mit immer weniger Aufwand bei zeitgleich immer besserer Qualität unter Zuhilfenahme künstlicher Intelligenz fälschen. Das ermöglicht Identitätsdiebstahl, beispielsweise bei der Fälschung von Pässen.
- Sie sind besorgt um die Abwehr von Terrorangriffen. Deepfakes könnten bei der gerichtlichen Bearbeitung von IS-Rückkehrer\*innen aus dem Irak und Syrien eine Rolle spielen, wobei auch Bild- und Videomaterial als Beweismittel verwendet wird. Wenn die beschuldigten IS-Mitglieder nun glaubwürdig behaupten könnten, dass Bilder und Video- und Audioaufzeichnungen ihrer Verbrechen elektronisch manipuliert wurden, stünde die Staatsanwalt vor neuen technischen Herausforderungen. Terrorgruppen im In- und Ausland könnten darüber hinaus Deepfakes nutzen, um insbesondere über soziale Medien neue Anhänger\*innen zu werben.

## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### Wissenschaftler\*in, die/der an synthetischer Stimmerzeugung arbeitet

- Sie arbeiten als Postdoc an einer führenden technischen Universität in Deutschland. In Ihrer Promotion haben Sie ein verbessertes Machine Learning Modell entwickelt, um die Gestik und Mimik einer Person auf das Video einer anderen Person zu übertragen (Facial Reenactment). Aktuell arbeiten Sie nun an einem Projekt zur Optimierung synthetischer Stimmerzeugung.
- Sie sind überzeugt, dass Ihre Grundlagenforschung für zahlreiche legitime Zwecke eingesetzt werden kann. Insbesondere in der Assistenztechnologie im medizinischen Bereich sehen Sie großes Potenzial. Menschen, die aufgrund von Behinderungen oder chronischen Erkrankungen ihre Stimme verloren haben, könnten durch Deepfakes eine authentische Stimme zur Kommunikation erhalten, oder sogar ihre eigene Stimme gewissermaßen behalten.
- Sie sehen in Deepfakes einen großen technischen Fortschritt, der physische Beeinträchtigungen von Menschen lindern könnte. Deepfakes könnten so auch zur Inklusion von stummen Menschen in der Gesellschaft beitragen.
- Ihre Forschung könnte darüber hinaus etwa benutzt werden, um Werbung oder Bildungsvideos in zahlreiche Sprachen zu übersetzen und damit auch gezielt Menschen anzusprechen, die seltene Sprachen sprechen.
- Sie sind sich der Sorgen um die schädliche Nutzung von Deepfakes bewusst. Auf die positiven Nutzungen von Deepfakes sollte jedoch nicht verzichtet werden. Zudem: weltweit wird an der Optimierung von Deepfake-Technologien gearbeitet – an Universitäten, aber auch von Seiten der Privatwirtschaft. Sie möchten, dass Deutschland hier Spitzenreiter wird und von den wirtschaftlichen und gesellschaftlichen Vorteilen der Technologie profitieren kann. Zudem sind Sie überzeugt, dass dadurch eher sichergestellt wird, dass die Technologie nicht in die falschen Hände gerät.

## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### Politische\*r Künstler\*in und Dokumentarfilmer\*in

- Sie sind ein/e politische\*r Künstler\*in und Dokumentarfilmer\*in aus den USA. Sie nutzen Deepfakes künstlerisch, um auf gesellschaftliche Missstände hinzuweisen und für Demokratie zu kämpfen.
- Sie erstellen hierfür Deepfakes von bekannten Persönlichkeiten und unterlegen die erstellten Deepfakes mit politischen Botschaften. Beispielsweise haben Sie Donald Trump in einem Video Worte über die Wichtigkeit des Klimaschutzes in den Mund gelegt. Marc Zuckerberg macht in einem Ihrer Deepfake-Videos auf den Einfluss von Facebook auf die Gesellschaft und die daraus entstehenden Gefahren aufmerksam.
- Ihr Ziel ist es, den politischen Diskurs zu beeinflussen und zur kritischen Reflexion beizutragen.
- Ihre synthetischen Videos beginnen mit den Worten "die folgende Szene hat nicht stattgefunden", damit klar ist, dass es sich um eine Fälschung handelt.
- Sie möchten somit das Denken und Verhalten mächtiger Personen beeinflussen und Impulse für Persönlichkeiten des öffentlichen Lebens schaffen, den Botschaften ihrer Deepfakes zu folgen. Sie bezeichnen diese Form als „prosoziale synthetische Medien“.
- Gemeinsam mit befreundeten Aktivist\*innen aus der LGBT-Szene möchten Sie Ihre technischen Kenntnisse nun für ein umfangreiches Dokumentarfilmprojekt nutzen: Sie planen, einen Dokumentarfilm über die Verfolgung von Schwulen, Lesben und Transgender-Personen in Tschetschenien zu drehen und die faceswap-Technologie einzusetzen, um die Identität der Protagonist\*innen, die vor schwulenfeindlichen Säuberungen flohen, zu maskieren. Gleichzeitig können Sie dadurch deren Emotionen und Ausdrucksformen bewahren (dieses Beispiel beruht auf dem Dokumentarfilm ‚Welcome to Chechnya‘ von David France, 2020).
- Für das benötigte Bild- und Videomaterial haben Sie Freiwillige angefragt, die durch das ‚Leihen‘ ihrer Gesichter die LGBTQ-Szene unterstützten und so eine Form von Aktivismus leben.
- Der Einsatz von KI ermöglichte es Ihnen, Geschichten zu erzählen, die sonst aus Schutz der Identität der Verfolgten nicht erzählt werden könnten
- Sie sind der Meinung, dass dies der LGBTQ Szene ihre Menschlichkeit auf eine Weise zurückgibt, die unter anderen Umständen nicht möglich ist. So kann das Publikum mit ihnen mitfühlen und sympathisieren.
- Sie sind der Meinung, dass für schädliche Zwecke eingesetzte Deepfakes ein Verbrechen und keine Technologie sind. Deshalb möchten Sie Ihre Technologie ‚Deeptrue‘ nennen, weil die Protagonist\*innen in Ihrem Dokumentarfilm die Wahrheit sagen und Sie diese durch den Einsatz von Faceswap in der Wahrung ihrer Anonymität schützen.
- Sollte ihr Dokumentarfilm erfolgreich sein, erwägen Sie, mit der Opposition in Belarus zusammenzuarbeiten, um deren Botschaften im Netz zu verbreiten und gleichzeitig die Identität der Sprechenden zu verschleiern.

## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### CEO in der Unterhaltungsindustrie

- Sie sind in der Unterhaltungsbranche tätig; schwerpunktmäßig interessieren Sie sich für Animationen.
- Sie sehen in der Verwendung von Deepfakes in der Filmindustrie große Chancen, da durch Deepfakes Videomaterial erstellt werden kann, ohne dass die Schauspieler\*innen vor Ort sein müssen. Da Covid-19 die Filmbranche stark getroffen hat, sehen Sie darin eine Möglichkeit, in neuer Form auch unter Pandemiebedingungen Filme zu drehen. Perspektivisch erhoffen Sie sich auch große Zeit- und Kostenersparnisse.
- Auch im Bereich von Actionfilmen sehen Sie neue Anwendungsmöglichkeiten. Die Technologie Faceswap kann verwendet werden, um das Gesicht der Hauptdarsteller\*innen auf den Körper eines Stunt-Doubles zu legen, um realistischer wirkende Action-Aufnahmen in Filmen zu erhalten.
- Sie haben erfahren, dass durch Deepfakes die Möglichkeit besteht, tote Darsteller\*innen für neue Rollen wiederzubeleben. Dies ist insbesondere für Filme interessant, bei denen Sequels geplant werden.
- Deepfakes ermöglichen es zudem, jüngere Versionen von Schauspieler\*innen realistischer (und kostengünstiger) darzustellen, wenn in Filmen Szenen aus der Vergangenheit gezeigt werden.
- Auch im Bereich von Videospielen sind Sie von den neuen technologischen Anwendungsmöglichkeiten fasziniert. Sie haben gelesen, dass ein Fan mit geringen Ressourcen für ein Videospiel Marvels Avengers Avatare kreiert hat, in denen die Spielfiguren die Prominenten aus den entsprechenden Filmen verkörpern. Nun überlegen Sie, welche weiteren Anwendungsmöglichkeiten Deepfakes für Ihre Branche bereithält.
- Durch die Zeit- und Kostenersparnisse dank Deepfakes erhoffen Sie sich auch, dass die Erstellung von Filmen und Videospielen in der Zukunft mehr Akteuren und damit auch neuen Filmstudios oder Firmen aus dem globalen Süden offenstehen könnte.

## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### Mitarbeiter\*in einer Nachrichtenagentur

- Sie sind in einer Nachrichtenagentur tätig. Faktencheck und seriöse Quellen sind für Sie von höchster Bedeutung.
- Sie sind besorgt über die rasante Veränderung der heutigen Massenkommunikation besonders bei Jugendlichen.
- Zu großen Teilen werden in sozialen Medien Filmsequenzen zunehmend rasch konsumiert und über Plattformen und Messengerdienste wie beispielsweise Youtube, TikTok und Whatsapp verbreitet.
- Sie sind der Meinung, dass Deepfakes eine Umgebung schaffen könnten, in der Zuschauer\*innen das Gefühl bekommen, überhaupt nichts mehr glauben zu können. Sie sehen in Deepfakes eine Bedrohung für grundlegende Diskurse und Prozesse in einer offenen und demokratischen Gesellschaft.
- Sie fürchten, die Zunahme von Deepfakes könnte zu einem Vertrauensverlust führen, was in Demokratieverdrossenheit enden könnte.
- Aktuell betrifft ca. ein halbes Prozent aller Deepfakes Menschen aus Nachrichten, Medien oder Politik.
- In der Rolle als Nachrichtenagentur ist Ihnen der hohe Wert der Meinungsfreiheit sehr bewusst. Sie sehen die Schwierigkeit einer Regulierung oder eines Verbots von Deepfakes im Allgemeinen, da dies oft im Widerspruch zu Meinungsfreiheit steht. Es ist Ihnen bewusst, dass jeder Regulierungsversuch von Deepfakes insbesondere im politischen Bereich einen Eingriff in die Meinungsfreiheit darstellt.
- Sie sind besorgt, dass Deepfakes langfristig zu einer Schwächung der Medien führen könnten und sehen neue Herausforderungen für den Journalismus, da Journalist\*innen mitunter nicht mehr wissen, wie sie Quellen bewerten sollen. Diese neue Herausforderung trifft Journalist\*innen in einer ohnehin schwierigen wirtschaftlichen Lage.
- Sie sind der Überzeugung, dass Deepfakes eine Herausforderung für Glaubwürdigkeit und Vertrauenswürdigkeit medial vermittelter Informations- und Kommunikationsinhalte werden können.

## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### Journalist\*in von Postillon

- Sie greifen regelmäßig aktuelle politische Debatten auf und verpacken diese in Satire, um sie auf der Seite des Postillon zu veröffentlichen.
- Sie sind der Meinung, dass Parodie und Satire ein wichtiges Elixier für eine Gesellschaft ist. Sie erlauben, Machtstrukturen zu hinterfragen und bringen Leser\*innen zum Schmunzeln. Dafür arbeiten Sie mit gezielter Fehlinformation (aber nicht Desinformation, denn Sie zielen nicht darauf ab, die Öffentlichkeit zu täuschen).
- Deepfakes, welche Sie auf dem Postillon verbreiten, dienen der Unterhaltung und nicht dazu, den abgebildeten Personen Schaden zuzufügen.
- Sie wollen damit zu einer kritischen öffentlichen Debatte beitragen.
- Für einen Beitrag des Postillon haben Sie Deepfakes eingesetzt und fühlen sich durch die Einschätzung des Expert\*innengremium *der High Level Expert Group on Fake News and Online Disinformation* der EU bestätigt, die der Auffassung ist, dass Deepfakes im Bereich von Satire und Parodie nicht unter die Definition von Desinformation fallen.
- Das Expert\*innengremium der *High Level Expert Group on Fake News and Online Disinformation* empfiehlt solche Deepfakes als Fake zu labeln, um verifizierbare falsche Inhalte zu präsentieren.
- In letzter Zeit werden Sie zunehmend dafür kritisiert, Deepfakes für Satire und Parodie zu verwenden. Kritiker\*innen werfen Ihnen vor, dass auch im Bereich von Satire und Ironie, ethische Fragen noch nicht abschließend geklärt sind: Wer bestimmt was Satire/Ironie ist, und was nicht? Zudem: Gerade in Bezug auf Inklusion müssen Sie sich der Frage stellen, wie Sie mit Personengruppen umgehen, die keine Satire verstehen (zum Beispiel aufgrund mangelnder Sprachkenntnisse, oder aufgrund von eingeschränkten kognitiven Fähigkeiten).
- Kritiker\*innen halten Ihnen vor, dass auch im Bereich von Satire und Ironie eine Demokratiegefährdung entstehen könnte, wenn eine Deepfake-Parodie aus ihrem ursprünglichen Kontext in einen anderen Kontext übertragen wird, ohne später deren Verwendung zu kennzeichnen.



## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### Ermittler\*in im Bereich Kinderpornografie

- Sie sind in der Kriminalpolizei speziell für die Bekämpfung von Kinderpornografie tätig. Um die Beteiligten an Pornoringen ausfindig machen zu können, brauchen Sie Zugang zu den Foren. Häufig werden Personen erst in einen Pornoring eingelassen, wenn sie ein gewisses „Startkapital“ an kinderpornografischem Videomaterial übergeben. Bislang blieb Ihnen daher der Zugang zu diesen Foren verwehrt.
- Seit wenigen Monaten dürfen Sie nun synthetisches Videomaterial im Einsatz gegen Kinderpornografie verwenden. Die Bundesregierung hat Gesetze erlassen, welche es Ihnen erlauben, Kinderpornos für Ermittlungen synthetisch zu erstellen, um Zutritt zu einschlägigen Foren zu bekommen. Derzeit ermitteln Sie nun gegen mehrere Dutzend Tatverdächtige.
- Sie stellen sich die Frage, ob dadurch die Menge an kinderpornografischem Bild- und Videomaterial insgesamt erhöht wird, oder ob so gezielter die Verbreitung und Verwendung von Kinderpornografie bekämpft werden kann.
- Zudem beschäftigen Sie sich mit der Frage, welche Auswirkungen dies auf Pädophile haben wird. Pädophile könnten Deepfakes nutzen, um Kinder zu sexuellen Aktivitäten zu ermutigen.
- Langfristig erhoffen Sie sich jedoch, dass durch den Einsatz von synthetischer Kinderpornografie durch die Kriminalpolizei und die damit einhergehenden Ermittlungserfolge die Erstellung von pornografischem Material mit Minderjährigen reduziert werden kann.
- Ihnen wurde die Frage gestellt, ob es auch von Seiten der Pornohersteller\*innen legitim sei, anstelle der echten Tat solche Videos zu erstellen, um das Bedürfnis nach Kinderpornografie zu stillen.

## Positionspapier für interaktiven Gesprächskreis über Deepfakes

### Museumspädagog\*in

- Sie sind Museumspädagog\*in und in Ihrer Abteilung dafür zuständig, eine neue historische Ausstellung über die deutsche Kolonialzeit interaktiv zu gestalten.
- Sie haben davon gehört, dass das Salvador Dalí Museum in Florida seit 2019 Deepfakes nutzt, damit Besucher\*innen im Rahmen der ‚Dalí Lives‘-Ausstellung mit dem 1989 verstorbenen Künstler ‚direkt‘ in Interaktion treten kann.
- Mithilfe von Deepfakes wollen Sie nun historische Figuren erstellen, die durch direkte Ansprache zu den Besucher\*innen Geschichte auditiv und visuell repräsentieren können.
- Um die Inhalte für Kinder und Jugendliche möglichst authentisch und lebensnah zu gestalten, lassen Sie persönliche Erlebnisse von Kolonialherren und Kolonialiserten durch Deepfakes repräsentieren. Die Kinder und Jugendlichen können mit den Deepfakes interagieren, indem sie ihnen beispielsweise Fragen stellen.
- Ihre Idee haben Sie unter anderem mit mehreren Organisationen von Lehrkräften unterschiedlicher Schulen besprochen. Diese sind begeistert vom Potenzial von Deepfakes für den Bildungsbereich und könnten sich vorstellen, in Zukunft auch für ihren Unterricht auf Dienstleistungen zurückzugreifen, die Deepfakes erstellen, um diesen greifbarer zu gestalten.
- Seit Covid-19 suchen Sie neue Formen, wie ein Museumsbesuch auch virtuell umgesetzt werden kann. Sie sehen durch die Verwendung von Deepfakes, möglicherweise in Kombination mit Virtual oder Augmented Reality, die Möglichkeit, einen Museumsbesuch dadurch auch im Freien oder von zu Hause aus stattfinden zu lassen.