

**Deepfakes im Marketing:
Eine Analyse von Deepfake-Marketing mit
exemplarischer Ausarbeitung eines eigenen
Deepfake-Modells zur Überprüfung von Effekten bei
Rezipienten**

Masterarbeit
zur Erlangung des Grades Master of Science
des Fachbereichs Informatik und Medien der
Technischen Hochschule Brandenburg

vorgelegt von:
Paulina Selent



Betreuer: Prof. Dipl. Des. Julia Schnitzer
Zweitgutachter: Prof. Alexander Peterhänsel

Brandenburg an der Havel, 28. Juli 2022

In dieser Masterarbeit wird der Einsatz von Deepfakes im Marketing untersucht. Es geht darum, Deepfake-Marketing zu analysieren und vier eigene Deepfakes zu erstellen, um mögliche Effekte bei Rezipienten zu erkennen.

Dazu wird die folgende Forschungsfrage „Welche Effekte lösen hyperpersonalisierte Werbebotschaften durch Deepfakes bei Probanden aus?“ aufgestellt.

Für die Beantwortung der Forschungsfrage wird zunächst die Personalisierung von Werbung, sowie Deepfakes im Allgemeinen thematisiert. Im Anschluss daran wurden die Vor- und Nachteile der Hyperpersonalisierung durch die Integration von Deepfakes in die Werbung analysiert und vier eigene Deepfakes auf Basis realer Video-Werbungen generiert.

*Zur Überprüfung der möglichen Effekte bei Rezipienten wurde eine Online-Umfrage erarbeitet, bei der 22 Probanden die vier selbst produzierten Deepfakes sahen und Fragen zu ihrer persönlichen Einstellung zur Personalisierung durch Deepfakes mitteilten. Die Ergebnisse der nicht repräsentativen Umfrage zeigen, dass sich die Auswirkungen bei hyperpersonalisierten Werbebotschaften durch Deepfakes primär in zwei Kategorien einordnen lassen: **positive Überraschung** und **Misstrauen**.*

Im Detail können die Probanden Deepfakes nicht eindeutig erkennen und wünschen sich eine offene Deklaration von Deepfake-Inhalten in der Werbung. Bereits 4 von 10 Teilnehmern würden sogar ihre Daten für die Produktion von Deepfake-Werbung zur Verfügung stellen. Doch auch die Angst vor dem Missbrauch persönlicher Daten und möglichen Gefahren, welche von Deepfakes ausgehen könnten, belasten die Teilnehmer.

Für zukünftige Forschung bietet sich an, zur Produktvermarktung mit Deepfake-Werbung weiterzuforschen. Dafür bedarf es in erster Linie eine umfangreichere Datenerhebung.

Schlüsselwörter

Deepfake, Marketing, Hyperpersonalisierung, Künstliche Intelligenz



Inhaltsverzeichnis

Abbildungsverzeichnis	6
Tabellenverzeichnis	7
1. Einleitung	12
1.1 Motivation	12
1.2 Ziel	12
1.3 Vorgehen	13
2. Personalisierung in der Online-Werbung	14
2.1 Einordnung in den Marketingkontext	14
2.2 Begriffsabgrenzungen	15
2.3 Arten und Ziele	16
2.4 Die Rolle der Customer Journey	17
2.5 Datengrundlagen	19
2.6 Akzeptanz von Personalisierungen	21
2.7 Bisherige Einsatzmöglichkeiten von personalisierter Werbung	23
2.8 Videowerbung	24
2.9 Werbung im Wandel	25
3. Deepfakes: Fake News oder Zukunftsmusik?	28
3.1 Definitionsversuche	28
3.2 Fundamentale Prozesse der Deepfake-Technologie	29
3.3 Deepfake-Evolution	31
3.4 Arten und Funktionsweisen von Deepfakes	33
3.5 Determinanten für erfolgreiche Deepfakes	37
3.6 Aktuelle Open Source Tools	38
3.7 Reale Deepfakes und deren Auswirkungen	39
3.8 Erkennungsansätze	42
4. Konzept der Hyperpersonalisierung von Werbebotschaften mit Hilfe von Deepfakes	45
4.1 Hyperpersonalisierung durch KI-Einsatz	45
4.2 Zielgruppendefinition	46
4.3 Abwägen der Vor- und Nachteile für Kunden und Marketer	51
4.4 Szenario Entwicklung	57
5. Forschungsdesign	58
5.1 Technologie und Arbeitsweise von DeepFaceLab	58
5.2 Dokumentation des Vorgehens	60
5.3 Visualisierung der Deepfakes	70
5.4 Umfrage	71
5.5 Deskriptive Auswertung	72
6. Fazit	80
Literaturverzeichnis	82
Anhang	98

Abbildungsverzeichnis

Abbildung 1 Übersicht der Begrifflichkeiten rund um Personalisierung (In Anlehnung an Stüber, 2013 und Schoblik et al., 2020)	15
Abbildung 2 Klassischer vs. onlinegestützter Kaufprozess (In Anlehnung an Deges 2020, S. 76)	19
Abbildung 3 Prinzip der neuronalen Netze (In Anlehnung an Buxmann & Schmidt 2019, S. 14)	30
Abbildung 4 Beispielhafte Darstellung eines Deep Neural Network (In Anlehnung an rsipvision, 2015)	31
Abbildung 5 oben: DeepMind KI (Quelle: Brock 2019, S.16); unten: Samsung KI (Quelle: Zakharov 2019, S.1)	32
Abbildung 6 Beispiel für FaceShifter (Quelle: Li et al., 2020, S.1)	32
Abbildung 7 Vergleich der Face Swapping Methoden aus Naruniec et al. (2020)	32
Abbildung 8 Methode des Face Swapping im Überblick (Quelle: Nirkin et al., 2017, S. 3)	33
Abbildung 9 Auto-Encoder-Decoder Prinzip: Nutzung desselben Encoder-Decoder-Paares für die Extrahierung von Gesichtsmerkmalen mit Tausch der Decoder für die Generierung des Deepfakes (In Anlehnung an Masood et al., 2021, S. 8)	34
Abbildung 10 Arbeitsweise der FACIAL-GAN (Quelle: C. Zhang et al., 2021, S. 3)	35
Abbildung 11 Gesichtssynthese mit StyleGan (Quelle: Karras et al., 2019, S. 4)	36
Abbildung 12 Facial Attribute Manipulation mit StarGan (Quelle: Choi et al., 2018, S. 1)	37
Abbildung 13 links: (Suwajanakorn et al., 2017); rechts: (BuzzFeedVideo, 2018)	40
Abbildung 14 links: Deepfake-Avatar (The Dali Museum (a), 2019); rechts: Deepfake-Satire (Sassy Justice, 2020)	41
Abbildung 15 Einteilung der Deepfake-Erkennungsmechanismen (In Anlehnung an Masood et al., 2021, S. 31f und Mirsky und Lee, 2020, S. 26ff)	42
Abbildung 16 DIVSI-Internetmilieus mit Grad der Eignung für Hyperpersonalisierung mit Deepfakes (In Anlehnung an DIVSI, 2016)	47
Abbildung 17 Anteile der Nutzertypen aus Deutschland sicher im Netz (2021, S. 17)	48
Abbildung 18 Photo by Andrea Piacquadio from Pexels	49
Abbildung 19 Photo by Ron Lach from Pexels	50
Abbildung 20 Hypothesenmodell	56
Abbildung 21 Übersicht der verfügbaren Skripte in DeepFaceLab	60
Abbildung 22 Bereinigung fehlerhafter Gesichter für src und dst	61
Abbildung 23 links: maskiertes Gesicht im XSeg-Editor, rechts: XSeg-Training für den Deepfake „Pia“	62
Abbildung 24 SAEHD Trainingsübersichten für die Deepfakes mit Eye-Priority	66
Abbildung 25 Benutzeroberfläche für das Merging in DeepFaceLab	67
Abbildung 26 Postproduktion in Adobe AfterEffects	68
Abbildung 27 Postproduktion Vorher-Nachher-Vergleich (links: Rohdaten aus DFL, rechts: nach Postproduktion)	69
Abbildung 28 Impressionen der Deepfakes (links = Original, rechts = Deepfake)	70
Abbildung 29 Antwortverteilung zur Frage „Waren die Videos Deepfakes?“	76

Tabellenverzeichnis

Tabelle 1 Zahl der Publikationen zum Stichwort „deepfake“ von 2015 bis 2022 generiert mit: https://app.dimensions.ai _____	40
Tabelle 2 Vergleich der empfohlenen und tatsächlichen Hardware für DeepFaceLab ____	58
Tabelle 3 Dauer und Frameanzahl der src-Videos _____	59
Tabelle 4 Übersicht der XSeg-Werte _____	63
Tabelle 5 Übersicht der SAEHD-Parameter _____	65

Hinweis im Sinne des Gleichbehandlungsgesetzes

Zur besseren Lesbarkeit wird in der vorliegenden Arbeit auf die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Es wird das generische Maskulinum verwendet, wobei alle Geschlechter gleichermaßen gemeint sind.

Ehrenwörtliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen oder Hilfsmittel benutzt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt wurde.

Brandenburg an der Havel, 28. Juli 2022
Ort, Datum



Unterschrift
(*Paulina Selent*)

I. Einleitung

I.1 Motivation

Kunden reagieren auf und akzeptieren heutzutage kaum unpersonalisierte Werbung, so Helmut Briggel von der Dialogagentur Defacto (Deutsche Post, 2019). Deshalb liegt Personalisierung auch in den Top 3 der Prioritäten für Medienunternehmen im Jahr 2020 (Adobe, 2020). Knapp 40% der Unternehmen gestalten ihre Onsite-Personalisierung bereits mit Hilfe von künstlicher Intelligenz (KI). Haupteinsatz der KI liegt hier bisher in der Datenanalyse. KI-gestützte Design-Projekte finden in den Unternehmen bislang wenig Zuspruch (Adobe & Econsultancy, 2020). Dies weist eine Diskrepanz zu den prognostizierten Umsätzen auf, welche sich in den kommenden 4 Jahren laut Tractica mit KI erzielen lassen können (Tractica, 2016). Hierbei soll sich der Umsatz fast versechsfachen lassen und bietet somit großes Potential für neuartiges Nischenmarketing, sowie einzigartige Wettbewerbsvorteile.

Deepfakes sind Bilder oder Videos, die mit Hilfe künstlicher Intelligenz erstellt wurden und täuschend echt wirken (Bendel, 2019). Mika Westerlund und weitere Autoren stellten fest, dass Deepfakes eine Bedrohung für Wirtschaft, Politik und Gesellschaft darstellen und mit Hilfe von Gesetzen, neuester Technik und Weiterbildung authentifiziert oder bekämpft werden können (Antova et al., 2021; Westerlund, 2019). Die ersten Agenturen nutzen bereits Deepfakes für die Erstellung ihres Videocontents. So wirbt der Energieanbieter Tibber in einem aktuellen Werbespot mit einem Angela

Merkel Deepfake für günstigen Strom und setzt auf Witz und Komik (Tibber Norge, 2021). In den Trailer zum Thriller Reminiscence von Warner Bros. kann man das eigene Gesicht einfügen (Haridy, 2021) und Deutschlands erfolgreichste Daily-Soap „Gute Zeiten, schlechte Zeiten“ thematisiert den Umgang mit Deepfakes in der Jubiläumsfolge (Kohlstedt, 2022). Ganz aktuell kursiert ein Fake-Videoanruf des Bürgermeister von Kiew – Vitali Klitschko mit den Bürgermeistern von Berlin, Madrid und Wien bezüglich des Ukraine-Krieges. Nach zweifelhaften Aussagen und Misstrauen bestätigte sich: „*Allem Anschein nach war es ein Deep Fake*“ (spiegel.de, 2022)! Deepfakes werden 2022 zu einem wachsenden Trend (Davies, 2021)! Whittaker et al. (2021) gehen sogar davon aus, dass Werbebotschaften mit Deepfake-Technologie in Zukunft effektiver als rein KI-basierte Werbebotschaften sein werden.

I.2 Ziel

Ziel dieser Masterthesis ist es, den aktuellen Stand der Forschung zum Thema Deepfakes mit dem der Personalisierung zu verknüpfen. Die Forschung zu Deepfakes beläuft sich momentan in hohem Ausmaß zur Erkennung und Prävention dieser neuen Technologie. Diese Masterarbeit soll die Erkenntnisse zur positiven Nutzung von Deepfakes im Marketing erweitern. In der Literatur werden dazu bereits theoretische Ansätze und Anregungen für zukünftige Forschung im Bereich Deepfakes und Marketing gegeben (Campbell et al., 2021; Whittaker et al., 2021). Diese Arbeit greift Aspekte daraus auf und erweitert sie durch das Prinzip der Hyperpersonalisierung. Für

die Untersuchung des Status quo wird eine Literaturrecherche durchgeführt, um aktuelle Forschungsansätze und deren Erkenntnisse zu untersuchen und miteinander in Verbindung zu setzen. Da sich die Thematik der Deepfakes erst innerhalb der letzten 5 Jahren entwickelte, besteht die Literatur dazu größtenteils aus Preprints. Umso wichtiger ist es, die Forschung in diesem Bereich voranzutreiben. Dafür wurde die zentrale Forschungsfrage „Welche Effekte lösen hyperpersonalisierte Werbebotschaften durch Deepfakes bei Probanden aus?“ aufgestellt. Im Laufe der Arbeit wird diese Forschungsfrage durch vier Hypothesen ergänzt und mit Hilfe einer Online-Umfrage beantwortet.

1.3 Vorgehen

Für die theoretische Grundlage dieser Arbeit werden in den ersten beiden Oberkapiteln zum einen die Thematik der Personalisierung im Onlinemarketing (Kapitel 2) mit ihren Spezifikationen und Einsatzgebieten erläutert. Zum anderen wird die Technologie der Deepfakes (Kapitel 3) beschrieben. Hier geht es darum, das Fundament dieser beiden Thematiken zu schaffen, welche für das Deepfake-Marketing von essentieller Bedeutung sind. Das Konzept der Hyperpersonalisierung von Werbebotschaften mit Hilfe von Deepfakes wird im vierten Kapitel dieser Arbeit diskutiert und die Effekte von Deepfakes bei Probanden mit einer Online-Umfrage erforscht.

Im Rahmen der Masterarbeit wird dabei die Werbeform des Deepfake-Marketings für die Hyperpersonalisierung von Werbung untersucht. Dafür werden anhand eines selbst entwickelten „Buddy-Systems“, die Auswirkungen von Deepfake-Werbung an Probanden erforscht. Die vier Person of Interest (POI) stellen dabei Videomaterial für die Erstellung eines Face-Swap Deepfakes mit dem Gesicht der eigenen Person zur Verfügung. Fünf Probanden aus dem Umfeld der jeweiligen POI, die sogenannten Buddys sehen diesen und zwei weitere high-end Deepfakes und müssen anhand von Fragen und Aussagen ihre Einstellungen und Gedanken mitteilen. Auf Grundlage dieser Ergebnisse sollen durch statistische Verfahren erste Tendenzen zur Eignung und Akzeptanz des Deepfake-Marketings erkannt werden.

Der Schwerpunkt dieser Arbeit liegt in der Untersuchung einer neuen Marketingform mit Deepfake-Inhalten durch ein Szenario und der Überprüfung von Effekten, die diese Marketingrichtung bei Probanden verursacht. Darüber hinaus sollen erste Tendenzen zur Akzeptanz bzw. des Erfolgs dieser Marketingrichtung, sowie Zukunftsausblicke gegeben werden.

2. Personalisierung in der Online-Werbung

„Um erfolgreich zu sein, müssen Marken ihre Kundendaten besser nutzen. Die Bereitstellung relevanter Kommunikation, die eine 360°-Sicht auf Kundinnen und Kunden vermittelt, ist hierbei von essenzieller Bedeutung.“, so Andrew Stephenson – Director of Marketing EMEA bei Treasure Data (von Piechowski, 2021).

Personalisierung von Werbemaßnahmen zählt zu einer der Grundproblematiken von Werbeschaffenden im Onlinemarketing. In den folgenden Kapiteln wird die Thematik der Personalisierung in den Kontext des Onlinemarketing eingeordnet und in Hinblick auf eine nützliche Erweiterung durch künstlich intelligente Systeme diskutiert.

2.1 Einordnung in den Marketingkontext

Das Onlinemarketing umfasst die Planung, Organisation und Kontrolle aller marktorientierten Aktivitäten, die über mobile oder stationäre Endgeräte mit Internetzugang zur Erfüllung der Marketingziele, den Rezipienten zugänglich gemacht werden (Kreutzer, 2021, S. 6). Der Geschäftsführer des Deutschen Marketing Instituts Prof. Dr. Michael Bernecker fügt in seiner Definition von Onlinemarketing noch hinzu, damit einen „[...] Mehrwert für Kunden, Online-Nutzer und das Unternehmen zu erreichen [...]“ (Bernecker, 2021).

Mit dem Erfolg der sozialen Medien und dem großen Angebot der Interaktionsmöglichkeiten im Internet (wie z.B. Shopping, Entertainment, Informationsgewinnung und sozialer Austausch) stieg auch die Relevanz des Onlinemarketings.

Nutzer konsumieren, reagieren auf und erschaffen Inhalte im Internet und sind dabei stets auf der Suche nach einem „guten Gefühl“, wie Kreutzer es beschreibt. Als Orientierungspunkt für effizientes und erfolgreiches Onlinemarketing dient das Oktogon des Marketings nach Kreutzer (Kreutzer, 2021, S. 14):

Zu den acht Erfolgsfaktoren des Onlinemarketing gehören:

1. Aufbau guter Gefühle,
2. Erwartungsmanagement,
3. Wertorientiertes Kundenmanagement,
4. Konsequente Ergebnisorientierung,
5. Empfängerorientierte Kommunikation,
6. Umsetzung der 4 Ks (Kontinuität, Konsistenz, Konsequenz, Kompetenz),
7. Bedienung aller Customer Touchpoints und die
8. Relevanz der Angebote.

Insbesondere auf den zuletzt genannten Punkt des Oktogons des Marketings wird in dieser Arbeit wert gelegt. Es geht darum, für den Nutzer relevante Produkte (und/ oder Dienstleistungen) anzubieten. Dies wird unter anderem mit personalisierten Angeboten erreicht.

2.2 Begriffsabgrenzungen

Da in der Literatur keine allgemeingültige Definition zu Personalisierung existiert und viele Begriffe synonym zu „Personalisierung“ genutzt werden, muss eine Abgrenzung erfolgen.

Der Duden versteht unter dem Begriff Individualisieren, etwas auf den Einzelfall beziehen (Duden (a), 2022). Personalisieren jedoch meint etwas auf eine einzelne Person auszurichten (Duden (b), 2022). Die Begriffe werden in wissenschaftlichen Publikationen häufig synonym genannt und haben zum Ziel, die individuellen Wünsche und Interessen der Kunden zu verstehen und wertzuschätzen. Schneider (2018) unterscheidet Individualisierung als Ermöglichen eines selbst zusammengestellten, einzigartigen Produkts nach individuellem Geschmack und gibt als Beispiel den Konfigurator für Nike Sportschuhe an.

Personalisierung hingegen ist eine direkte und maßgeschneiderte Kundenansprache, die sich den individuellen Bedürfnissen und Anforderungen eines Kunden anpasst (Schneider, 2018).

Eva Stüber zählt in „Personalisierung im Internethandel“ auch Begriffe wie Segmentierung, One-to-One-Marketing oder Customization zu Synonymen von Personalisierung (Stüber, 2013, S. 10ff). Auf weitere Abgrenzungen wird an dieser Stelle jedoch verzichtet und lediglich eine Begriffsübersicht dargestellt:

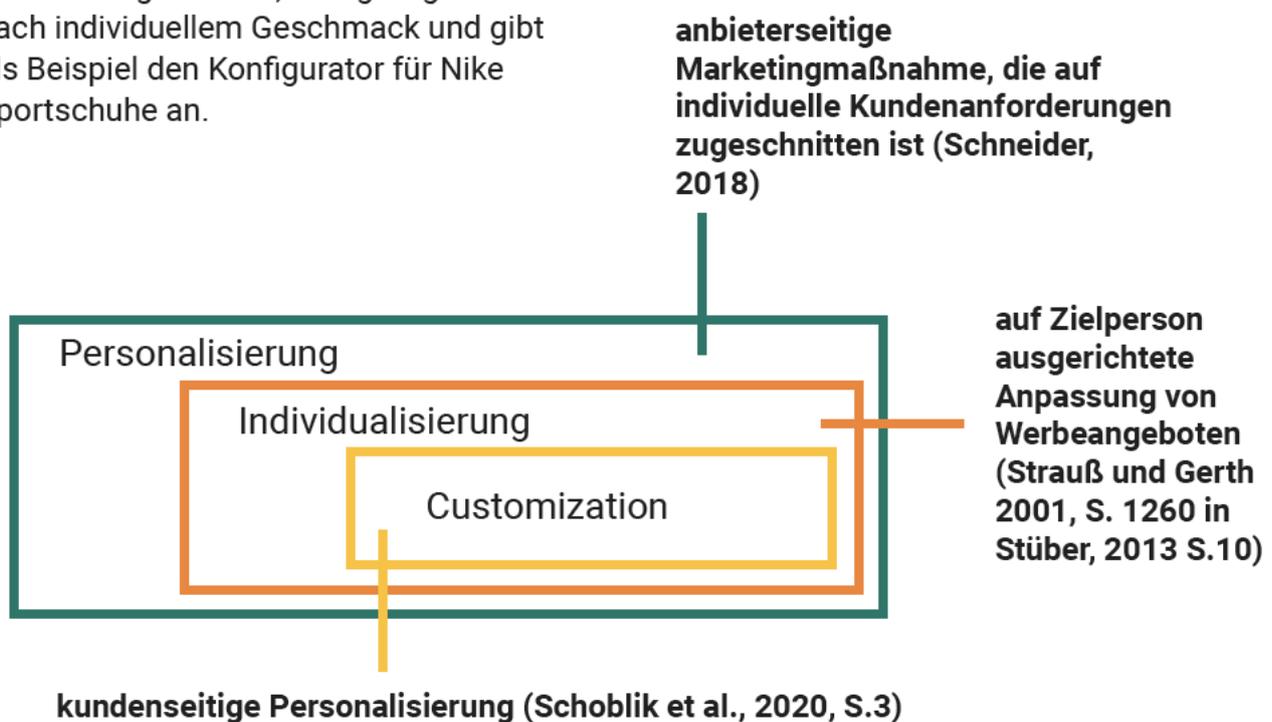


Abbildung 1 Übersicht der Begrifflichkeiten rund um Personalisierung (In Anlehnung an Stüber, 2013 und Schoblik et al., 2020)

2.3 Arten und Ziele

Die Forschungsergebnisse von Lee & Park (2009) aus dem Jahr 2009 stellen dabei drei wesentliche Personalisierungsarten heraus. Demnach ergibt sich nach Wichtigkeit der Probanden der Studie geordnet folgende Reihenfolge (vgl. auch Stüber, 2013, S. 23):

- **Persönliche Beratung „Personal Advice Personalization“**

Die wichtigste Funktionsausprägung zur Personalisierung von Online-Diensten stellt die persönliche Beratung dar. Hierzu zählen z.B. personalisierte Produkt- oder Preisvergleiche und auch Entscheidungshilfen.

- **Erkennung „Recognition Personalization“**

Bei dieser Art der Personalisierung werden Daten gespeichert und bei Bedarf beim nächsten Kontakt des Nutzers mit dem Unternehmen genutzt (z.B. für die persönliche Anrede).

- **Angebot „Offer Personalization“**

Als weniger relevante Art der Personalisierung zeigte sich bei den Konsumenten die Angebotspersonalisierung. Ein Beispiel wären z.B. personalisierte Preise.

Blickt man über den Tellerrand in eine andere Disziplin des Marketings – dem Content Marketing so verspricht laut Volz & Griep (2020, S. 144 –157) vor allem die Kombination zweier Personalisierungsarten Erfolg:

1. Personalisierung durch Personas:

Hierfür muss eine solide Kundendatengrundlage geschaffen werden, um den aktuellen Kundenstamm und somit die Zielgruppe der Maßnahmen identifizieren zu können. Nützliche Daten sind hierbei neben unternehmensfremden (Second-Party-Data/Third-Party-Data) vor allem die unternehmensinternen Daten (First-Party-Data). Beispiele hierfür sind z.B. Käufe, Kaufabbrüche, Frequenz der Webseitenbesuche, Preissegment. Aus diesen persönlichen Informationen in einer Datenbank können mit Hilfe der Clusteranalyse Personengruppen identifiziert werden.

2. Personalisierung anhand der Marketingfunnelstufen:

Im Sinne des Marketingtrichters (engl. Sales Funnel) ist der Prozess, den ein Kunde durchläuft in verschiedene Stufen eingeteilt: **Awareness** (Sehen, was der Markt bietet), **Interest** (Produkte fokussieren), **Consideration** (favorisiertes Produkt auswählen), **Purchase Intent** (Auswahl ist getroffen, letzter Kaufimpuls fehlt noch), **Purchase** (Kauf des Produkts), **Upselling** (Nutzen von Zusatzoptionen zum Produkt abwägen) und **Repeated Purchase** (Wiederholung des Kaufs). Insofern die Nutzerbedürfnisse je Funnelstufe ermittelt wurden, wird es möglich Marketingkanäle und Content an die Bedürfnisse des Kunden anzupassen.

Volz & Griep (2020) empfehlen die Kombination dieser beiden Ansätze, um „[...] sowohl den Nutzeransprüchen, die aus der Persona resultieren, gerecht zu werden als auch die verschiedenen Bedürfnisse entlang des Sales Funnels zu erfüllen.“ (S.

154f) und somit die Unternehmensziele – insbesondere Marketingziele – zu erreichen. Die Zielgruppe des Deepfake-Marketings wird in Kapitel 4.2 Zielgruppendefinition weitergehend erörtert.

„[...] Wir erwarten auf uns zugeschnittene Inhalte und Botschaften.“, so betiteln Kröger & Marx (2020, S. 9) die Folge unseres Wohlstandes und den damit verbundenen Möglichkeiten die den Nutzern heutzutage zur Verfügung stehen. Murthi & Sarkar (2003) stellen in ihrer Forschung zwei Hauptgründe für die große Bedeutung von Personalisierung im Marketing heraus:

1. Für Konsumenten wird die Komplexität des Kaufentscheidungsprozesses durch eine gezieltere Auswahl verringert.
2. Für Unternehmen entstehen Wettbewerbsvorteile durch die Differenzierung zu (wohlmöglich nicht personalisierten) Konkurrenzangeboten.

Auch für weitere Forscher (Coelho & Henseler, 2012; Fan & Poole, 2006) steht deshalb das obergeordnete Ziel von personalisierten Angeboten fest: Unternehmen setzen sich durch Personalisierungen von relevanten Mitbewerbern ab. Doch auch für die Kunden erkennen Forscher positive Einflüsse z.B. in den Bereichen Kundenzufriedenheit, Kundenbindung und Kundenvertrauen (Joecks-Laß, 2020, S. 79).

Diese Kundenreise – auch Customer Journey genannt – spielt eine zentrale Rolle im Onlinemarketing. Das folgende Kapitel soll das Prinzip der Customer Journey und den Zusammenhang zur Personalisierung von Werbeangeboten darstellen.

2.4 Die Rolle der Customer Journey

Für eine erfolgreiche Personalisierung von Inhalten wird es also nötig, die Reise eines potentiellen Kunden, vom ersten Kontakt mit dem Produkt/der Dienstleistung bis hin zur letztendlichen Transaktion und einem möglichen Wiederkauf zu kennen und zu verstehen.

Für Marketer haben – insbesondere nach zwei Jahren COVID-19 Pandemie – Kundenbeziehungen einen großen Stellenwert im täglichen Geschäft eingenommen, um schnell auf sich ändernde Kundenbedürfnisse eingehen zu können (Carrel, 2022, S. 3). *„Kunden erwarten von Unternehmen, dass sie nahtlos von einem Interaktionsmodus in den nächsten wechseln können, dass sie sich ihre Präferenzen über alle Touchpoints hinweg merken und dass ihnen in jeder Phase ihrer Customer Journey herausragende Erlebnisse geboten werden.“* (Carrel, 2022, S. 8), so der Vizepräsident Marketing bei Adobe weiter.

Kaufinteressierte kommen mit den Angeboten der Unternehmen sowohl online als auch offline in Berührung. Diese sogenannten Berührungspunkte – Touchpoints – nähren die Customer Journey. Dabei gilt es besondere „Moments of Truth“, also spezielle Zeitpunkte für ein Kaufverhalten zu unterscheiden. Der First Moment of Truth beschreibt dabei das Aussuchen und Erwerben eines Produkts. Der Second Moment of Truth liegt vor, wenn das Produkt ausgepackt und genutzt wird (Kruse Brandão & Wolfram, 2018, S. 92).

Allerdings sollte hierbei berücksichtigt werden, damit eine strenge Abgrenzung zwischen online und offline Touchpoints oft ungenau definiert werden kann.

Digitale Endgeräte sind täglicher Begleiter von 80% aller Deutschen (Initiative D21 e.V., 2021, S. 11) und ermöglichen einen permanenten Wechsel zwischen online und offline. Kreutzer schlägt deshalb vor, die Customer Journey „noline“ zu gestalten. Die Differenzierung von online und offline verliert an Bedeutung und bedarf einer Kombination im Marketingkonzept (Kreutzer, 2021, S. 11). Aufgrund des Online-Zeitalters fügt Kreutzer (a) (2021, S. 36f) deshalb noch den Zero Moment of Truth hinzu. Dieser umfasst alle Informationen, die Dritte online zur Verfügung stellen (user-generated-content). Die eigene Customer Journey wird somit noch vor dem eigenen Auseinandersetzen mit dem Produkt von Erfahrungen anderer Kunden beeinflusst. Deshalb ist es wichtig auch diese Komponente auf der Kundenreise zu berücksichtigen.

Die Customer Journey wird in der Literatur vermehrt als Trichterprozess dargestellt: Interessenten begeben sich in den Trichter und jene, die durch den Trichter wieder heraus kommen sind zu realen Kunden geworden. Aufgabe des Unternehmens ist es, die Reise durch den Trichter so angenehm und wirkungsvoll wie möglich zu gestalten. Einen Ansatz dafür bietet das noch heute Anwendung findende Werbewirkungsmodell „AIDA-Modell“ von Elmo Lewis aus dem Jahr 1989 (Strong, 1925, S. 76).

AIDA steht demnach für:

- **Attention (Aufmerksamkeit):** Aufmerksamkeit potentieller Kunden gewinnen,
- **Interest (Interesse):** Interessen der Kunden am Produkt fördern und ausbauen,
- **Desire (Verlangen):** Begierde nach dem Produkt beim Kunden auslösen und
- **Action (Handlung):** Kunden zum Kauf verleiten.

Kreutzer jedoch erweitert dieses Modell zu **ASIDAS**: Mit seinem dynamischeren Charakter zielt das Modell darauf ab, damit die einzelnen Abschnitte der Customer Journey nicht mehr strikt in Reihe ablaufen und sich gegenseitig beeinflussen können. **ASIDAS** bedeutet: **Attention – Search - Interest - Desire - Action - Share**. Zusätzlich zu den herkömmlichen AIDA-Prozessen können z.B. intensive Suchphasen (**Search**) auftreten oder Erfahrungen mit Anderen geteilt werden (**Share**) (Kreutzer (a), 2021, S. 36f). Deges (2020, S. 76) hat für den onlinegestützten Kaufprozess fünf Phasen identifiziert.

Die folgende Abbildung soll die Unterschiede zum offline Kaufentscheidungsprozess verdeutlichen:

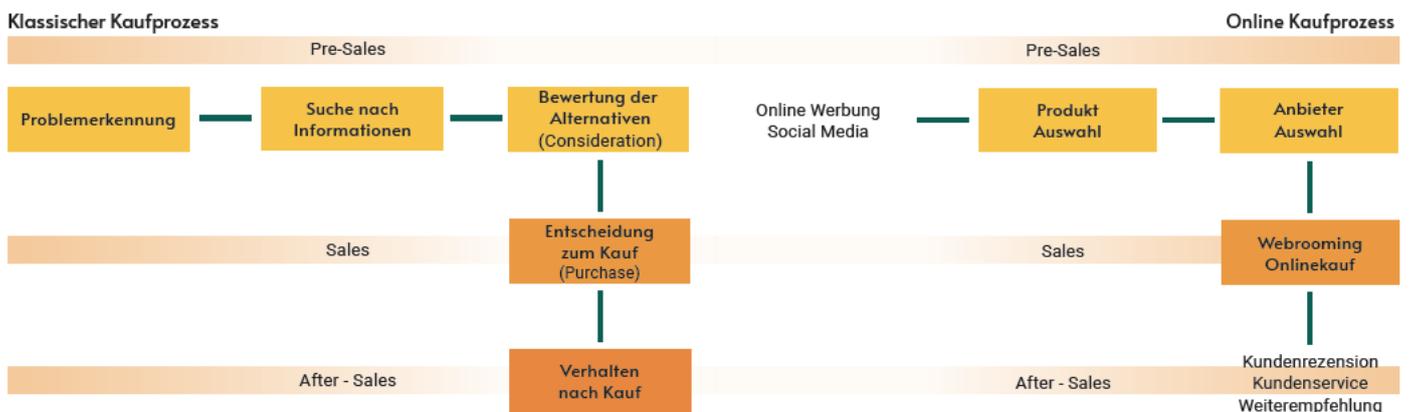


Abbildung 2 Klassischer vs. onlinegestützter Kaufprozess (In Anlehnung an Deges 2020, S. 76)

Nutzer kommen sowohl online als auch offline durchschnittlich mit 6 bis 8 Touchpoints in Berührung, bevor sie sich für ein Produkt entscheiden (Glynn, 2015). Aus diesem Grund ist das Zusammenspiel aller Touchpoints ausschlaggebend für eine zielführende Customer Experience (Meffert et al., 2019, S. 146).

Wie jedoch generiert man ausreichend wertvolle Informationen über die potentiellen Kunden und welche Daten werden dafür benötigt? Ohne solide Datengrundlage können die Präferenzen der Rezipienten des Deepfake-Marketings nur errahnt werden. Einen Überblick auf mögliche Datenquellen wird das nächste Kapitel geben.

2.5 Datengrundlagen

Sobald Nutzer in irgendeiner Form – durch Informationssuche, Teilen von persönlichen Inhalten auf Social Media Plattformen, Scannen von QR-Codes oder Interaktion mit Sprachassistenten – mit dem Internet in Berührung kommt, werden absichtlich und unabsichtlich Daten hinterlassen.

Die International Data Corporation (IDC) prognostiziert für das Jahr 2025 eine Generierung neuer Daten in Höhe von 175 Zettabyte (1 Zettabyte entspricht 1.000.000.000.000.000.000.000 Byte) (Reinsel et al., 2018, S. 3).

Für differenzierte Personalisierung werden verschiedene Daten benötigt. Jedoch kommt es neben der Quantität viel mehr auf die Qualität dieser an. Bereits vor über 20 Jahren zeichnete sich ab, dass man mit vielen gesammelten Nutzerdaten nicht automatisch bessere Ergebnisse erzielt. Davenport et al. (2001) erkannten schon damals: „Companies [...] may know more about their customers, but they don't know the customers themselves.“

Zuerst wird zwischen individuellen Kundendaten und aggregierten Marktdaten unterschieden. Kundendaten werden aus internen Prozessen oder Befragungen generiert und umfassen beispielsweise Transaktionsdaten oder Kontaktdaten. Marktdaten wiederum umfassen Informationen über z.B. Konkurrenten oder Kundensegmente, die über klassischen Marktforschungsmethoden

erfasst werden können (Hippner & Wilde, 2006, S. 404ff). Keuper et al. (2018, S. 40) untergliedern Daten noch einmal hinsichtlich ihrer technischen Struktur: Mit dem Aufkommen von moderner Informationsverarbeitungssoftware wurde es möglich neben strukturierten – also von Computern eindeutig lesbaren Informationen – auch unstrukturierte Daten wie z.B. handgeschriebene oder gesprochene Informationen computergestützt auszuwerten und vor allem zu nutzen.

Um diese Daten noch besser zu spezifizieren, lassen sich Daten in First-, Second-, und Third-Party-Daten einteilen: First-Party-Daten umfasst die Informationen, die ein Unternehmen selbst über seine Kunden durch seine eigenen Tools wie z.B. Website oder App erfasst. Sie gelten als besonders wertvoll, da diese Daten direkt vom und im Endeffekt auch wieder für den Kunden gewonnen werden (Kamps & Schetter, 2018, S. 90). Als Beispiele sind hier folgende Arten zu nennen (vgl. auch Keuper et al., 2018, S. 40):

- Kontaktdaten (z.B. Name, Anschrift, Land, E-Mail)
- Zielgruppeninformation (z.B. Geburtstag, Familienstand, Kinder, Interessen)
- Transaktionsdaten (z.B. Produkte, Kaufhäufigkeit, Menge, Umsatz)
- Registrierungsdaten (z.B. Permissions)
- Responsedaten (z.B. Öffnungsraten)
- Social-Media-Verhalten (z.B. Likes, Suchverlauf, Follower)
- User-Verhalten (z.B. Verweildauer, Klickverhalten)

Als Second-Party-Daten beschreiben Kamps & Schetter (2018) weiterhin exklusive Kundendaten von anderen Unternehmen – sozusagen externe First-Party-Daten. Diese können z.B. durch direkte Kooperationen geteilt oder über Data-Management-Plattformen ausgetauscht werden. Der noch fehlende Datentyp Third-Party-Daten umfasst anonymisierte Datensätze, die von Datenhändlern gesammelt und zum Kauf (oder Miete) angeboten werden. Aufgrund der Erweiterung von Tiefe und Umfang der Userdaten eignen sich diese beiden Datentypen besonders für Optimierungen der Customer Journey (Keuper et al., 2018, S. 42).

Marketer zählen laut Studie von Lebo (2019) vor allem First-Party-Daten wie Aktivitäten auf der Website, bei Transaktionen und innerhalb von Kampagnen zu den wichtigsten Datenarten für die Personalisierung ihrer Produkte. Der Spezialist für Personalisierung bei Dynamic Yield, Anoop Vasisht, sieht vor allem in den folgenden Datenquellen sinnvolle Ansatzpunkte: Location, Endgerät, Traffic Source, Drittanbieterdaten, Besucherverhalten, Customer-Relationship-Management (CRM), Uhrzeit und aktuelle Landingpage (Vasisht, 2020).

Zur Erhebung der Daten können verschiedene Herangehensweisen genutzt werden. Klassischerweise können Daten aus dem CRM genutzt oder anhand von Interaktionen mit dem Unternehmen ermittelt werden. Zu weiteren Ansätzen gehört aber auch der visuelle Ansatz durch Elemente aus dem Bereich der Gamification wie z.B. Bilder zur Auswahl von Lebensmitteln zur Ermittlung

von potentiellen Lieblingsgerichten. Bei der Gamification werden spieletypische Elemente in spielfremde Anwendungsgebiete z.B. mit dem Ziel der Motivationssteigerung übertragen (Bendel, 2021).

Nachdem eine erste Hürde – die Erhebung von qualitativ hochwertigen Kundendaten – erfolgreich überwunden wurde, stellt sich im Anschluss die Frage, wie die Akzeptanz von personalisierten Angeboten bei den Zielpersonen gebildet bzw. beeinflusst wird.

2.6 Akzeptanz von Personalisierungen

Die Akzeptanz wird in der Akzeptanzforschung durch diverse Modelle und Definitionen beschrieben. Im Allgemeinen kann die Akzeptanz als Bereitschaft, einen Sachverhalt, ein Objekt oder eine Person in Kauf zu nehmen betrachtet werden (Markgraf, 2018; Pressmar, 1982, S. 324). Das Akzeptanzmodell von Ajzen & Fishbein (1980) auch bekannt als Theory Of Reasoned Action (TRA) geht davon aus, dass die Akzeptanz mit der subjektiven Einstellung zu einem Sachverhalt einhergeht. Ein Verhalten wird dabei komplett durch eine Verhaltensabsicht bestimmt, welche durch die eigene Einstellung und vermutete Erwartungen beeinflusst werden (Graf, 2007, S. 34f).

Kollmann's Forschung zum Thema Akzeptanz (1998, S. 67ff) bildet im deutschsprachigen Raum das empirische Fundament und umfasst insgesamt drei Phasen des Akzeptanzprozesses, die sich

gegenseitig beeinflussen:

- **Einstellungsphase:** Der Konsument lernt die Innovation kennen, wertet Informationen aus, setzt sich mit Einsatzmöglichkeiten auseinander, sucht aktiv nach Zusatzinformationen und bewertet die Innovation subjektiv anhand von Vor- und Nachteilen.
- **Handlungsphase:** Der Konsument wandelt Interesse in eine konkrete Handlung wie z.B. Besitz oder Kauf um.
- **Nutzungsphase:** Der Konsument nutzt die Innovation freiwillig, wiederholt und mit dem Ziel bestimmte Aufgaben oder Probleme zu lösen.

(vgl. auch Stüber, 2013, S. 44ff; Schlohmann, 2012, S. 128ff)

Auf die Phasen des Akzeptanzprozesses wirken wiederum Determinanten ein, die diese beeinflussen. Zu nennen sind dort Einflüsse wie z.B. demografische Merkmale; Einstellung bedeutender Personen oder Marken; individuelle Persönlichkeitsmerkmale; Faktoren, die das Produkt, den Konsumenten oder das Unternehmen betreffen und Umwelt- bzw. Wettbewerbsfaktoren (Stüber, 2013, S. 53).

Rogers (2003, S. 248ff) bildet in seiner Diffusionstheorie – je nachdem wie schnell die Akzeptanz einer Innovation beim Probanden eintritt – fünf Kundensegmente, die insbesondere für die Einordnung neuer Technologien wie Deepfakes interessant sein können. Dazu gehören die folgenden Innovationsakzeptanztypen mit ihrem jeweiligen Marktanteil:

- **Innovatoren**

Unternehmungslustig – 2,5 % Marktanteil

Die Innovatoren gelten neugierig und probieren gern Neues aus. Voraussetzung, um ein Innovator zu sein, ist die Kontrolle über seine finanziellen Mittel, falls sich eine Innovation als Fehler erweist und die Fähigkeit, komplexe, technische Prozesse zu verstehen bzw. zu nutzen. Weiterhin darf der Umgang mit Ungewissheit und Risiken kein Problem darstellen. Durch die Ideen und Anregungen, die Innovatoren in das soziale Umfeld einbringen, spielen sie eine entscheidende Rolle bei der Verbreitung von Innovationen (Rogers, 2003, S. 248ff).

- **Early Adopters**

Seriös – 13,5 % Marktanteil

Early Adopters gelten als Meinungsführer. Rogers beschreibt sie als Personen, bei denen man sich erkundigen muss, bevor man eine neue Idee anwendet. Der Early Adopter muss die Unsicherheit von Anderen gegenüber der Innovation reduzieren, indem er die Innovation testet und seine eigenen Erfahrungen mit anderen teilt (Rogers, 2003, S. 248f).

- **Early Majority**

Bewusst – 34 % Marktanteil

Die Gruppe der Early Majority übernimmt neue Ideen erst kurz vor dem Durchschnittsbürger. Zugehörige nehmen selten Führungspositionen ein und tauschen sich lieber mit Gleichaltrigen aus. Die Funktion dieser Gruppe im Diffusionsprozess ist die Vernetzung. Sie können lange überlegen und profitieren von den Erfahrungen der Early Adopters mit der neuen Idee (Rogers, 2003, S. 249).

- **Late Majority**

Skeptisch – 34 % Marktanteil

Die Late Majority steht Innovationen skeptisch und vorsichtig gegenüber. Die „späte Mehrheit“ übernimmt neue Ideen erst, wenn der Großteil der Gesellschaft diese bereits angenommen hat. Gründe dafür könnten gesellschaftlicher Druck oder wirtschaftliche Notwendigkeit sein. Rogers spricht davon, dass ein Druck von Gleichgesinnten unbedingt nötig ist, um die anfängliche Überzeugung in eine reale Übernahme der Innovation zu wandeln (Rogers, 2003, S. 249f).

- **Laggards**

Traditionell – 16 % Marktanteil

Laggards – also Nachzügler – neigen dazu neuen Ideen misstrauisch zu sein und gelten als die Letzten, die neue Ideen akzeptieren. Ihre Meinungsbildung fundiert meist auf Entscheidungen von vergangenen Generationen und Traditionen. Aufgrund meist begrenzter Ressourcen muss sicher gegangen werden können, damit sich die Innovation lohnt (Rogers, 2003, S. 250).

Auch die Ergebnisse der Studie von Lee & Park (2009, S. 89) zeigen, dass die Einstellung der Kunden zu personalisierten Online-Angeboten stark vom sozialen Druck Angehöriger, Freunden und Kollegen abhängig ist und auch die bisherige Erfahrung mit Online-Einkäufen (in der Studie: Bekleidung) eine bedeutende Rolle spielt.

Diese vorgenommene Einteilung im Sinne der Diffusionstheorie kann mit der Unterteilung der Sinus Milieus respektive DIVSI Internet-Milieus (DIVSI, 2016) ergänzt werden. Hierbei werden Internetnutzer in

internetfern und internetnah unterschieden. Für das Erschließen der möglichen Zielgruppe von Deepfake-Marketing ist es nötig herauszufinden, welche Personengruppen dafür statistisch überhaupt in Frage kommen.

Die Bevölkerungsbefragung von PWC (2019) zu personalisierter Werbung ergab, dass Dreiviertel der Internetnutzer Sorgen haben, dass ihre Daten in falsche Hände geraten könnten. Auf das Thema Datenschutz wird im weiteren Verlauf dieser Arbeit im Kapitel 4.3. Abwägen der Vor- und Nachteile für Kunden und Marketer noch genauer eingegangen. Die Hälfte der Internetnutzer hat nichts gegen personalisierte Werbung, solange sie bei der Nutzung nicht stört. Immerhin schauen drei von zehn Personen lieber personalisierte statt standardisierte Werbung.

Diese Resultate decken sich mit jenen der aktuell veröffentlichten YouGov & Global Witness (2021) Studie. Nutzer scheinen derzeit von personalisierten Angeboten nicht begeistert zu sein. 57% der Befragten lehnt personalisierte Werbung sogar ab – nur 11% gaben an zufrieden zu sein, wenn ihre persönlichen Daten für die Ansprache mit personalisierten Anzeigen genutzt werden. Als Gründe wurden fehlende Transparenz bei der Datenbeschaffung und Datenschutzbedenken genannt.

2.7 Bisherige Einsatzmöglichkeiten von personalisierter Werbung

Zunächst werden die Formen und Einsatzgebiete von personalisierten Angeboten, mit denen Nutzer bereits Kontakt gehabt haben können analysiert. Der Data Scientist – Laurenz Wuttke – sieht je nach Anwendungsfall und technischen Möglichkeiten den Einsatz von Personalisierung in den Elementen Ansprache, Form, Inhalt, Kanal und Medium, Uhrzeit und Ort als sinnvoll (Wuttke, 2022).

Neben personalisierten Anreden in Newslettern oder auf Webseiten („Hallo Paulina“) können Werbebanner oder -filme gezielt ausgespielt werden. Befindet sich der Nutzer z.B. an einem bestimmten Ort können spezielle Werbeinformationen (beliebt: Rabattcoupons) versendet werden. Das sogenannte Location-based Marketing nutzt dafür Daten wie IP-Adresse, Systeme wie Beacons oder das Prinzip der Geolokalisierung via Mobilfunk, WLAN oder GPS (Schürmann, 2019).

Weit verbreitet für das Element „Inhalt“ sind sogenannte Recommender Systeme – also Empfehlungssysteme. Diese individualisierten Empfehlungssysteme bieten Vorschläge aufgrund der Präferenzen des Nutzers (Müller, 2005, S. 25). Vorteil dieser Systeme ist die Zeitersparnis, da die Auswahl der relevanten Informationen eingegrenzt wird. Kollmann (2019, S. 727) spricht in dem Zusammenhang von zwei unterschiedlichen Modellen, die das Recommendation-Konzept darstellen: **Content-based Recommendation**, generiert Empfehlungen durch die Auswertung von

bisher betrachteten und ausgelassenen Inhalten (Dörner 2003, S. 20 zitiert nach Kollmann, 2019, S. 727). **Collaborative Recommendation** hingegen ist die Prognose von möglichen „sinnvollen“ Produkten basierend auf den Bewertungen anderer Personen wie z.B. bei Amazon oder Netflix mit Anzeigen wie z.B. „andere Kunden kauften auch XY“ oder „Nutzer die diesen Film sahen, schauten auch XY“ (Kollmann, 2019, S. 728). Zur Generierung von dafür nützlichen Daten werden beispielsweise Data Mining, Informationsrückgewinnung aus Datenbanken und auch aktuelle Informationen zum Kauf/der Suche eingesetzt (Heinemann, 2017, S. 11). Das große Ziel dieser Empfehlungen ist die Maximierung des Kundeneinkaufs im Sinne von Up-Selling (hochwertigere Produkte), Cross-Selling (weitere andere Produkte) und More-Selling (große Mengen eines Produkts) und somit auch die Maximierung des Umsatzes (Heinemann, 2017, S. 11ff).

Insbesondere der Bewegtbildmarkt stellt für die im weiteren Verlauf dieser Arbeit näher erläuterten Deepfakes großes Potential für Personalisierungsmöglichkeiten dar, weshalb das nächste Kapitel Aufschluss über die Relevanz von Videowerbung geben soll und einige Modelle dazu näher erläutert.

2.8 Videowerbung

Beisch et al. (2021, S. 518) nennen den Trend einen regelrechten „Bewegtbildboom“, der sich auch im Jahr 2022 weiter ausbreitet. Fast 90% der Menschen werden täglich mit Bewegtbildinhalten erreicht, was laut Beisch et al. (2021) vor allem an der Corona Pandemie liege.

Diesen Boom nutzen Marketer nicht erst seit Ausbruch von COVID-19. Bereits vor 10 Jahren riefen Reisig et al. (2011, S. 245) dazu auf, die Vorteile von Videowerbung nicht zu unterschätzen. Laut diverser Studien ließen sich mit Online-Video-Werbung bessere Ergebnisse als mit klassischer TV-Werbung verzeichnen. Als Gründe nannten die Autoren die erhöhte Teilhabe der Nutzer und eine höhere Erinnerung an die Werbeinhalte, da die Nutzer Inhalte häufig eigenständig abrufen (Reisig et al., 2011, S. 245; Uyenco & Kaplan, 2010). Und sogar über die Hälfte der Befragten einer HubSpot Studie (Kolowich Cox, 2017) gaben an, mehr Videoinhalte von Unternehmen sehen zu wollen.

Für die später folgende Produktion der Deepfakes in dieser Arbeit wird mit bereits existierenden Werbeclips gearbeitet, weshalb zuvor kurz auf die Ausspielungsmöglichkeiten dieser visuellen Werbung eingegangen wird. Videos lassen sich generell in in-Page und in-Stream Werbeformate einteilen. Bei in-Page Anzeigen handelt es sich um klassische Flächen auf Webseiten, die mit Werbespots versehen werden (z.B. Banner). In-Stream Anzeigen werden vor, während oder nach dem Video-Content eingeblendet und haben einen hohen Aufmerksamkeitswert, denn der eigentliche Video-Content wird durch die in-Stream Anzeige zum Teil unterbrochen (Reisig et al., 2011, S. 246). Die Varietät des Einsatzes ist somit hoch.

Auf inhaltlicher Ebene wird in Anlehnung an die Studie von Coker et al. (2017) an dieser Stelle zwischen straight-sell (faktenbasierten) Videospots und Storytelling-Videospots unterschieden.

Belch & Belch (2018, S.312) beschreiben die straight-sell Strategie als eine „[...] a straightforward presentation of information concerning the product [...]“ mit Fokussierung auf Merkmalen, Eigenschaften und Vorteilen des beworbenen Produkts. Demgegenüber steht der Storytelling Ansatz, welcher mit dem Aufkommen von sozialen Medien interessanter für Unternehmen geworden ist, da sie in ihren Werbevideos ihre eigene Geschichte erzählen können und in letzter Instanz Kunden gewinnen und an sich binden können (Pulizzi, 2012, S. 1). Es scheint zu funktionieren, denn Lundqvist et al. (2013) erkannten in ihrer Studie, dass Probanden, die mit Storytelling konfrontiert waren, die Werbung deutlich positiver beurteilten und sogar bereit waren mehr Geld für das Produkt zu zahlen.

Für das Deepfake-Marketing bedeutet das also, dass informative, spannende und im besten Fall auch emotional ansprechende Videoinhalte generiert werden müssen.

2.9 Werbung im Wandel

Wie in den letzten Kapiteln ersichtlich wurde, entwickelt sich die Werbebranche ständig weiter. Durch technische Innovationen und gesellschaftliche Ereignisse verändern sich zeitgleich auch die Anforderungen der Nutzer an die Werbung. So entwickelten sich mit dem Aufkommen der sozialen Netzwerke neue Strukturen der Vernetzung und Informationsgewinnung im Internet. Nutzer generieren selbst Informationen und andere Nutzer konsumieren diese. Nicht erst mit dem digitalen Zeitalter werden deshalb auch die Rufe bezüglich Fehlinformationen bzw. Fake News lauter. Nach der analogen Manipulation von Werbung (z.B. durch

Make-Up, Licht oder Kameralinsen) ist der Wandel über die digitale Bildmanipulation (z.B. durch Photoshop oder Computer Generated Imagery (CGI)) jetzt in der dritten Generation – der synthetischen Manipulation (z.B. durch GANs und KI) angelangt (Campbell et al., 2021, S. 2 ff).

Einige Autoren sprechen in dem Zusammenhang auch von einem „postfaktischen Zeitalter“, in dem sich die Welt aufgrund fortschreitender Digitalisierung der Gesellschaft gerade befindet (Appel, 2020; Hendricks & Vestergaard, 2018; Temir, 2020). Die Gesellschaft für deutsche Sprache wählte bereits 2016 das Wort „postfaktisch“ zum Wort des Jahres mit der Begründung, „[...] dass es in politischen und gesellschaftlichen Diskussionen heute zunehmend um Emotionen anstelle von Fakten geht. (...) Nicht der Anspruch auf Wahrheit, sondern das Aussprechen der „gefühlten Wahrheit“ führt im ‚postfaktischen Zeitalter‘ zum Erfolg.“ (Gesellschaft für deutsche Sprache e. V., 2016). Auch die damalige Bundeskanzlerin Angela Merkel versuchte sich an einer Definition. Mit ihren Worten heißt es, „[...] wir lebten in postfaktischen Zeiten. Das soll wohl heißen, die Menschen interessieren sich nicht mehr für Fakten, sie folgen allein den Gefühlen.“ (Der Tagesspiegel, 2016) und stellt mit dieser Aussage Emotionen ganz klar vor Fakten. Postfaktisch bedeutet also, Fakten sind in der heutigen Zeit unwichtiger geworden und legen damit die Weichen für die Ausbreitung von Fake News, Clickbaiting, Deepfakes und synthetischen Medien.

Fake News sorgten insbesondere seit der US-Wahl von Donald Trump im Jahr 2016 für Diskussionen über die Glaubwürdigkeit von (politischen) Aussagen. Die Definition zu Fake News von Allcott & Gentzkow (2017, S. 213) wird in der Literatur am häufigsten zitiert und beschreibt Fake News „[...] to be news articles that are intentionally and verifiably false, and could mislead readers.“. Der Duden geht in seiner Definition sogar noch einen Schritt weiter und nennt explizit Orte, an denen Fake News kursieren: „In den Medien und im Internet, besonders in sozialen Netzwerken, in manipulativer Absicht verbreitete Falschmeldungen.“ (Duden (c), 2022). In der Literatur wird der Fake News Begriff noch ausführlicher diskutiert, was für diese Arbeit jedoch zu weit führt (Pritzkau & Schade, 2021, S. 116ff).

Wardle (2017) kategorisiert Fake News in sieben Formen: irreführender Inhalt, hochstaplerischer Inhalt, erfundener Inhalt, falsche Beziehung, falscher Kontext, manipulierter Inhalte und Satire. Die in dieser Arbeit angestrebte Forschung zum Thema Deepfakes reiht sich laut Dobber et al. in die Kategorie „Desinformation“ ein (Dobber et al., 2021, S. 71). Die Vielfältigkeit von Fake News unterstreicht an dieser Stelle die Wichtigkeit des Abwägens von Falschnachrichten und der Wahrheit.

Für Jaster & Lanius (2019, S. 33f) bieten Fake News gleich zwei große Problemfelder, die es im Sinne der Identifikation zu lösen gilt: (1) Fake News sind falsch oder irreführend **[Mangel an Wahrheit]** und (2) Verfasser von Fake News agieren mit Täuschungsabsicht **[Mangel an Wahrhaftigkeit]**. Viel wichtiger als die

Definition ist deshalb die Frage der Erkennung von Fake News. Wie können Laien Fake News im Internet erkennen und welche Anhaltspunkte können für die Erkennung berücksichtigt werden?

Zhou & Zafarani (2020, S. 1) definieren in ihrer Übersicht zu Fake News vier Anhaltspunkte zur Erkennung von Fake News: (1) das enthaltene falsche Wissen, (2) der Schreibstil, (3) die Verbreitungsmuster und (4) die Glaubwürdigkeit der Quelle. Nutzer müssen die Fakten anhand von Experten und/oder einer Gruppe von Personen, die als Faktenchecker fungieren auf ihren Wahrheitsgehalt überprüfen (S. 8ff). Der Schreibstil von Fake News kann durch Charakteristika des maschinellen Lernens identifiziert werden. Dadurch kann z.B. die Quote oder Anzahl von negativ konnotierten Wörtern ermittelt und zur Bewertung durch neuronale Netzwerke herangezogen werden (S. 14ff). Im dritten Anhaltspunkt – der Erkennung durch Verbreitungsmuster – werden neuronale Netzwerke darauf trainiert bestimmte Verbreitungsmuster zu erkennen. Durch diese Nachrichtenkaskaden (baumartige Strukturen zur Verbreitung von Inhalten) können Artikel als wahr oder falsch eingeordnet werden (S. 20f). Der letzte Punkt umfasst die Glaubwürdigkeit der Quelle. Zhou & Zafarani (2020, S. 26) gehen davon aus, dass die Bewertung der Glaubwürdigkeit der Nachrichtenquellen (z.B. Autoren, Webseiten, Vereine) ein erster Hinweis darauf sein kann, ob diese Quellen häufiger im Verdacht der Falschmeldungen stehen.

Die aktuelle Studie von Meßmer et al. (2021) zeigt, dass zwar oft richtig eingeschätzt wird, ob eine Quelle vertrauenswürdig ist, jedoch Zweifel an der Unabhängigkeit des Journalismus gegenüber anderen Institutionen besteht. So bekamen Aussagen wie „Lügenpresse“-Vorwürfe; die Medien würden Hand in Hand mit Institutionen arbeiten, um die Meinungen zu manipulieren oder systematisches Belügen Zuspruch bei den Probanden. Allerdings zeigten die Ergebnisse der Studie auch den Willen, nur für wahr befundene Informationen an andere weiterzuleiten und Falschinformationen zu ignorieren (S. 4ff).

Medienschaffende nutzen zur Erreichung von Werbezielen verschiedene KPIs (Key-Performance-Indicator) zur Messung des Erfolgs von z.B. Bannern oder Werbeanzeigen. Einer davon ist die Klickrate/Click-Through-Rate (CTR), die das Verhältnis von angeklickten zu angesehenen Werbemitteln angibt (Kreutzer, 2021, S. 106). *„Potenzielle Leserinnen und Leser zu ködern, um Klicks zu generieren, ist heute für viele Medienunternehmen – unabhängig von deren Seriosität – alltäglich.“*, so Mayer (2020, S. 69). Dafür werden vor allem Überschriften und Bilder aufmerksamkeitswirksam verfasst, um in erster Instanz viele Klicks zu generieren und zweitrangig die geweckten Interessen zu befriedigen (Mayer, 2020, S. 67). Der Trick dahinter ist das Nutzen der menschlichen Neugierde. Nach der Loewenstein'schen Theorie der Informationslücke verspüren Menschen eine Lücke zwischen dem, was sie aktuell wissen und dem, was sie wissen wollen (Loewenstein, 1994, S. 75; vgl. dazu auch Anand et al., 2019, S. 1; Chen et al., 2015, S.

16). Dieser Effekt ist wegen der Neugierde auf das Deepfake-Marketing ebenfalls möglich.

An dieser Stelle wird es nötig, tiefer in das Thema Deepfakes und synthetische Werbung einzusteigen. Unter synthetischer Werbung wird manipulierte Werbung, die durch künstliche und automatische Erstellung und Veränderung von Daten erzeugt oder bearbeitet wird verstanden (Campbell et al., 2021, S. 1).

Einsatzgebiete von Personalisierungen können nach Luber & Litzel (2021) insbesondere im Bereich künstlich generierter, visueller Inhalte liegen. Das Hauptaugenmerk der synthetischen Medien liegt in dieser Arbeit jedoch auf der Thematik der Deepfakes. Das folgende Oberkapitel beschäftigt sich mit diesen. Es werden in den kommenden Abschnitten Begriffe abgegrenzt, technische Fortschritte eingeordnet, Funktionsweisen erklärt, auf Erkennungsmechanismen eingegangen, aktuelle Deepfake-Beispiele gegeben und Risiken sowie Chancen dieser neuen Technologie erörtert. Das folgende Kapitel zu Deepfakes stellt neben der Personalisierung die zweite Grundsäule, der im Anschluss diskutierten Werbeform des Deepfake-Marketings dar.

3. Deepfakes: Fake News oder Zukunftsmusik?

Deepfakes werden zu den synthetischen Medien zugeordnet. Digital manipulierte Bilder, Videos und Audiosequenzen werden in der heutigen Zeit nicht nur hinter den Kulissen von großen, internationalen Filmproduktionen oder eigens dafür spezialisierten Unternehmen generiert. Durch den Einsatz von künstlicher Intelligenz (KI) sind die Eintrittsbarrieren für die Manipulation und Generierung von Inhalten gesunken. Mit gut ausgestatteter Hardware, Grundwissen und Geduld lassen sich bereits jetzt täuschend echte Ergebnisse erzielen. Dabei entsteht ein Problem: Diese neuen, digitalen Technologien erschweren die Erkennung von „echten“ und „falschen“ Inhalten. Eine der aktuell kontrovers diskutierten Beispiele dafür sind Deepfakes. In den folgenden Kapiteln wird diese Technologie näher erläutert.

3.1 Definitionsversuche

Das Phänomen der Deepfakes existiert noch nicht einmal zehn Jahre. Abgeleitet aus den beiden Begriffen „Deep Learning“ und „Fake“ entstanden so die „Deepfakes“, um die sich die folgenden Kapitel drehen werden. Ende 2017 veröffentlichte ein Reddit Nutzer namens „deepfakes“ den ersten Deepfake in Form pornografischen Materials mit prominenten Gesichtern wie z.B. Emma Watson oder Scarlett Johansson (Kietzmann et al., 2020, S. 2). Dadurch, dass Deepfakes noch nicht so lange existieren ist der Literaturbestand zu diesem Zeitpunkt auch überschaubar. Neben wissenschaftlichen Ausarbeitungen wird sich für diese Arbeit deshalb auch auf andere Medienquellen zum Thema Deepfake wie z.B. Pressemitteilungen,

technische Berichte und audio-visuellem Material auf Internetplattformen berufen. Eine einheitliche Definition von Deepfakes kann in der Literatur zurzeit noch nicht gegeben werden. Im Folgenden werden deshalb verschiedene Definitionsansätze zusammengetragen:

Deepfakes sind laut Westerlund (2019, S. 39) hyperrealistische Videos, die durch den Einsatz von KI audio-visuelle Situationen schaffen, die so nie passiert sind. Mirsky & Lee (2020, S. 3) verstehen Deepfakes als „*Believable media generated by a deep neural network*“. Masood et al. (2021, S. 1) fügen hinzu, damit Deepfakes synthetisch generierte Videos und Audios mit dem Einsatz von künstlicher Intelligenz sind. Deepfakes nutzen KI also, um Attribute (wie z.B. Gesicht, Stimme, Geschlecht) von einer Person auf eine andere Person zu übertragen (Campbell et al., 2021, S. 1; Floridi, 2018; Karnouskos, 2020). Basierend auf der maschinellen Lerntechnik namens Deep Learning werden Deep Neural Networks (DNNs) trainiert, um gefälschte Inhalte möglichst authentisch wirken zu lassen (Kietzmann et al., 2021, S. 2). Weiter beschreiben Kietzmann et al. DNNs als künstliche, von menschlichen Gehirnen inspirierte Konstrukte, die individuell trainiert werden können, damit Daten aus Bildern, Audio und Videos automatisch gekoppelt, ersetzt oder über Videodaten darübergerlegt werden können. Das große Ziel ist dabei, einen hyperrealistischen, aber trotzdem gefälschten alternativen Inhalt erstellen zu können. Möglich wird dies, da der Algorithmus kontinuierlich dazu lernt und so das Output Material verbessert (Maras & Alexandrou, 2019, S. 1f). Die Algorithmen trainieren dabei auf Grundlage von mehreren

hundert bis tausenden Bilddaten der Zielperson (Koopman et al., 2018, S. 133).

Anhand dieser Sammlung von Definitionsversuchen wird bereits deutlich, dass das Phänomen Deepfakes komplex ist. Im nächsten Kapitel werden einzelne Aspekte spezifischer erklärt.

3.2 Fundamentale Prozesse der Deepfake-Technologie

In diesem Abschnitt geht es darum, grundlegende Prozesse innerhalb der Deepfake Technologie zu erklären und die Zusammenhänge abzubilden.

Der Begriff der künstlichen Intelligenz steht für Laien oftmals synonym zu den sich unter dem Deckmantel befindlichen Prozesse wie Neuronale Netzwerke oder Deep Learning. Um diese Begriffe abzugrenzen, werden im Folgenden die Prozesse dahinter kurz erklärt:

- **Machine Learning**

Das Machine Learning (ML) umfasst als Bestandteil der KI verschiedene Methoden des Selbstlernens. Tom Mitchell (US-Informatiker) definiert Machine Learning beispielhaft als *„A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E .“* (Mitchell, 1997, S. 2), welches er mit einem Schachspiel vergleicht. Die Performance des Computers (P) beim Schachspiel (T) wird durch Erfahrung (E) stets verbessert, wenn er viel spielt und daraus Erkenntnisse gewinnen und analysieren kann.

Beim ML kommen sogenannte selbst-adaptive Algorithmen zum Einsatz. Diese lernen selbstständig, ohne Einwirken eines Menschen und ziehen Rückschlüsse aus dem Gelernten mit dem Ziel sich kontinuierlich zu verbessern (Kreutzer & Sirrenberg, 2019, S. 6). Für den Lernprozess sind folglich große Mengen an Daten (Big Data) nötig. Alles, was digital hinterlegt ist, kann in einen ML-Algorithmus eingespeist werden (Hao, 2018). Innerhalb des ML werden drei große Lernmethoden unterteilt (Kreutzer & Sirrenberg, 2019, S. 7ff):

- **Supervised Learning (beaufsichtigt)**

Bei dieser Methode werden der KI bereits „richtige“ Lösungsmuster vordefiniert. Die Daten sind beschriftet und die KI muss die Algorithmen nur noch darauf trainieren, diese zuverlässig zu erkennen. Beim Abspielen eines Films auf Streaming-Plattformen wird dem Algorithmus so direkt gesagt, dass er ähnliche Angebote zu genau diesem Film finden soll.

- **Unsupervised Learning (unbeaufsichtigt)**

Anders als beim supervised Learning kennt die KI beim unsupervised Learning keinerlei Lösungen vorab. Die Algorithmen müssen Ähnlichkeiten in den unbeschrifteten Daten (Muster) selbst identifizieren. Innerhalb des Marketings kann diese Methode Muster im Einkaufsverhalten erkennen (Stichwort: Recommender Systeme).

- **Reinforcement Learning (verstärkend)**

Am Anfang dieser Methode existiert keine mustergültige Lösung. Die Algorithmen arbeiten hier nach dem Trial-and-Error Prinzip, indem sie verschiedene Lösungswege testen und bei Erfolg

weiterverfolgen (und bei Niederlage verwerfen und einen neuen Weg suchen). Ein Beispiel dafür ist das Go-Spiel mit AlphaGo.

- **Künstliche neuronale Netze**

Künstliche neuronale Netze (KNNs) arbeiten auf Grundlage von ML-Algorithmen und werden so trainiert und optimiert. Dabei handelt es sich genauer betrachtet um ein Geflecht, welches mit der Informationsverarbeitung des menschlichen Gehirns zu vergleichen ist. Im übertragenen Sinne bestehen Neuronale Netze, deshalb aus Knoten (Neuronen) und Kanten (Synapsen). Der Vorteil dieser mehrschichtigen Netzwerke ist die verbesserte Erkennungsfähigkeit von Zusammenhängen gegenüber einfachen Algorithmen des ML (Kreutzer & Sirrenberg, 2019, S. 12ff).

Neuronale Netze bestehen aus drei Ebenenarten, in welchen sich Knotenpunkte befinden, die Informationen aufnehmen oder modifizierte Informationen an den nächsten Knotenpunkt weitergeben:

Input-Units sind die Eingangsdaten, die es zu analysieren gilt. Die darauffolgenden **Hidden-Units** verarbeiten die Informationen aus der vorherigen Ebene und stellen das „Innere“ eines neuronalen Netzes dar. Die Informationen können in den Hidden-Units in mehreren Schichten hintereinander angeordnet sein. Die **Output-Units** stellen die Ausgangsinformationen bereit und präsentieren das vorläufige Ergebnis des KNNs (z.B. bei einem Erkennungsalgorithmus die Klassifikation „Mann“ oder „Frau“). Der eigentliche Lernprozess wird innerhalb des neuronalen Netzes durch die Gewichtung der

Verbindungen (positiv/negativ für das Ziel), der Entwicklung neuer bzw. Eliminierung alter Verbindungen und der iterativen Optimierung bestehender Verbindungen erreicht (Buxmann & Schmidt, 2019, S. 14; Gentsch, 2019, S. 36; Kietzmann et al., 2020, S. 5; Kreutzer & Sirrenberg, 2019, S. 5; Luber & Litzel, 2017; Mirsky & Lee, 2020, S. 5).

Die nachfolgende Grafik soll das Prinzip der Ebenen, Schichten und Kanten an dieser Stelle verdeutlichen:

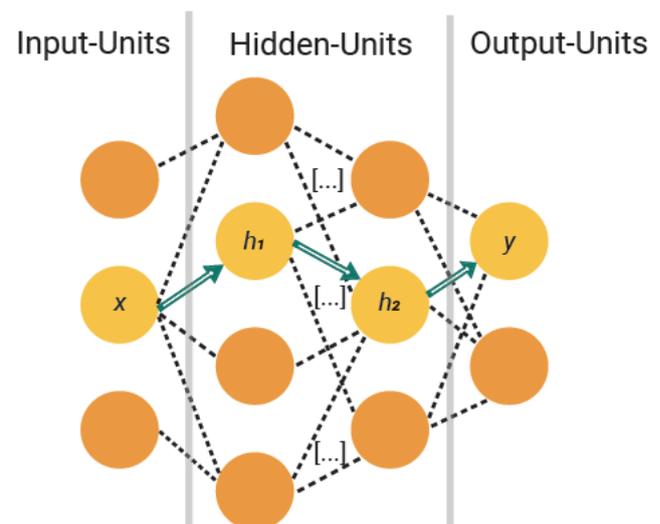


Abbildung 3 Prinzip der neuronalen Netze (In Anlehnung an Buxmann & Schmidt 2019, S. 14)

- **Deep Learning**

Die zuvor erwähnten neuronalen Netzwerke bestehen meist aus nur 2 bis 3 Hidden-Units. Tiefe neuronale Netze, die auch als Deep Learning bezeichnet werden bestehen aus mehr als 10 dieser Hidden-Layers (Treder & Eter, 2018, S. 714f). Der Zusatz „deep“ steht somit für die Erweiterung der Komplexität der Schichten des neuronalen Netzes (Kreutzer & Sirrenberg, 2019, S. 8). In der Literatur wird in dem Zusammenhang auch von Deep Neural Networks (DNNs)

gesprachen. Der Vorteil von DNNs und dem Deep Learning besteht darin, das tief verborgene Muster und Korrelationen in den Daten erkannt und die Daten miteinander verbunden werden (Kreutzer & Sirrenberg, 2019, S. 8). Das Ziel kann z.B. sein, hyperrealistische aber dennoch gefälschte Inhalte in Form von Bild, Video oder Audio zu erzeugen (Kietzmann et al., 2021, S. 2). Die folgende Grafik verdeutlicht die Komplexität von DNNs gegenüber herkömmlichen künstlichen neuronalen Netzen:

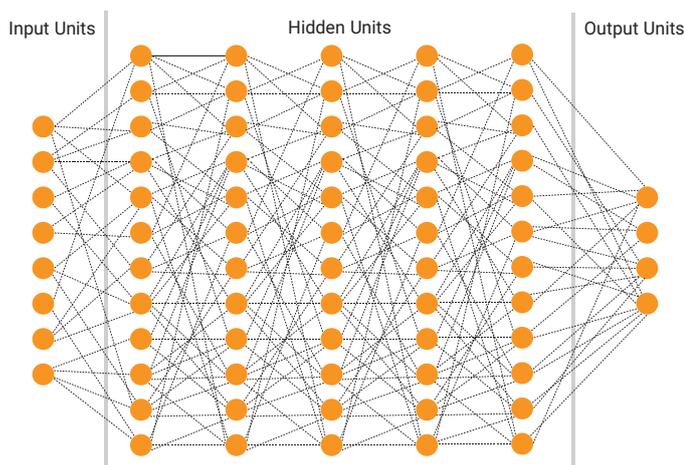


Abbildung 4 Beispielhafte Darstellung eines Deep Neural Network (In Anlehnung an rsipvision, 2015)

- Generative Adversarial Networks**
 GANs (zu deutsch: generierendes gegnerisches Netzwerk) bestehen aus gleich zwei künstlichen neuronalen Netzwerken. Creswell et al. (2018) veranschaulichen die Funktionsweise anhand der exemplarischen Darstellung von Kunstfälscher und Kunstexperte. Der Fälscher – also der Generator – erstellt Fälschungen mit dem Ziel realistische Bilder zu generieren. Der Kunstexperte – Diskriminator in der Fachsprache genannt – bekommt sowohl die gefälschten als auch echte Daten und muss diese unterscheiden. Beide Modelle sind logisch gekoppelt und

trainieren sich gegenseitig. Durch jede Iteration lernen die Modelle voneinander und verbessern den Output (Albahar & Almalki, 2019, S. 3245; Creswell et al., 2018, S. 53; Goodfellow et al., 2014; Luber & Litzel, 2021; Mirsky & Lee, 2020, S. 6f; Tong et al., 2020, S. 266).

Im folgenden Kapitel geht es um die zeitliche Entwicklung von Deepfakes. Bevor sich Deepfakes erstmalig über Reddit ausbreiten konnten, forschten wissenschaftliche Teams weltweit an künstlich intelligenten Systemen.

3.3 Deepfake-Evolution

Nach der Definition wichtiger Begriffe innerhalb der Deepfake-Technologie wird nun die Evolution der Deepfakes zusammengefasst.

Die technische Grundlage für Deepfakes schuf die wissenschaftliche Arbeit von Goodfellow et al. (2014) zu GANs – der Technologie hinter den Deepfakes. Seit dem Jahr 2014 ist jedoch viel passiert. Viele Forscher verbesserten die Technologien in diesen Bereichen stetig, mit dem Ziel die Qualität der Resultate weiter zu steigern (Karras et al., 2018; Liu & Tuzel, 2016; Radford et al., 2016). Auch Technik-Giganten wie Google und Samsung entwickeln eigene KI-Verfahren und treiben die Deepfake-Forschung voran. So generiert die Google-Schwester-KI DeepMind 2018 täuschend echte Burger (Brock et al., 2019) und Samsungs KI modifiziert ein Jahr später statische Fotos anhand realer Mimik (Zakharov et al., 2019). Der rot markierte Burger stellt dabei den KI-generierten Burger

aus den Beispielfoto-Burgern dar. Im Fall der Samsung KI werden die Gesichtszüge einer Person identifiziert und auf dessen Grundlage auf das Kunstwerk übertragen.

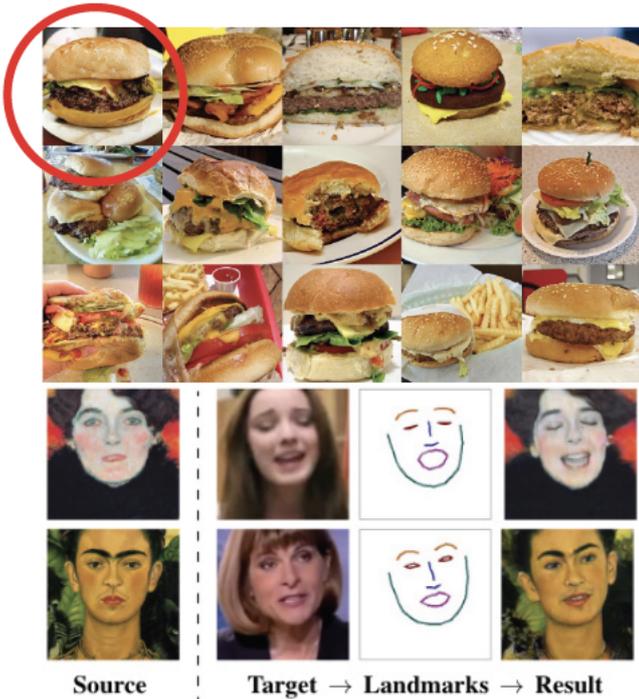


Abbildung 5 oben: DeepMind KI (Quelle: Brock 2019, S.16); unten: Samsung KI (Quelle: Zakharov 2019, S.1)

Ein weiterer Meilenstein gelang Microsoft und chinesischen Forschern mit der FaceShifter KI. Das Framework legt großen Wert auf hohe Genauigkeit und berücksichtigt Okklusionsfehler, indem die KI verschwommene oder von Brillen oder Haaren verdeckte Bereiche selbstständig wiederherstellt (L. Li et al., 2020). Die Quelle (Source) befindet sich im folgenden Beispiel auf der linken Seite, das Zielfoto (Target) in der Mitte und rechts ist das KI-generierte Ergebnis. Insbesondere am rechten Beispiel lässt sich die Berücksichtigung von Okklusionsfehlern anhand des Netzes vor dem Gesicht gut erkennen.



Abbildung 6 Beispiel für FaceShifter (Quelle: Li et al., 2020, S.1)

Dem Team von Disney Research in Kooperation mit der Technischen Hochschule Zürich gelang im selben Jahr ein weiterer Clou: Sie entwickelten die erste Methode, mit der Deepfakes in Megapixel-Auflösung generiert werden (Naruniec et al., 2020). Bisherige Ergebnisse lagen meist in 256 x 256 Pixel Auflösung vor und Disneys Ansatz liefert nun Deepfakes in 1024 x 1024 Pixel Auflösung. Im folgenden Beispiel erkennt man deutliche Qualitätsunterschiede der Face Swap Methoden. Verschiedene Autoren spekulieren außerdem darüber, dass sich diese Deepfake-Technik bald in Filmproduktionen etablieren könnte (Schreiner, 2021) und Potential zum „Game-Changer“ (Drees, 2020) in Hollywood habe.



Abbildung 7 Vergleich der Face Swapping Methoden aus Naruniec et al. (2020)

3.4 Arten und Funktionsweisen von Deepfakes

Deepfakes lassen sich in zwei Kategorien einteilen. Diese Kategorisierung schließt Mischformen jedoch nicht aus. Demnach gibt es Audiomanipulation und optische Manipulation (Masood et al., 2021). Zur Audiomanipulation zählen z.B. Text-To-Speech und Voice Swapping. In dieser Arbeit wird sich jedoch hauptsächlich auf die visuelle Manipulation von Inhalten insbesondere Face Swapping fokussiert, weshalb die Untergruppierungen der visuellen Deepfakes umfangreicher ausfallen.

Die Gruppe der visuellen Deepfakes lässt sich in fünf Typen unterteilen. Im Folgenden wird die Grundidee dahinter erklärt und anhand von Beispielen verdeutlicht. In der Literatur findet man dazu weitere unterschiedliche Versuche die verschiedenen Deepfake-Typen zu differenzieren (Rahman et al., 2022, S. 13; Tong et al., 2020, S. 265ff; Xinwei et al., 2021, S. 367ff; T. Zhang et al., 2020, S. 67ff). Im Anschluss dazu sollen die technischen Prozesse der einzelnen Deepfake-Arten spezifiziert werden.

- **Face Swapping**

Beim Face Swap wird das Gesicht einer Person im Quellvideo mit dem Gesicht einer Zielperson ausgetauscht (Nguyen et al., 2022, S. 1). In der Regel werden dabei drei Stufen durchlaufen.

In der **ersten Stufe** soll das Gesicht eines Videos erkannt und eben dieses Gesicht aus der Sammlung der entstandenen, einzelnen Gesichtsframes ausgewählt werden. Das Kriterium ist hierfür, die Ähnlichkeit des Aussehens und der Posen des Gesichts. Dies geschieht für das Ausgangsgesicht und Zielgesicht. Im **zweiten Schritt** ersetzt das KI-Modell die Gesichtskomponenten (Augen, Nase, Mund) und nimmt Anpassungen zur Beleuchtung und Farbe des Zielbildes vor, damit die Unterschiede zwischen den beiden Materialien möglichst gering sind und ein im Idealfall nahtloser Übergang entsteht. **Zuletzt** wird die Rangfolge der Bilder des Zielgesichts ermittelt, indem berechnet wird, wie viel sich die Übereinstimmungsabstände überlappen müssen (Bitouk et al., 2008; Masood et al., 2021, S. 7; Nirkin et al., 2017, S. 2).

Neben dem bereits aufgeführten Beispiel von Naruniec et al. (2020) – Abbildung 7 – bieten Nirkin et al. (2017) in ihrer Publikation ebenfalls eine übersichtliche Darstellung dieses Prozesses:

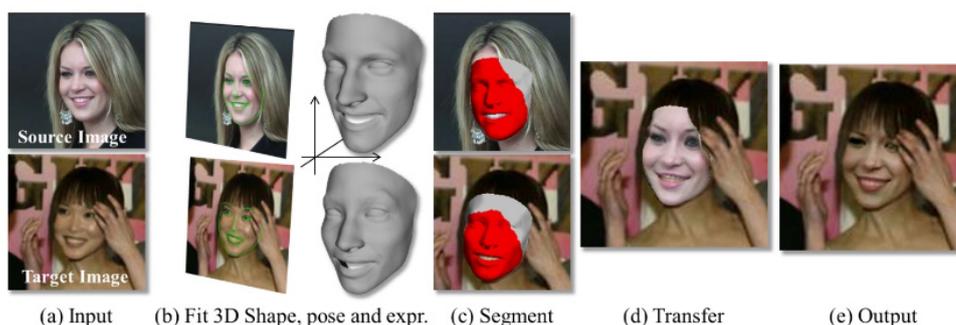


Abbildung 8 Methode des Face Swapping im Überblick (Quelle: Nirkin et al., 2017, S. 3)

Der technische Hintergrund dieser Prozesse liegt in der Nutzung von Deep Neural Networks (DNNs). Das Ziel dieser gehirnähnlichen Strukturen liegt darin die Stärke oder Gewichtung von Verbindungen stetig zu aktualisieren, um die Ergebnisse zu optimieren (Kietzmann et al., 2020, S. 5). Im Falle der Deepfakes funktioniert dies durch die gemeinschaftliche Arbeit von Autoencoder-Decoder-Paaren. Als Encoder wird dabei das Extrahieren von Gesichtsmerkmalen bezeichnet. Der Decoder ist für die Konstruktion eines Gesichts zuständig. Der latente Raum bezeichnet den verborgenen Raum bzw. Informationsengpass, den ein Autoencoder benötigt, um ernsthaft zu lernen und nicht nur die Merkmale auswendig zu lernen. Für einen Austausch des Gesichts werden zwingend zwei Encoder-Decoder Paare

benötigt. Laut Masood et al. liegt der Trick darin, jeden Encoder zunächst auf dem Quellmaterial und dann auf dem Zielmaterial zu trainieren. Anschließend werden die Decoder vertauscht und so bildet der Decoder des Zielbildes mit den Merkmalen des Quellbild-Encoders im latenten Raum das manipulierte Ergebnis. Vereinfacht gesagt bedeutet das: Zwei Paare nutzen das gleiche Encoder-Netzwerk mit dem Ziel, Ähnlichkeiten im Gesicht zu identifizieren und daraus zu lernen (Kietzmann et al., 2020, S. 5f; Y. Li et al., 2022, S. 75; Masood et al., 2021, S. 7f; Mirsky & Lee, 2020, S. 6; Nguyen et al., 2022, S. 3; T. Zhang et al., 2020, S. 68).

Die folgende Übersicht zeigt die Verwobenheit dieses Prinzips noch einmal visuell:

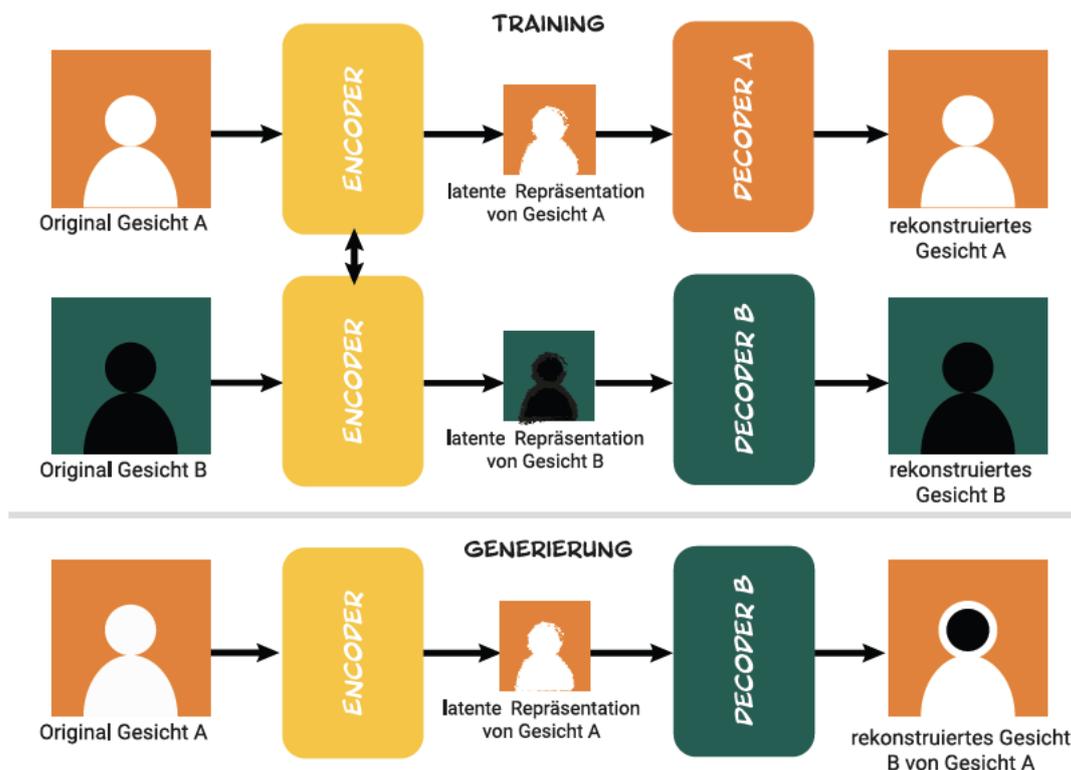


Abbildung 9 Auto-Encoder-Decoder Prinzip: Nutzung desselben Encoder-Decoder-Paares für die Extrahierung von Gesichtsmerkmalen mit Tausch der Decoder für die Generierung des Deepfakes (In Anlehnung an Masood et al., 2021, S. 8)

- **Lip Syncing**

Die Lippensynchronisation umfasst das Synthetisieren von Videomaterial aus Audiospuren im Bereich des Mundes (Suwajanakorn et al., 2017, S. 2). Hierbei wird der Mund so manipuliert, damit es aussieht, als würde eine Person einen beliebigen Text sprechen, den die Person so nie gesagt hat. Für ein natürliches und überzeugendes Ergebnis müssen passende Lippenbewegungen und das Aussehen des Mundes erzeugt werden (Masood et al., 2021, S. 11f). Die Methode von C. Zhang et al. (2021) nimmt beispielsweise eine Audioaufnahme und einen Videoclip dessen Gesicht manipuliert werden soll als Input. Die FACIAL-GAN synthetisiert Lippenbewegungen, Kopfhaltungen und das Blinzeln der Augen und erzeugt ein realistisches Video.

Die technischen Ansätze beim Lip Syncing sind vielfältig, weshalb die zuvor aufgeführte FACIAL-GAN an dieser Stelle aufgrund qualitativ hochwertiger Ergebnisse als Beispiel dienen wird. Die FACIAL-GAN analysiert phonetische, kontextbezogene

und personale Informationen. Dabei werden implizite und explizite Informationen zu z.B. Kopfhaltung, Action Units für Augenblinzeln (AU45), Texturen oder Beleuchtung gesammelt, um von den zeitlichen Zusammenhängen und phonetischen Merkmalen bei der Konstruktion des Zielvideos zu profitieren. Vorab muss das Quellvideo deshalb in ein 3D-Modell überführt werden, damit die Parameter (Augenblinzeln, Kopfhaltung, Ausdruck, Identität, Textur, Beleuchtung) im Anschluss kombiniert werden und zum Rendering des Zielvideos führen (C. Zhang et al., 2021).

Die folgende Grafik veranschaulicht diesen Prozess:

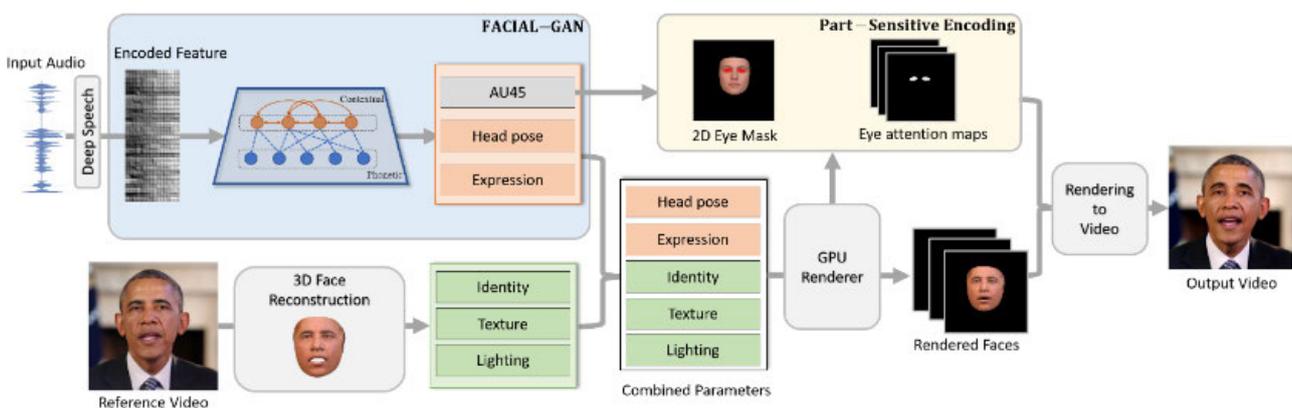


Abbildung 10 Arbeitsweise der FACIAL-GAN (Quelle: C. Zhang et al., 2021, S. 3)

- **Puppet Mastery**

Eine weitere Ausprägung der Deepfakes kann als Marionettenspiel bezeichnet werden. In der Literatur findet auch der Name Reenactment-Ansatz Einsatz und verfolgt das Ziel, Gesichtsbewegungen einer Quellperson auf Videos oder Bilder einer anderen Person zu übertragen (Thies et al., 2018, S. 1). Limitationen liegen bei verschiedenen Ansätzen vor allem darin, dass nur die Mimik sich verändert und die Kopfhaltung starr und unnatürlich wirkt, so Thies et al. weiter.

Eine der bekanntesten Techniken zu Marionetten-Deepfakes geben Thies et al. (2016) mit dem Face2Face Modell (vgl. auch T. Zhang et al., 2020, S. 68). Dieser neue Ansatz zur Gesichtsdarstellung in Echtzeit erfordert die Analyse der Gesichtsausdrücke im Quell- und Zielvideo unter Berücksichtigung von photometrischen Übereinstimmungen. Anschließend wird der Gesichtsausdruck von der Quelle durch einen Deformationstransfer auf das Zielvideo übertragen. Für die finale Version wird das synthetische Gesicht noch an die Beleuchtungssituation des Hintergrundes angepasst und final in Echtzeit gerendert, so Thies et al. weiter. Probleme treten bei dieser Deepfake

Variante vor allem bei kleinen Verzerrungen des Körpers auf, die zu einer unnatürlichen Verzerrung des Zielvideos führen (Xinwei et al., 2021, S. 369). So wird die Echtzeit Puppet Mastery von Thies et al. (2018) mit einem Kamerainput als Datenquelle gespeist. Durch Mimik wird die „Marionette“ des Mannes (Zielperson) und seine Gesichtskomponenten gesteuert.

- **Face Synthesis**

Die Gesichtssynthese beschreibt das Zusammenstellen von fotorealistischen Bildern für das Gesicht einer Person. Das große Ziel dabei ist, realistisch wirkende Portraits von nicht existierenden Personen zu erschaffen (Masood et al., 2021, S. 18; Tong et al., 2020, S. 267). Als der beliebteste Ansatz dieser Variante gilt StyleGan. Diese 2019 entwickelte Erweiterung herkömmlicher GANs beinhaltet einen Stiltransfer und ermöglicht so einen Eingriff in die Steuerung der Synthese nicht existierender Gesichter, so die NVIDIA Forscher (Karras et al., 2019, S. 1). Diese folgenden Personen existieren nicht, sehen aber täuschend echt aus. Sie sind durch die Extraktion bestimmter Attribute aus den Quellen A und B im latenten Raum und der anschließenden Synthese zu jeweils neuen Bildern entstanden (Karras et al., 2019):



Abbildung 11 Gesichtssynthese mit StyleGan (Quelle: Karras et al., 2019, S. 4)

- **Facial Attribute Manipulation:**

Bei der Facial Attribute Manipulation werden Teile des Gesichts verändert. Diese selektive Auswahl von Veränderungen kann sich z.B. im Hinzufügen von Bärten oder Emotionen, Alterungsprozessen, Änderung von Haar- oder Hautfarbe oder einer Änderung des Blickwinkels äußern (Masood et al., 2021, S. 20). Die folgenden Deepfakes zeigen modifizierte Eigenschaften wie z.B. blondes Haar, Geschlecht, Alterung und blassere Hautfarbe.

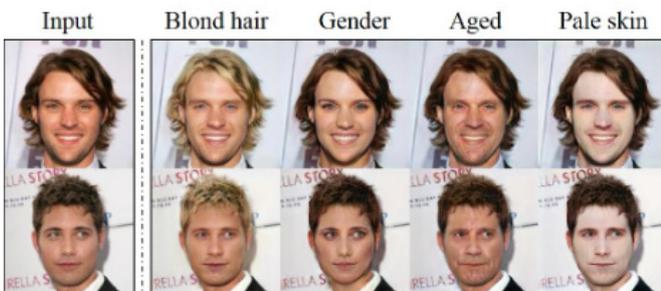


Abbildung 12 Facial Attribute Manipulation mit StarGan (Quelle: Choi et al., 2018, S. 1)

StarGan basiert auf der Technologie der StyleGan, indem es Attribute von einer Quelle auf das Ziel überträgt. Der Unterschied hierbei liegt in der Fähigkeit des Generators zusätzlich zu den Gesichtskomponenten Bereichsbezeichnungen zu erlernen (Choi et al., 2020, S. 2). Auf Grundlage dieser Fähigkeit kann ein Bild dann in bestimmten Bereichen modifiziert werden (z.B. blond hair aus dem oben genannten Beispiel). Qualitätsprobleme gibt es vor allem dann, wenn die Gesichter nicht frontal, sondern seitlich vorliegen (Xinwei et al., 2021, S. 368).

Masood et al. (2021, S. 12ff) diskutieren in ihrer Publikation bereits diverse Ansätze zu den hier aufgeführten Deepfake-Arten anderer Forscher und geben Hinweise zu

Limitationen. Auch in der Veröffentlichung von Mirsky & Lee (2020) werden Forschungsansätze ausgiebig diskutiert und Zusammenfassungen durch Tabellen zu den einzelnen Modellen und deren Spezifikationen gegeben, weshalb an dieser Stelle explizit darauf verzichtet und für Zusatzinformationen auf diese Literatur verwiesen wird.

3.5 Determinanten für erfolgreiche Deepfakes

Im folgenden Abschnitt werden Faktoren ermittelt, die für die Erstellung qualitativ hochwertiger und fotorealistic wirkender Deepfakes bis dato beachtet (und in Zukunft optimiert) werden müssen.

- **Menge der verfügbaren Daten**

Dadurch dass die Deep Learning Algorithmen mit einer großen Anzahl an Trainingsdaten für qualitativ hochwertige Ergebnisse gespeist werden müssen, stellt die Notwendigkeit von ausreichend Material des Objekts einen Knackpunkt in der Produktion von Deepfakes dar. Daher liegt ein Fokus in der Forschung in der Reduktion von notwendigem Material für das Training mit zeitgleicher Verhinderung von Qualitätsverlusten durch wenig Trainingsmaterial (Mirsky & Lee, 2020, S. 9).

- **Occlusions**

Werden Gesichtsteile durch Haare, Hände oder Brillen verdeckt, spricht man von sogenannten Occlusions. Erste Ansätze testen die Behebung der Probleme durch Segmentierung und gezieltes In-Painting, also Übermalen der verdeckten Stellen

(Masood et al., 2021, S. 33; Mirsky & Lee, 2020, S. 9).

- **Artefakte**

Deepfake-Netzwerke interpretieren die Videos nicht als Gesamtheit, sondern Frame für Frame, was zu Flimmern oder Zittern führt. Forscher arbeiten an Lösungen durch die Implementierung von Kohärenzverlusten oder die Nutzung von Recurrent Neural Networks (RNNs) (Masood et al., 2021, S. 33; Mirsky & Lee, 2020, S. 9). RNNs sind besonders für die Verarbeitung von Zeitreihendaten und anderen sequentiellen Daten geeignet (DiPietro & Hager, 2020) und kommen in der Deepfake-Technologie vor allem bei Audiodaten zum Einsatz (Mirsky & Lee, 2020, S. 9).

- **Seitenansicht und großer Kameraabstand**

Bei Frontalaufnahmen werden bereits gute Ergebnisse produziert, problematischer ist Material von der Seite oder bei dem die Entfernung des Gesichts groß ist. Beide Aspekte verursachen Artefakte oder eine fehlerhafte Gesichtssynthese, die es für gute Ergebnisse zu vermeiden gilt (Masood et al., 2021, S. 32).

Forscher legen außerdem nahe, dass Deepfakes durch die Optimierung von Determinanten wie z.B. Gesichtsausdruck, Lichtverhältnissen und Emotionen überzeugender wirken können (Albahar & Almalki, 2019, S. 3245). Zusammenfassend lässt sich festhalten und eine Brücke zur Glaubwürdigkeit in Videospielen schlagen: Deepfakes sollten so realistisch wie möglich erzeugt werden, um Glaubwürdigkeit bei Konsumenten zu erzeugen (Tencé et al., 2010, S. 2).

3.6 Aktuelle Open Source Tools

Im kommenden Kapitel werden Open Source Anwendungen vorgestellt, mit denen visuelle Deepfakes generiert werden können. Der Fokus liegt dabei auf der Erstellung von Face Swap Deepfakes, da diese die beliebteste Art von Deepfakes darstellen, die im Internet publiziert werden (Agarwal et al., 2019; Pu et al., 2021, S. 982) und Gegenstand für die selbst generierten Deepfakes im praktischen Teil dieser Arbeit ist.

In der Literatur werden die folgenden zwei Anwendungen häufig als die populärsten Tools zur Erstellung von Face Swap Deepfakes genannt: *FaceSwap* und *DeepFaceLab* (Bursic et al., 2021, S. 6090f; Khalil & Maged, 2021, S. 1f; Kötke, 2021, S. 2; Nguyen et al., 2022, S. 4; Pu et al., 2021, S. 982f; Tong et al., 2020, S. 268f; Xinwei et al., 2021, S. 368; T. Zhang et al., 2020, S. 69).

FaceSwap ist eine kostenlose Open Source Multiplattform für Deepfake Software, die 2017 entwickelt wurde. Die Software basiert auf Google's KI-Framework Tensorflow, der Deep Learning Bibliothek Keras und ist in Python gecodet. Es läuft auf den Betriebssystemen Windows, macOS und Linux (*FaceSwap*, 2022) und benötigt für optimale Performance eine moderne GPU mit CUDA Unterstützung (GitHub *FaceSwap*, 2022). Die Software arbeitet nach dem Autoencoder Prinzip, bei dem je ein Autoencoder für die Quelle und einer für das Zielvideo trainiert werden, um Gesichter zu erkennen und neue Gesichter aus einer latenten Darstellung rekonstruieren zu können. Durch die

Benutzung eines gemeinsamen Encoders wird im gemeinsamen latenten Raum die Gesichtsidentität gelöst, damit das Zielgesicht in diesem gemeinsamen Encoder mit dem Decoder des Quellvideos vertauscht wird (Pu et al., 2021, S. 982f). Laut Bursic et al. (2021, S. 6091) ist FaceSwap das vollständigste Tool, da es neben vieler Modelle anderer Tools auch mehrere hochauflösende Modelle zur Auswahl stellt. Diese Anwendung bietet zwei Wege es zu nutzen: über Kommandozeile oder GUI.

DeepFaceLab (DFL) betitelt sich selbst als „*the leading software for creating deepfakes*“ (GitHub DeepFaceLab, 2022) und gibt an, dass mehr als 95% aller Deepfake Videos mit DeepFaceLab produziert wurden. Als Abspaltung und Verbesserung von FaceSwap liegt bei dem 2018 entwickelten DFL dieselbe Funktionsweise des Autoencoder Prinzips zugrunde (Pu et al., 2021, S. 983). Das funktionsreiche und konfigurierbare Tool nutzt spezielle neuronale Netze (Multi-Task Cascaded Convolutional Networks, kurz: MTCCN), die drei verschiedene Phasen der Erkennung nutzen, um Gesichtsidentitäten (linkes Auge, rechtes Auge, Nase, linke Munddecke, rechte Munddecke) zu identifizieren (Bursic et al., 2021, S. 6091; K. Zhang et al., 2016). Mit DFL können Probleme wie Occlusion, schlechte Lichtverhältnisse, starke Schwankungen der Mimik und seitliche Aufnahmen des Gesichts besser gelöst werden (Xinwei et al., 2021, S. 368f). Als Systemvoraussetzungen wird die Notwendigkeit einer leistungsstarken Hardware mit adäquater Kühlung aufgrund der teilweise langen Trainingsprozesse (Stunden bis Tage) genannt (TMBDF, 2020).

DFL ist über die Kommandozeile oder durch das Anwählen von Skripten zu steuern und ggf. durch eigens programmierte Module zu erweitern (Perov et al., 2021, S. 3). Mit DFL ist Face Swapping auch in Kinoqualität möglich, so Perov et al. weiter. DFL erfährt große Unterstützung von der Community, da sie dabei helfen, die Software kontinuierliche weiterzuentwickeln und gegenseitige Unterstützung in Form von Forumsbeiträgen (TMBDF, 2020) und Tutorials mit weit über 1 Millionen Klicks (Cinecom.net, 2019) bieten. Für die selbst generierten Deepfakes dieser Masterarbeit wird aufgrund der Marktführerschaft und der vielfältigen Konfigurationsmöglichkeiten innerhalb des Tools die Software DeepFaceLab genutzt.

Weitere beliebte Tools sind die beliebteste jedoch aus Datenschutzgründen (Peteranderl, 2019) umstrittene chinesische App ZAO oder *Reface*, mit der man sein Gesicht in Videos, GIFs oder Bilder deepfaken kann. Auf die Benutzung von mobile Apps wird aufgrund der eingeschränkten Modifikations- und Steuerungsmöglichkeiten an dieser Stelle verzichtet.

3.7 Reale Deepfakes und deren Auswirkungen

Für ihre Forschung sammelten Pu et al. (2021) den bislang größten Datensatz an Deepfake-Videos von YouTube und Bilibili mit 1.896 Deepfakes. Die Dunkelziffer ist vermutlich viel höher. Dennoch sieht man parallel zum Anstieg der Deepfake-Literatur ein wachsendes Interesse an dieser neuen Technologie.

Jahr	2015	2016	2017	2018	2019	2020	2021	2022*
Publikationen	1	0	1	71	374	1.367	1.741	1.184

* Stand:
13.07.2022

Tabelle 1 Zahl der Publikationen zum Stichwort „deepfake“ von 2015 bis 2022 generiert mit: <https://app.dimensions.ai>

Die bisher populärsten Deepfakes können in verschiedene Kategorien eingeteilt werden. Eine dieser Kategorien ist Pornografie. Deepfakes werden für das Erstellen und Teilen von pornografischem Material mit den Gesichtern von berühmten Persönlichkeiten auf unzähligen einschlägigen Plattformen veröffentlicht. Angefangen mit dem ersten Deepfake des Reddit Nutzers werden z.B. die Gesichter von Scarlett Johansson, Angelina Jolie oder Kendall Jenner etc. bis heute böswillig und ohne erteiltes Einverständnis missbraucht (Harwell, 2018). Problematisch ist dabei der Fakt, dass Missbrauch des eigenen Bildmaterials für pornografische Zwecke in der Theorie jede Person treffen kann, wenn ausreichend Bild- oder Videomaterial z.B. von sozialen Kanälen zur Verfügung steht. Zur Erstellung von und dem Umgang mit pornografischen Deepfakes wird sich an dieser Stelle ausdrücklich distanziert.

In einer der berühmtesten Deepfakes von Suwajanakorn et al. (2017) wird das Gesicht des früheren US-Präsidenten Barack

Obama synthetisiert. Dabei wurde das Gesicht – insbesondere die Mundpartie – auf eine Audiodatei abgestimmt und am Ende sieht so aus, als würde der echte Obama in dem Video erzählen. Zum Zeitpunkt der Veröffentlichung war dies ein bahnbrechender Erfolg (Suwajanakorn et al., 2017).

Ein Jahr später wurden wieder Obama fremde Worte in den Mund gelegt. So nannte er den damaligen US-Präsidenten Donald Trump „[...] a total and complete dipshit [...]“ (BuzzFeedVideo, 2018) und warnte vor den Auswirkungen der Deepfake-Technologie. Dieses Video hat bis heute fast 9 Millionen Aufrufe auf YouTube und bestätigt die Macht und Anziehungskraft von Deepfakes. Doch gleichzeitig kann durch manipulierte Aussagen die Glaubwürdigkeit leiden oder noch gefährlicher: das Vertrauen in eine Person. Insbesondere für Politiker stellt diese Desinformation eine unkalkulierbare Bedrohung dar.



Abbildung 13 links: (Suwajanakorn et al., 2017); rechts: (BuzzFeedVideo, 2018)

Chesney & Citron (2019) sprechen in der Hinsicht von einem neuen Zeitalter des Desinformationskrieges. Aktuell kursieren Videos und Bilder aus dem Ukrainekrieg, die aus dem Zusammenhang gerissen wurden, um mögliche Angriffe oder Opfer zu zeigen (Sardarizadeh & Robinson, 2022). Viel schwerwiegender wären dabei die Konsequenzen des Einsatzes von Deepfakes, um Kriegspropaganda zu betreiben. So warnen Militärexperten auch vor Fälschungen von Satellitenbildern mit z.B. nicht existenten Brücken oder Militärbasen, die durch das Training von Deepfakesoftware generiert werden könnten (Fish, 2019; Tucker, 2019).

Auch im Wirtschaftssektor können Deepfakes großen Schaden anrichten, so wurde ein CEO beispielsweise um 243.000 Dollar betrogen, indem Betrüger seine Stimme synthetisierten (Stupp, 2019).

Für große Aufregung in der Kunstszene sorgte auch der Deepfake von Salvador Dalí im Dali Museum in Florida, bei dem ein Deepfake-Avatar in Gestalt von Dalí selbst erstellt wurde. *„Visitors to the Museum will soon have the opportunity to learn more about Dalí’s life and work from the person who knew him best: the artist himself.“* (The Dali Museum, 2019), so das Museum

über dieses bis dato einmalige Feature. Als interaktive Installation bereichert der Deepfake das Museum und seine Besucher, indem es spielerisch Interesse an Kunst und Geschichte weckt.

So macht künstliche Intelligenz in Form von Deepfakes es auch möglich, Donald Trump eine spezielle Version der Geschichte von Rudolf mit der roten Nase in einem Weihnachtspullover vorlesen zu lassen (Sassy Justice, 2020).

Insgesamt lassen sich die Intentionen der berühmtesten Deepfakes derzeit in vier Kategorien einteilen: Pornografie, böswillige Manipulation, Spaß, sowie Kommerz. Doch wie kann man den problematischen Risiken von Deepfakes entgegenwirken?



Abbildung 14 links: Deepfake-Avatar (The Dali Museum (a), 2019); rechts: Deepfake-Satire (Sassy Justice, 2020)

3.8 Erkennungsansätze

Nachdem im vorangestellten Kapitel sowohl die positiven als auch die momentan stärker überwiegenden negativen Aspekte der Deepfake-Technologie erläutert wurden, geht es in diesem Kapitel darum, Ansätze zur Erkennung von Deepfakes vorzustellen. Da der Fokus dieser Arbeit auf der Erstellung von Deepfakes liegt, werden die Erkennungsmechanismen und jeweilige Beispiele nur kurz erläutert.

Bislang gibt es kein allgemeingültiges Tool zur 100-prozentigen Erkennung von Deepfakes. Die bisherigen Ansätze zur Erkennung von Deepfakes lassen sich in artefaktbasierte und indirekte Erkennungsansätze einteilen.

Deepfake-Erkennungsansätze

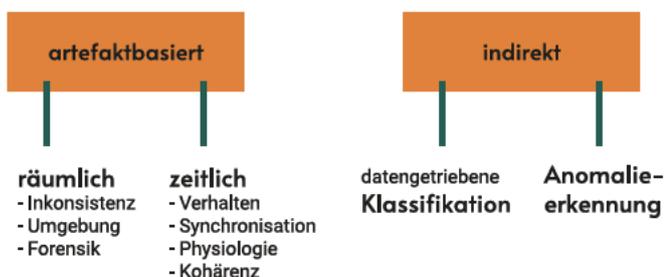


Abbildung 15 Einteilung der Deepfake-Erkennungsmechanismen (In Anlehnung an Masood et al., 2021, S. 31f und Mirsky und Lee, 2020, S. 26ff)

Methoden der Manipulationserkennung für statische Bilder lassen sich nicht pauschal auf Videomaterial anwenden, da diese auf die Erkennung von Einzelframes ausgelegt sind und die zeitlichen Merkmale von Videos nicht berücksichtigen. Ein weiterer Grund, wieso für Videoerkennung andere Methoden nötig sind liegt darin, weil sich die Qualität der Bilder nach der

Videokomprimierung rapide verschlechtert (Afchar et al., 2018, S. 1). Für erfolgreiche Erkennungsmechanismen werden robuste, skalierbare generische Algorithmen und hochwertige Datensätze für das Training benötigt (Weerawardana & Fernando, 2021, S. 80).

Zu der in der Literatur häufig angewandten Einteilung (Masood et al., 2021; Mirsky & Lee, 2020) der Erkennungsmechanismen zählen sowohl **räumliche** als auch **zeitliche** artefaktbasierte Aspekte. Erkennen lassen sich räumliche Artefakte vor allem durch:

- **Inkonsistenz:** Sichtbare visuelle Artefakte z.B. ungleichmäßige Kopfbewegungen und Orientierungspunkte (Masood et al., 2021, S. 31)

Yang et al. (2019) entwickelten dafür aufgrund der Erkenntnis, dass beim Einfügen des Fake-Gesichts in das Video kleine Fehler entstehen eine Methode, bei dem die 3D-Gesichtspose geschätzt wird. So sind Abweichungen Indizien für das Vorliegen eines Deepfakes.

- **Umgebung:** unnormale Erscheinungen im Bild z.B. Belichtung oder andere Details (Masood et al., 2021, S. 31) Der Erkennungsansatz von Y. Li & Lyu (2019) basiert zum Beispiel auf der Grundlage, dass damalige Deepfakes nur in niedriger Auflösung erzeugt werden konnten und verzerrt werden müssen, um auf das Zielgesicht zu passen (z.B. durch Skalieren oder Rotieren). Die dadurch erzeugten Artefakte können dann mit Hilfe eines CNNs entdeckt werden. (vgl. dazu auch „Deep Classifier“ in Nguyen et al., 2022, S. 9; Rahman et al., 2022, S. 18)

- **Forensik:** Hinterlassen von Merkmalen oder Mustern durch die Erstellung (Masood et al., 2021, S. 31)

Diese Merkmale sind zum Beispiel „Fingerabdrücke“ die von GANs, die bei der Synthese von Deepfakes hinterlassen wurden und im Nachgang identifiziert werden können (Yu et al., 2019). Koopman et al. (2018) verfolgen mit ihrer Methode die Analyse der PRNU (Photo Response Non Uniformity), also das einzigartige Sensorrauschen einer Kamera zur Erkennung von fremden Inhalten. (vgl. dazu auch „Shallow Classifier“ in Nguyen et al., 2022, S. 10f; Rahman et al., 2022, S. 19)

Zur zweiten Gruppe, den **zeitlichen** Ansätzen zählen:

- **Verhalten:** Unnatürliche Gesten und Gesichtsausdrücke (Masood et al., 2021, S. 31)

Agarwal et al. (2019) haben bewiesen, damit Zusammenhänge zwischen Gesichtsausdrücken und Kopfbewegungen dafür genutzt werden können, um Personen zu unterscheiden und ebenso manipulierte Videos zu identifizieren. Diese Zusammenhänge sind nicht sichtbar, werden jedoch bei der Synthese von Deepfakes häufig nicht beachtet und sind daher für eine Identifikation von Fakes nützlich (Agarwal, El-Gaaly, et al., 2020).

- **Synchronisation:** Unstimmigkeiten zwischen benachbarten Frames des Videos (Masood et al., 2021, S. 31)

Ein Ansatz in diesem Bereich konzentriert sich dabei auf das Erkennen von Unstimmigkeiten zwischen der visuellen Mundform und dem tatsächlich gesprochenen Laut, insbesondere bei Buchstaben für die der Mund geschlossen

sein muss (z.B. M, B, P) (Agarwal, Farid, et al., 2020).

- **Physiologie:** Mangel oder Fehler in biologischen Vorgängen wie z.B. Augenzwinkern oder dem Herzschlag (Masood et al., 2021, S. 31)

Zum Beispiel werden Deepfake-Algorithmen meist nur auf Bilddaten mit offenen Augen trainiert und sind somit schlecht in der Synthese von Blinzeln. Sowohl das Fehlen von Augenblinzeln als auch zu häufiges Blinzeln sind Anzeichen für Manipulation (Y. Li et al., 2018). Andere Forscher zählen z.B. auch das Fehlen von Details und Reflexionen bei Zähnen zu Anhaltspunkten für Deepfakes (Matern et al., 2019).

- **Kohärenz:** Fehlen von Optical Flow Fields und Auftreten von Artefakten wie z.B. Flimmern und Zittern zwischen den Frames (Masood et al., 2021, S. 31)

Der Erkennungsansatz von Amerini et al. (2019) nutzt optische Flussfelder, also das Maß zur Repräsentation von Bewegungsinformationen in Bewegtbildern, um Ungleichmäßigkeiten zwischen den einzelnen Frames zu identifizieren. Fei et al. (2021) schlagen wiederum eine Methode vor, die zunächst die Amplitude der Gesichtsbewegungen in dem potentiellen Deepfakevideo erhöht. Da Fakevideos stärkeres Flimmern und Verzerrungen als Originale aufweisen, können CNNs die Fälschungen erkennen.

Die indirekten Erkennungsansätze umfassen Deep Learning Methoden, wie Klassifikation und Anomalieerkennung.

- **Klassifikation:** CNNs, die auf datengetriebenen Modellen basieren (Masood et al., 2021, S. 31)

Der Ansatz von Hsu et al. (2020) trainiert z.B. zur Identifikation von Fake oder Nicht-Fake ein siamesisches neuronales Netz und nutzt Kontrastverluste als Erkennungsmerkmal.

- **Anomalieerkennung:** Ausreißer identifizieren (Masood et al., 2021, S. 31)
Mittal et al. (2020) analysieren Ähnlichkeiten bzw. Unähnlichkeiten von Gesicht, Sprache und Emotionen, um Anomalien zu erkennen.

Die verschiedenen Erkennungsmethoden für die Authentifizierung von Videomaterial geben Aufschluss über die Komplexität der Deepfake-Erkennung und mögliche Optimierungsansätze für zukünftige Forschung. Insbesondere für die Deepfake-Produktion werden Aspekte geliefert, auf die für authentische, fotoreale Ergebnisse besonders geachtet muss und die es in den Deepfake-Tools wie z.B. DeepFaceLab zu optimieren gilt. Im Umkehrschluss bedeutet dies aber auch, damit Deepfakes nicht mehr von echten Videos zu unterscheiden sind, wenn die zuvor genannten Aspekte gelöst werden.

4. Konzept der Hyperpersonalisierung von Werbebotschaften mit Hilfe von Deepfakes

Der folgende Teil dieser Arbeit beschäftigt sich mit der Ausarbeitung des Deepfake-Marketing-Konzepts, also der Hyperpersonalisierung von Werbemaßnahmen durch den Einsatz der Deepfake-Technologie in der Werbebranche. Gleichzeitig werden Hypothesen zur Überprüfung der Forschungsfrage ausgestellt. Im Anschluss an das Konzept wird die Produktion der eigenen Deepfakes begleitet und im Anschluss die Online-Umfrage zu den Effekten bei Rezipienten durchgeführt.

4.1 Hyperpersonalisierung durch KI-Einsatz

Die Hyperpersonalisierung kann als Weiterentwicklung der zu Beginn dieser Arbeit erläuterten Personalisierung im Werbesektor angesehen werden. Hyperpersonalisierung beschreibt die echtzeitgetriebene, kontextrelevante Nutzung von Browsing- und Verhaltensdaten von Nutzern für individuelle Kundenkommunikation (Capgemini Consulting & ESSEC, 2016, S. 5; Gouthier & Kern, 2021, S. 129). Diese hochgradig personalisierte Form der Kundenansprache ist derzeit die höchste Entwicklungsstufe im Bereich der Personalisierung (Gouthier & Kern, 2021, S. 133) und wird durch den Einsatz von KI für die Echtzeitkommunikation weiter ausgebaut (dotCMS, 2018).

Die in Kapitel 2.4 angesprochene Customer Journey und das wichtige Zusammenspiel der Marketingmaßnahmen über aller Touchpoints hinweg ist die Voraussetzung für zielführende Kundenerlebnisse. Mit

dem Einsatz von Big Data Analysen innerhalb der Hyperpersonalisierung werden so neue Ansatzpunkte zur Optimierung geliefert (Capgemini Consulting & ESSEC, 2016). So gelten die wichtigsten drei V's von Big Data als Wegbereiter für die Hyperpersonalisierung. Erst die Verfügbarkeit der großen Datenmenge (**Volume**) ermöglicht das Aufspüren von versteckten statistischen Zusammenhängen. Gleichzeitig kann dies durch z.B. KI-gestützte Anwendungen in einer Geschwindigkeit (**Velocity**) erreicht werden, die manuell nie erreicht werden würde. Und auch die Vielfalt (**Variety**) der zur Verfügung stehenden Daten ermöglicht neue Insights in Gebiete, die bisher nicht oder schlecht für Datenauswertungen genutzt wurden (Capgemini Consulting & ESSEC, 2016, S. 8f; Gouthier & Kern, 2021, S. 140).

Sodhi (2018) spricht davon, dass Marketer sich einen 360°-Blick über den Kunden verschaffen müssen. Gemeint ist damit die Sicherstellung eines ganzheitlichen, konsistenten und stets aktuellen Verständnisses der Kundenbedürfnisse (Gouthier & Kern, 2021, S. 137). Erreicht wird dies durch Bündelung und ständige Aktualisierung von Kundendaten aus CRM, sozialen Medien und Internetaktivitäten, so die Forscher weiter.

Das große Ziel der Hyperpersonalisierung ähnelt dem der herkömmlichen Personalisierung: Die Kundenansprache muss gezielt und initiativ durch relevante Informationen und Angeboten erfolgen, um einen Mehrwert zu bieten (Shukla & Nigam, 2018, S. 132). Persönliche Daten werden dabei genutzt, um jedem einzelnen Kunden passgenaue Informationen zu

liefern (Kerner, 2021). Die große Hürde ist dabei, alle Informationen über eine Person vollständig zu erfassen, um damit anschließend sinnvoll zu agieren, so Kerner weiter. Problematisch dabei sei allerdings die Bündelung von Daten in verschiedenen Systemen und daraus resultierende Informationslücken. Dabei ergaben die Ergebnisse einer Studie, dass 91% der Konsumenten eher bei Marken kaufen, „*die sie erkennen, sich an sie erinnern und ihnen relevante Angebote und Empfehlungen machen*“ (Accenture, 2018, S. 3), was klar für den Ausbau der Personalisierungsangebote stimmt. Melzer (2022) zählt zu den Vorteilen der Hyperpersonalisierung für Unternehmen vor allem die langfristige Reduktion von Marketingkosten und gleichzeitige Maximierung des Umsatzes. Für Kunden wird laut Melzer das Erlebnis mit der Marke gesteigert und die Markentreue erhöht.

Dort setzen Deepfakes an: Mit dem Einsatz von Deepfakes können Werbeinhalte dynamisch an die jeweiligen Präferenzen der Nutzer auf audio-visueller Ebene in Echtzeit adaptiert werden. Kietzmann et al. (2021, S. 5) sprechen in dem Zusammenhang von „*the ultimate in personalization*“. Das Nonplusultra der Personalisierung könnte somit durch Deepfakes erreicht werden.

So spezialisieren sich bereits erste Medienunternehmen auf die Generierung von hyperrealem, synthetischen Content. Dabei geht es nicht nur um die Produktion solcher Medien, sondern auch Lösungen für dadurch verursachte Probleme zu finden (syntheticfutures.org, 2022). Der VFX-Profi Chris Ume ist in den sozialen Medien für seine Tom Cruise Deepfakes bekannt (deptomcruise, 2022). Für die Deepfakes

arbeitet er mit einem Double zusammen, das ohnehin schon Ähnlichkeit mit Tom Cruise besitzt und kann so täuschend echte Ergebnisse erzielen die tausendfach gelikt und kommentiert werden. Insgesamt 3 Millionen Follower zählt die komplett digital erzeugte Persönlichkeit @lilmiquela auf Instagram. Der als junge Frau dargestellte Avatar eines amerikanischen Start-Ups fungiert als Influencerin und bewirbt erfolgreich Markenprodukte sowie eigene Musik (Daly, 2019). Die Beliebtheit synthetischer Medien scheint in den sozialen Medien bereits groß zu sein.

4.2 Zielgruppendefinition

Doch welche Menschen können durch hyperpersonalisierte Deepfake-Werbung im Sinne eines positiven Werbeerfolgs erreicht werden?

Die repräsentative Studie von PWC (2019) zum Thema personalisierte Werbung zeigt deutlich, dass Personalisierungen wahrgenommen werden. Besonders häufig ist den unter 30-Jährigen personalisierte Werbung aufgefallen (51%) (S. 20). Doch auch bei den 30 bis 50-Jährigen fällt mehr als 4 von 10 Personen Personalisierung auf. Bei der Geschlechterunterteilung zeigt sich, dass insbesondere Männer empfänglicher für personalisierte Werbung sind (S. 22). Trotzdem fällt auf, dass sich nur 25% der Befragten zu Käufen durch personalisierte Inhalte verleiten lassen (S. 23). Dies zeigt noch einmal deutlich, dass die Personalisierung von Werbeeinblendungen optimiert werden muss. „*Personalisierte Online-Werbung sticht ins Auge*“ (PWC (a), 2019), fasst Werner Ballhaus von PWC zusammen. Doch dabei spielt vor allem das

Alter eine Rolle, so Ballhaus weiter: „Vor allem Ältere zeigen sich skeptisch, während die jüngeren Deutschen personalisierte Werbeeinblendungen häufiger als eine Art Service verstehen, der ihnen interessante Produkte vorschlägt“ (PWC (a), 2019), ergänzt er.

Auf die Hyperpersonalisierung übertragen bedeutet dies, dass jüngere Deutsche eher empfänglich dafür sind als ältere Generationen. Aufgrund der Ergebnisse der PWC-Studie wird dabei die Altersgrenze zwischen jung und alt für erfolgreiches Deepfake-Marketing bei 30 Jahren gezogen.

Betrachtet man zusätzlich die DIVSI-Internetmilieus, so befindet sich die potentielle Zielgruppe der Hyperpersonalisierung mit Deepfakes im Bereich der internetaffinen Nutzer (DIVSI, 2016). Die internetaffinen Nutzer lassen sich in vier Untergruppen (in der

folgenden Grafik gelb und grün markiert) einteilen. Dabei passen insbesondere die Netz-Enthusiasten charakterisiert durch ihre Vorliebe für soziale Medien in die Zielgruppe. Doch auch die souveränen Realisten könnten für Hyperpersonalisierung mit Deepfakes interessant sein. Menschen, die in dieses Milieu eingeordnet werden, wollen trotz ihrer intensiven Internetnutzung durch das Vordringen des Internet in den privaten Lebensbereich nicht die Selbstbestimmtheit aufgeben. Die anderen beiden Internet-Milieus sind durch ihre weniger (Unbekümmerte Hedonisten) und ausgeprägte (Effizienzorientierte Performer) Sicherheitsbedenken gekennzeichnet, weshalb sie nur eingeschränkt für Personalisierung durch Deepfakes geeignet sind. Die Übersicht zeigt die einzelnen DIVSI-Internetmilieus mit ihrem Grad der Eignung für die Hyperpersonalisierung mit Deepfakes.

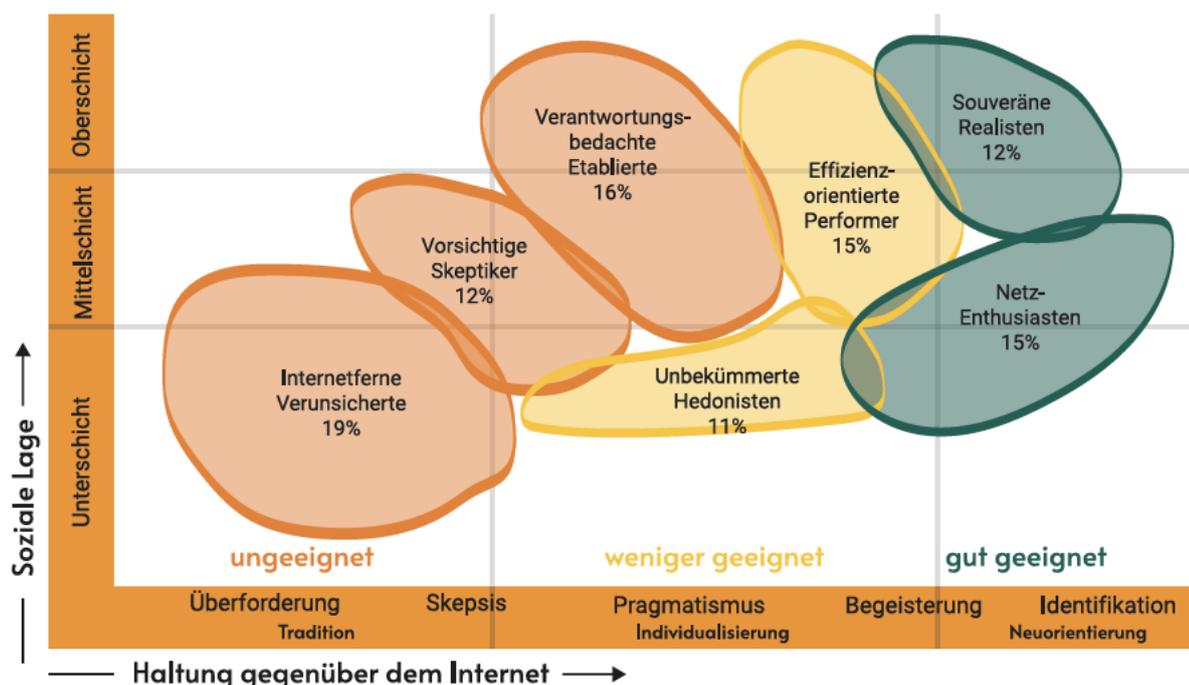


Abbildung 16 DIVSI-Internetmilieus mit Grad der Eignung für Hyperpersonalisierung mit Deepfakes (In Anlehnung an DIVSI, 2016)

Zieht man zusätzlich noch die Internet-Verbrauchertypen von Deutschland sicher im Netz (2021, S. 16ff) in Betracht, so sind insbesondere die antreibenden Nutzer für Deepfake-Marketing interessant. Die Antreibenden Nutzer zeichnet ein hohes Sicherheitswissen und gleichzeitige Begeisterung beim Ausprobieren neuer digitaler Anwendungen aus (S. 24f). Weniger geeignet sind die Bedachtsamen Nutzer, aufgrund ihrer prüfend-kritischen Art neuer Angebote, sowie Außenstehende Nutzer (S. 18f), die von neuen digitalen Diensten oft überfordert sind. Die Fatalistischen (S. 20f) und Gutgläubigen Nutzer (S. 22f) könnten aufgrund ihrer unzureichenden Kenntnis über Sicherheitsmaßnahmen und deren Anwendung zum Selbstschutz problematisch für Deepfakes sein. In der Übersicht der Anteile der Nutzergruppen zeigt sich, damit nur knapp 2 von 10 Personen für Hyperpersonalisierung durch Deepfakes geeignet sind.

Aufgrund der Annahmen werden im Anschluss zwei fiktive Nutzer der Zielgruppe gezeigt. Diese Personas stehen stellvertretend für die mögliche Zielgruppe von Hyperpersonalisierung mit Deepfakes.

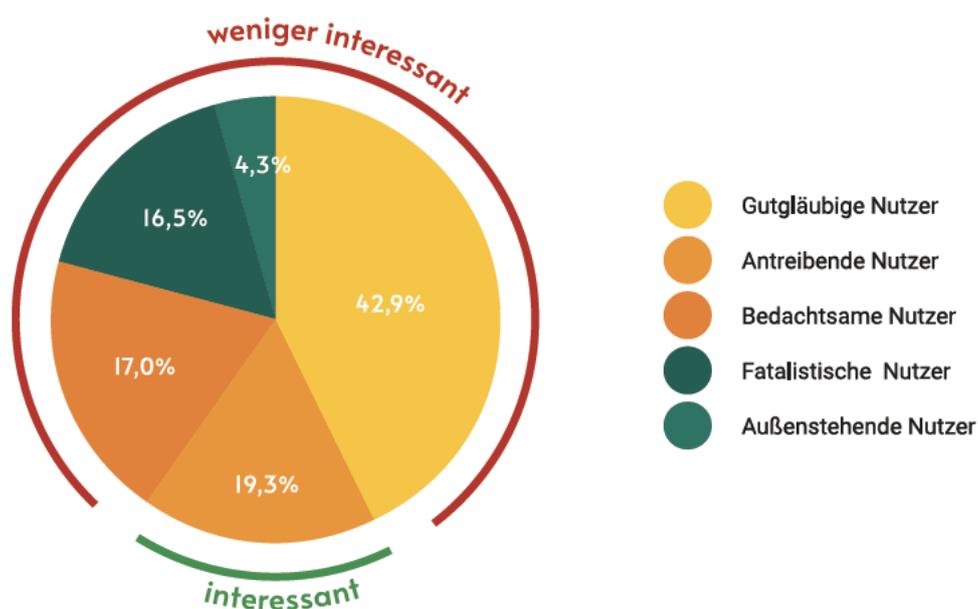


Abbildung 17 Anteile der Nutzertypen aus Deutschland sicher im Netz (2021, S. 17)

Tom, 30 Jahre

in einer Beziehung, wohnt in Hamburg,
2-Zimmer-Wohnung

- Account Manager bei einem Sportartikelhersteller
- **Lieblingsendgeräte:** Smartphone und Spielekonsole
- **Umgang mit dem Internet:** passiver Nutzer soziale Medien (um mit Freunden in Kontakt zu bleiben, kein aktives posten)
- **Gedanken zum Werbekonsum:** *„Auf der Arbeit und am PC nerven mich Werbeanzeigen, wenn ich konzentriert arbeiten möchte. In den sozialen Medien oder beim surfen in meiner Freizeit bin ich schon häufig durch Werbeanzeigen zufällig auf coole Produkte gestoßen, die zu meinem Lifestyle passen.“*
- **Gedanken zum Datenschutz:** *„Ich achte darauf, was ich im Internet mit wem teile. Bisher ist mir aber noch nichts passiert und ich denke, damit meine Schutzmaßnahmen und die der Internetseiten dazu beitragen. Außerdem gibt es in Europa unzählige Gesetze die meine Rechte schützen.“*
- **Gedanken zu neuen Medien:** *„Ich probiere gern Neues aus und sehe, ob es mich weiterbringt oder doch schlecht für mich ist. Probieren geht eben über studieren und am Ende zählt meine Erfahrung und das was ich damit anstellen kann. Ich bin mit Bandsalat aufgewachsen, hab mich mit zerkratzten CD's rum geärgert und jetzt kann ich mich theoretisch beim Zocken via VR-Brille mit meinen Freunden unterhalten. Was kommt als Nächstes?“*

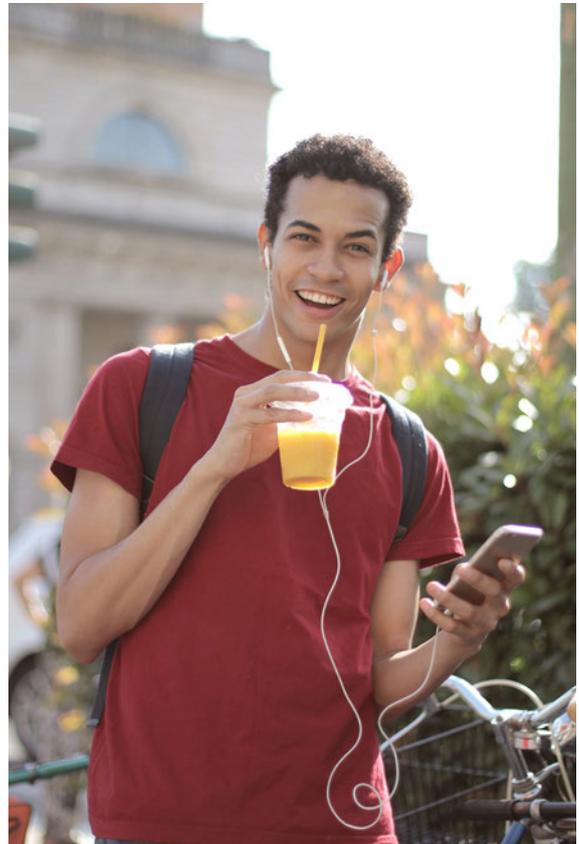


Abbildung 18 Photo by Andrea Piacquadio from Pexels

Luna, 18 Jahre



Abbildung 19 Photo by Ron Lach from Pexels

Single, wohnt in Hannover, 4er-WG im Studentenwohnheim

- Bachelorstudentin 1. Semester Hotelmanagement
- **Lieblingsendgeräte:** iPhone und iPad
- **Umgang mit dem Internet:** Content Creator in sozialen Medien mit großer Reichweite, liebt es ständig online zu sein und kann nicht ohne
- **Gedanken zum Werbekonsum:** *„Mein halbes Leben besteht aus Werbung. Seitdem ich mich in den sozialen Medien rum treibe sehe ich überall Menschen die Produkte bewerben. (Ich bin ja auch einer davon!) Ich finde das cool, weil man dadurch selbst viele Vorteile erhält und andere Menschen auf kreative und humorvolle Art inspiriert und so auch deren Leben bereichert. Ich schaue (und kaufe) gern das, was andere empfehlen!“*
- **Gedanken zum Datenschutz:** *„Wer heute im Internet unterwegs ist hinterlässt überall seine Daten. Für mich ist das voll ok und was will man dagegen auch machen? Einschränken möchte ich mich jedenfalls nicht. Es gibt ja fast überall die Möglichkeit Beiträge o.Ä. den Betreibern einer Seite zu melden.“*
- **Gedanken zu neuen Medien:** *„Als Digital Native bin ich für jeden Trend zu begeistern! Wenn etwas cool ist, teile ich es gern mit meinen Freunden und Followern.“*

4.3 Abwägen der Vor- und Nachteile für Kunden und Marketer

Das folgende Kapitel bringt die Vor- und Nachteile der Hyperpersonalisierung durch die Integration von Deepfakes in die Werbung zusammen. Dabei wird zwischen der Unternehmens- und der Kundenseite unterschieden. Begonnen mit den Vorteilen, werden im Anschluss die Nachteile des Konzepts erläutert.

- **Vorteile für Marketer bzw. Unternehmen**

Der wohl ausschlaggebendste Aspekt für Unternehmen ist die Kosteneinsparung beim Einsatz von Deepfakes. Models und Schauspieler müssen nicht physisch am Set anwesend sein. Die Gesichter der Protagonisten könnten vorab digitalisiert werden und in einer Gesichtsdatenbank abgelegt werden. Ähnliches Vorgehen wird bereits von Hour One (2022) für die Aufnahme neuer Charaktere als Präsentatoren praktiziert. Auch das bekannte Modeunternehmen Zalando rollte 2018 eine Kampagne mit Cara Delevingne aus, bei der über 60.000 Versionen des Spots (zugeschnitten auf verschiedene Orte) mit der Deepfake-Technologie produziert wurden (Chitrakorn, 2021; Langhart, 2021). Durch die Ressourceneinsparung können Teammitglieder andere Aufgaben übernehmen, was zu einer gesteigerten Produktivität des Teams führt (Mari, 2019, S. 6).

Unternehmen können dank künstlicher Intelligenz und Deepfakes Big Data effizient nutzen. L’Oreal Paris nutzt Daten wie Präferenzen und Interessen von Kaufinteressenten, die von Google durch

KI-gestützte Verfahren erhoben wurden, um ihre Werbevideos zu personalisieren. Zwölf Versionen wurden dabei, individuell zugeschnitten auf die Vorlieben der Zielgruppen produziert. Der Erfolg dieser Vorstufe der Hyperpersonalisierung zeigt sich prompt: Mit der Kampagne konnte ein Anstieg des Produktinteresses um 144% und bei der Kaufabsicht um 30% erzielt werden (Mari, 2019, S. 11; Think with Google, 2018). Deepfakes können hier genutzt werden, um Protagonisten, Sprache oder Text zu synthetisieren, um auf noch spezifischere Präferenzen reagieren zu können.

Neben dem Ausbau der effizienten Datennutzung nimmt die Integration von Deepfakes in das Werbekonzept auch auf den Bereich der Kreativität Einfluss. Insbesondere für die Filmbranche stellen Deepfakes deshalb großes Potential aufgrund der Erweiterung der Möglichkeiten für Spezialeffekte dar. So können bereits verstorbene Personen zum Leben erweckt (The Dali Museum, 2019), das Äußerliche von Personen stark verjüngt oder gealtert (Whittaker et al., 2020) oder durch visuelle Effekte schwer umsetzbare Szenen produziert werden, um z.B. Fortsetzungen von Filmen oder Serien mit den „echten“ Personen aus früheren Zeiten zu realisieren oder Filmcharaktere für Commercials oder anderweitige externe Produktionen einzusetzen. Und auch für noch lebende Personen könnte der Einsatz von Deepfakes Vorteile haben: Die „Malaria Must Die“ Kampagne mit David Beckham wurde durch die Deepfake-Technologie in neun Sprachen realisiert. Obwohl David Beckham diese neun Sprachen nicht beherrscht, sehen die Mundbewegungen aus, als würde

er sie selbst sprechen (Butcher, 2019). Die Deepfake-Technologie ermöglicht es Unternehmen auch, den Kunden die Möglichkeit zu geben das eigene Gesicht in den echten Filmtrailer zu integrieren (Moon, 2021), um damit z.B. im Bekanntenkreis passiv für den Film zu werben.

- **Vorteile für Kunden**

Kunden profitieren beim Einsatz von Deepfakes für die Hyperpersonalisierung von Werbebotschaften vor allem darin, dass sie genau auf sie zugeschnittene Werbeeinblendungen erhalten. So könnte durch Nutzerdatenanalyse ermittelt werden, welche Art von Protagonist der jeweilige Kunde bevorzugt. Unpassende Merkmale könnten so aktiv außer Acht gelassen. Der Kunde sieht am Ende z.B. keine Werbung mit Morgan Freeman, den er überhaupt nicht mag und bekommt dafür Werbung mit Mr. Bean, den er sehr gern sieht. So können Emotionen wie Ärger und Frust über nichtzutreffende Werbung vermieden werden und stattdessen positive Gefühle wie Freude und Bewunderung angeregt werden, welche im Idealfall zur Akzeptanz der Werbung und zum Kauf führen können.

Weiterhin fördert die Konfrontation von Deepfakes und synthetisch generierten Medien den Umgang mit neuen Medien. So könnte z.B. das Erkennen von Deepfakes nach den im vorangestellten Kapitel 3.8 Erkennungsansätze aufgelisteten Aspekten zur Reduktion der überschätzten eigenen Fähigkeiten bei der Deepfake-Erkennung (Köbis et al., 2021) beitragen und mögliche Bedenken verringern.

Insgesamt zeigt sich, dass sowohl auf Kunden- als auch auf Unternehmensseite,

mit Deepfakes einzigartige Produkt- und Markenerlebnisse geschaffen werden, die erstens klare Wettbewerbsvorteile schaffen und zum anderen das Erleben von Produkten und Marken aufregender und überraschender gestalten.

Neben den positiven Effekten müssen aber auch die Nachteile dieser neuen Technologie erörtert werden.

- **Probleme für Marketer bzw. Unternehmen**

Vorab wurden bereits die Voraussetzungen für die Umsetzung von Deepfakes kurz angerissen. Die Notwendigkeit leistungsstarker Hardware, sowie Geduld und fundiertes Know-How stellen insbesondere kleine Agenturen vor Probleme. Die Möglichkeit der Datenerhebung und die viel wichtigere angepasste Datennutzung kann momentan auch nicht so erfolgen, wie sie für eine Hyperpersonalisierung z.B. in Echtzeit nötig ist. Software und Tools dafür müssen erst konzipiert, mit bestehenden Anwendungen verbunden und anschließend in eine sinnvolle Deepfake-Marketing-Pipeline integriert werden. Wann dies möglich ist, bleibt unklar.

Für die Generierung von Deepfakes müssen Deepfake-Spezialisten geschult werden. So könnten sich z.B. VFX-Spezialisten mit Wissen über Programmierung und Deep Learning dafür qualifizieren.

Eine andere Problemquelle birgt die Rechtsgrundlage in Deutschland, denn dort herrscht Uneinigkeit. Gesetze zum Umgang mit Deepfakes fehlen. Soziale Netzwerke wie Facebook entfernen Deepfakes (Benrath & Theile, 2020) aufgrund eigener

Richtlinien, um ihre Nutzer vor Fake News zu schützen. In Kalifornien wurden anlässlich der US-Wahlen 2020 das Gesetz „AB 730“ (California Assembly Bill, 2019) erlassen, welches verbietet innerhalb 60 Tagen vor der Wahl manipulierte audio-visuelle Daten zu veröffentlichen. Dieses Anti-Deepfake-Gesetz wurde allerdings nur in politischem Zusammenhang erlassen und es bleibt unklar, ob weitere Gesetze (insbesondere in Deutschland) folgen. Die Unklarheit im rechtlichen Umgang mit Deepfakes stellt ein großes Hindernis für deutsche Unternehmen dar.

Auch das schlechte Image von Deepfakes trägt zum Unmut bezüglich Deepfakes dar. In wissenschaftlichen Ausarbeitungen überwiegt die Darstellung der negativen Konsequenzen von Deepfakes und die Probleme, die Personen dadurch erleiden deutlich (Albahar & Almalki, 2019; Buo, 2020; Hancock & Bailenson, 2021), was ein negatives Abbild in Bezug auf Deepfakes impliziert.

- **Probleme für Kunden**

Die Probleme mit dem Deepfake-Marketing für Kunden unterteilen sich in zwei Lager. Das erste große Problem betrifft die rechtliche Absicherung.

Wie die Unternehmen können sich auch Privatpersonen nur schwer rechtlich gegen den missbräuchlichen Nutzen privater Daten absichern. Laut § 22 Satz 1 Kunsturhebergesetz (KUG) dürfen Bilder und Videos einer Person nur nach deren Einwilligung verbreitet und öffentlich zur Schau gestellt werden. Stellen Personen ihre privaten Gesichtsdaten Unternehmen wie Hour One zur Verfügung können diese z.B. für Deepfakes genutzt werden

ohne, dass die Person Einfluss auf den Inhalt dessen hat, für den das eigene Gesicht gerade in der Werbeanzeige wirbt. Diese rechtlichen Grauzonen müssen für erfolgreiches, sicheres Deepfake-Marketing eliminiert werden. Auch durch die Datenschutz-Grundverordnung (DSGVO) sind Personen vor der unbefugten Verbreitung personenbezogener Daten geschützt. So besteht laut Art. 17 DSGVO das Recht auf Löschung und in begründeten Fällen Schadensersatzansprüche (Art. 82 DSGVO). Die DSGVO greift hier, da Personen insbesondere in visuellen Deepfakes wie z.B. Face Swaps in der Regel nach Art. 4 DSGVO identifizierbar sind. Eine erste Lösung und Sicherheitsmaßnahme für Deepfake-Ersteller wäre somit die schriftliche Einwilligung der Personen unter Erfüllung der in Art. 7 DSGVO vorgeschriebenen Bedingungen einzuholen.

Das andere große Problem betrifft die Skepsis der Nutzer gegenüber Neuem bzw. dieser neuen Medienform.

Dadurch, dass Deepfakes erst seit Ende 2017 veröffentlicht wurden und mit den stetigen Entwicklungen jetzt langsam der breiten Masse zur Verfügung stehen entstehen immer mehr Berührungspunkte im Alltag. Snapchatfilter mit Face Swapping Technologie, virtuelle Models beim Online-Shopping, Parodievideos mit US-Präsidenten, die Geschichten vorlesen und ganz aktuell Deepfake-Videos der beiden beteiligten Präsidenten des Ukraine-Kriegs: Mit all dem kommen Nutzer derzeit in Berührung und sehen, was möglich ist und in Zukunft noch qualitativ hochwertiger sein wird. Aufkommende Skepsis und Ängste in Bezug auf Datensicherheit, verschwimmende Grenzen zwischen

Realität und Illusion, sowie schädlicher Betrug stellen das Sinnbild von „Seeing is Believing“ in Frage. Deepfakes verwirren Menschen (Vaccari & Chadwick, 2020). Die aktuellen Aufklärungsmechanismen der Bundesregierung illustrieren die verschiedenen Arten von Deepfakes, geben praktische Beispiele und Hinweise, an welchen Artefakten Deepfakes zu erkennen sind (BSI, 2022). Doch laut Köbis et al. (2021) können Menschen Deepfakes nicht zuverlässig erkennen. Das Bundesamt für Sicherheit in der Informationstechnik (BSI, 2022) gibt jedoch an, dass die Erkennung von Deepfakes durch Aufklärung signifikant gesteigert werden kann.

Unterstützung könnte die Aufklärung durch öffentliche Deklaration von Deepfakes erfahren. Die Offenlegung von Werbeinhalten ist im Marketing kein Neuland, so sind z.B. Content Creator und Influencer in sozialen Medien dazu verpflichtet, laut Medienstaatsvertrag §8 und §22 gesponserte und bezahlte Beiträge zu kennzeichnen (Medienstaatsvertrag, 2020). Das große Ziel liegt darin, das Werbevorhaben klar zu kommunizieren und dem Nutzer sichtbar zu machen, da Verbraucher ein Recht darauf haben zu wissen, wann sie mit werblichen Inhalten in Kontakt treten (Boermann et al., 2018, S. 146). So zeigt bisherige Forschung, dass die Offenlegung zwar die Bewertung, Glaubwürdigkeit und Einstellung gegenüber der Marke verringert, sich gleichzeitig aber die Anerkennung und das Überzeugungswissen erhöht (Naderer et al., 2021, S. 689). So schafft die Offenlegung von Werbung den Grundstein für Transparenz von Werbeabsichten auf Nutzerebene. Übertragen auf die

Deepfake-Technologie und in Kombination mit den lauter werdenden Forderungen nach Regulierungsansätzen könnte die Offenlegung von Deepfake-Werbung einen ersten Schritt zum Abbau von Skepsis gegenüber Deepfakes beitragen und in nächster Konsequenz die Akzeptanz hyperpersonalisierter Werbung erhöhen. In Fortführung der Gedanken von Whittaker et al. (2021, S. 21f) zur öffentlichen, transparenten Kennzeichnung von Deepfakes wird deshalb folgende erste Hypothese aufgestellt:

H1: Die offensichtliche Deklaration als Deepfake wirkt sich positiv auf die Akzeptanz von hyperpersonalisierter Werbung aus.

Auf Grundlage der Theory of Reasoned Action von Ajzen und Fishbein besteht die Vermutung, dass die Einordnung des persönlichen Empfindens zu Deepfakes als glaubwürdig/authentisch zur Akzeptanz der Deepfake-Werbemaßnahmen führen kann. Kietzmann et al. (2021, S. 9) betonen in dem Zusammenhang ausdrücklich: „[...] *it is not yet clear how people will respond to hyper-personalized ads.*“. In seiner Veröffentlichung forscht Reichelt (2021) darüber, was Deepfakes glaubhaft wirken lässt. Deepfakes müssen so realistisch wie möglich sein, um Glaubwürdigkeit auszustrahlen, wobei dann immer noch die subjektive Wahrnehmung auf die persönliche Entscheidung einwirkt (Reichelt, 2021, S. 6). Auch aus der VXF-Perspektive lässt sich durch einen gesteigerten Realitätsgrad zu erhöhter Akzeptanz der Filme (z.B. Jurassic World) schließen (Fontana, 2018). So gilt es die in Kapitel 3.5. aufgestellten Determinanten bei der

Produktion der Deepfakes für die Umfrage dieser Arbeit zu berücksichtigen und folgende Hypothese zu überprüfen:

H2: Je authentischer der Deepfake ist, desto niedriger ist die Gefahr von Skepsis ggü. der Werbemaßnahme.

Die Ergebnisse von Ham's (2017) Forschung zum Umgang mit Online-Werbung deutete darauf hin, dass sich Nutzer bei stark personalisierten Inhalten unwohl und irritiert fühlen, was gleichzeitig mit Privatsphärebedenken und Skepsis einhergeht. Nutzer nehmen personalisierte Werbung als Überredungsversuche und Manipulation wahr, so Simonson (2003). Trotzdem stellt Ham (2017, S. 20) fest, dass personalisierte Werbung einen positiven Effekt auf die Überzeugungskraft, den empfundene Nutzen hat und sogar eine Steigerung der gewünschten Aktion auslösen kann (Kietzmann et al., 2021, S. 5). Stefan Herzog vom Max-Planck-Institut ergänzt zu dem Thema: „Hier zeigt sich eine Diskrepanz in der Einstellung. Einerseits akzeptiert die Mehrheit maßgeschneiderte [...] Werbeanzeigen. Andererseits aber lehnt sie das derzeitige Sammeln von Daten ab, die eine solche Personalisierung ermöglicht“ (Max-Planck-Gesellschaft, 2020). Hier zeigt sich also wieder die in Kapitel 2.6 angesprochene Abhängigkeit der Akzeptanz personalisierter Werbung von personengebundenen Faktoren. Zu diesen personengebundenen Faktoren können auch emotionale Bindungen gehören. Dies bestätigt auch die Studie „Make A Lasting Impression“, in der eine Steigerung des Absatzes bei Werbung, die Menschen emotional und kreativ anspricht, erkannt wurde (Kantar Millward Brown, 2017, S. 12).

So zeigte sich in der Studie von Nielsen (2015, S. 4), dass 8 von 10 Konsumenten persönlichen Empfehlungen mehr vertrauen als allen anderen Werbeformen. Die Akzeptanz von personalisierter Werbung ist somit von der emotionalen, sowie kreativen Beschaffenheit der Werbung in Kombination mit persönlichen Empfehlungen von z.B. Influencern oder Personen aus dem privaten Umfeld abhängig. Daraus wird geschlussfolgert, dass die Produktion von Deepfakes mit Gesichtsmaterial aus dem privaten Umfeld (z.B. Freunde, Bekannte, Familie) eine ähnliche Emotionalisierung und Kreativitätssteigerung darstellen kann, wie in der Studie von Kantar Millward Brown (2017). Demnach wird die nächste Hypothese aufgestellt, die es im Anschluss zu untersuchen gilt:

H3a: Deepfakes mit Material aus dem privaten Umfeld haben einen positiven Effekt auf die Akzeptanz hyperpersonalisierter Werbung.

Neben den positiven Effekten muss an dieser Stelle wieder auf die Datenschutzbedenken der Konsumenten geachtet werden. In der repräsentativen Studie von Deutschland sicher im Netz (2021) zeigt sich, dass Menschen der Internetnutzung im Allgemeinen skeptisch bezüglich der Sicherheit gegenüber stehen und in Konsequenz Abstriche bei der Nutzung von Angeboten machen würden. Gleichzeitig geben aber über 20% der Befragten an, gern Neues auszuprobieren (S. 28). Sie sind also offen für Neues trotz Sicherheitsbedenken. Mit der Umfrage soll untersucht werden, ob Probanden bereit sind, Bildmaterial für die Produktion der Deepfakes zur Verfügung zu stellen und sich

selbst wissentlich manipulieren zu lassen. So lässt sich an dieser Stelle die letzte Hypothese definieren:

H3b: Deepfake-Werbung mit dem eigenen Gesicht schürt Bedenken zum Datenschutz und missbräuchlicher Nutzung.

Resultierend aus den vorangestellten Untersuchungen lassen sich zwei Oberkategorien zu den Effekten auf Rezipienten aufstellen: **Skepsis** und **Akzeptanz**. Für das Deepfake-Marketing bedeutet dies das zweigeteilte Ziel der Reduktion von Skepsis gegenüber der Deepfake-Technologie und eine Erhöhung der Akzeptanz von hyperpersonalisierten Werbeangeboten.

• **Hypothesenmodell**

Folglich kann das Hypothesenmodell aufgestellt werden. Darin werden die möglichen Zusammenhänge von Authentizität, Deklarierung und Inhalten der Deepfakes zur Senkung der Skepsis gegenüber Deepfakes und der Erhöhung der Akzeptanz gegenüber hyperpersonalisierter Werbung grafisch dargestellt. Dabei stellen die letzten beiden genannten Variablen die annehmbaren Effekte dar, die Deepfake-Marketing im Idealfall bei Rezipienten auslöst.

Deklarierung der Deepfakes

H1: Die offensichtliche Deklarierung als Deepfake wirkt sich positiv auf die Akzeptanz von hyperpersonalisierter Werbung aus.

Authentizität der Deepfakes

H2: Je authentischer der Deepfake ist, desto niedriger ist die Gefahr von Skepsis gegenüber der Werbemaßnahme.

Inhalte der Deepfakes

H3a: Deepfakes mit Material aus dem privaten Umfeld haben einen positiven Effekt auf die Akzeptanz hyperpersonalisierter Werbung.

H3b: Deepfake-Werbung mit dem eigenen Gesicht schürt Bedenken zum Datenschutz und missbräuchlicher Nutzung.

Effekte für das Deepfake-Marketing

Akzeptanz hyperpersonalisierter Werbung erhöhen

Skepsis gegenüber Deepfakes senken

Abbildung 20 Hypothesenmodell

4.4 Szenario Entwicklung

Für die geplante Umfrage sowie die Produktion der Deepfakes soll an dieser Stelle ein Szenario entwickelt werden, wie die Deepfakes in Zukunft im Marketing eingesetzt werden könnten. Der in dieser Arbeit angestrebte Einsatz von hyperpersonalisierten Werbemaßnahmen durch Deepfakes könnte sich in Zukunft wie folgt gestalten:

Die Nutzerdaten von Konsumenten wie z.B. Tom und Luna werden durch KI-basierende Anwendungen gesammelt, gespeichert, ausgewertet und intelligent weiterverarbeitet. Dadurch, dass die Daten aus verschiedenen Quellen stammen können die Algorithmen tiefgreifende Zusammenhänge erkennen und voraussagen, welche Präferenzen, Wünsche oder Bedenken die jeweiligen Nutzer haben.

Luna begeistert sich ganz besonders für die Marke M und zeigt dies durch Suchen von Inhalten, Liken, Kommentieren von Beiträgen und/oder dem Posten von kurzen Videos mit einem Produkt auf sozialen Medien. Aufgrund spezieller Algorithmen, die eine Auswahl treffen, welche Personen für Unternehmen interessant werden, um mit dem Produkt zu werben, bekommt Luna das Angebot in der Deepfake-Werbekampagne mitzuwirken. Sobald Luna bestätigt hat, dass ihr Gesicht (und Stimme) für die Deepfake-Werbung genutzt werden darf, muss qualitativ hochwertiges Videomaterial angefertigt werden. Dafür nimmt sie zuhause oder in einer speziellen Partner-Agentur der Marke M ein 2-minütiges Video vor neutralem Hintergrund nach speziellen

Vorgaben auf, indem sie sich von allen Seiten filmt und bestimmte Sätze spricht. Nach Freigabe des Videos durch die Marke M wird Lunas Gesicht in den Deepfake-Werbepot der Marke M eingefügt und für nachfolgende Kampagnen gespeichert.

Tom, ein Freund aus der Heimatstadt folgt Luna in den sozialen Medien. Da der Algorithmus erkannt hat, dass Tom nach Schuhen der Marke M Ausschau hält, erhält er eine Werbeanzeige von Marke M mit dem Deepfake-Werbepot, der auf Lunas Gesicht basiert und für ähnliche Schuhe wirbt, die er sich bereits angesehen hat.

Welche Reaktionen löst Deepfake-Werbung bei Tom bzw. den Konsumenten aus? Dies gilt es nun zu untersuchen.

5. Forschungsdesign

Das Forschungsdesign umfasst die Dokumentation der Produktion der Deepfakes und die Online-Umfrage. Mit dem Vorgehen sollen die Hypothesen verifiziert oder falsifiziert und die Forschungsfrage beantwortet werden. Die Vorgehensweise wird dokumentiert, um den Prozess transparent darzustellen und die Reproduzierbarkeit dieser Studie zu gewährleisten.

5.1 Technologie und Arbeitsweise von DeepFaceLab

DeepFaceLab ist eine Open-Source-Software zur Erstellung von Deepfakes. Über GitHub können die Skripte sowie die komplette Anwendung kostenfrei bezogen werden. Zu beachten ist dabei der richtige *build* für die jeweilige Grafikkarte. Empfohlen wird die Anwendung insbesondere für NVIDIA-Grafikkarten. An einer Workstation in der Technischen Hochschule Brandenburg im

Netzwerk- und Sicherheitslabor werden die Deepfakes produziert. Die Ausstattung der Workstation wird im Folgenden mit der empfohlenen Ausstattung aus dem DeepFaceLab-Guide in Vergleich gesetzt.

Anders als für das macOS-Betriebssystem kommt DeepFaceLab für Windows als ausführbare Datei, die nicht explizit über die Kommandozeile installiert und aufgerufen werden muss. Lediglich Parameteranpassungen während der Benutzung werden dabei über die Kommandozeile gesteuert. Die Basis der DFL-Kenntnisse stellt der offizielle DFL-Guide dar. Sollte im weiteren Verlauf von Empfehlungen gesprochen werden, handelt es sich dabei um Empfehlungen aus dem offiziellen DFL-Guide (TMBDF, 2020).

Die Deepfakes werden aus zwei Dateien generiert: `data_src` beinhaltet dabei Videomaterial, das die Gesichtsquelle (source) zeigt. Das Zielvideo wird mit

	Workstation im Labor	Empfehlung von DFL (TMBDF, 2020)
Prozessor (CPU)	4 Kernel	mind. 4 Kernel, besser 8 Kernel
Arbeitsspeicher (RAM)	32 GB	32 GB
SSD-Speichergröße	1 TB	500 GB, besser 1 TB
verfügbarer Speicher	498 GB frei	mind. 32 GB, besser 128 GB
Grafikkarte (GPU)	NVIDIA GTX 1080 Ti	NVIDIA GTX 900-1000 Serie NVIDIA RTX 2000 Serie NVIDIA RTX 3000 Serie AMD Radeon

Tabelle 2 Vergleich der empfohlenen und tatsächlichen Hardware für DeepFaceLab

data_dst für (destination) bezeichnet. Als Ausgangsmaterial für data_src dienen die selbst angefertigten Videos mit den Probanden. Die Zielvideos sind die bestehenden Werbeclips, welche unter dem Namen data_dst abgespeichert sind.

Das Anfertigen der src-Videos erfolgt für eine einheitliche Ausgangsbasis nach einem vorher definierten Schema. Dieses kann im Anhang eingesehen werden. Um eine große Varietät an Material zu bekommen, wurden verschiedene Winkel, Gesichtszüge und Mundbewegungen im Video erzeugt. In der folgenden Tabelle wird die Dauer der einzelnen Videos mit Angabe der Einzelframes bei der genutzten Framerate 25fps angegeben.

Für die dst-Videos werden zwei Werbespots ausgewählt. Diese beinhalten Szenen mit den Schauspielern Christoph Maria Herbst und Malin Steffen. Für die männlichen Deepfakes mit Christoph Maria Herbst werden drei Sequenzen aus verschiedenem Videomaterial mit seiner Figur „Bernd Stromberg“ genutzt (Youtube (1), 2021; Youtube (2), 2021; Youtube (6), 2016). Diese dst-Sequenz beträgt 1:30min (bei 25fps = 2.250 Frames).

Für die Stepstone Deepfakes von Pia und Paulina werden drei Sequenzen aus verschiedenem Videomaterial mit Malin Steffen genutzt (Youtube (3), 2022; Youtube (4), 2021; Youtube (5), 2017). Diese dst-Sequenz beträgt 1:43min (bei 25fps = 2.575 Frames).

Im offiziellen DFL-Guide (TMBDF, 2020) wird empfohlen die Größe des Datensatzes bei 3.000-8.000 Gesichtern zu halten. Größere Mengen erhöhen die Trainingsdauer erheblich, laut Guide. Hauser (2018) fand jedoch durch Experimente heraus, dass die Ähnlichkeit des Materials in Bezug auf Beleuchtung und die Gesichtszüge wichtiger ist als die reine Anzahl der Gesichter. Aufgrund dieser Erkenntnis liegt die Anzahl im Experiment der Masterarbeit in einem Bereich von 2.250 bis 3.400 Gesichtern, da eine Diversität bei der Aufnahme sichergestellt wurde.

src-Werte	Proband 1 „Paulina“	Proband 2 „Pia“	Proband 3 „Micha“	Proband 4 „Alex“
Dauer in min	2:02	2:16	2:16	2:09
Frames gesamt	3.050	3.400	3.400	3.225

Tabelle 3 Dauer und Frameanzahl der src-Videos

5.2 Dokumentation des Vorgehens

Im Folgenden wird der Prozess der Generierung eines Deepfakes erläutert. Schritt für Schritt werden so die einzelnen Komponenten zur Finalisierung des Deepfakes erklärt.

Schritt 1: Extrahieren der Einzelframes aus dem Videomaterial

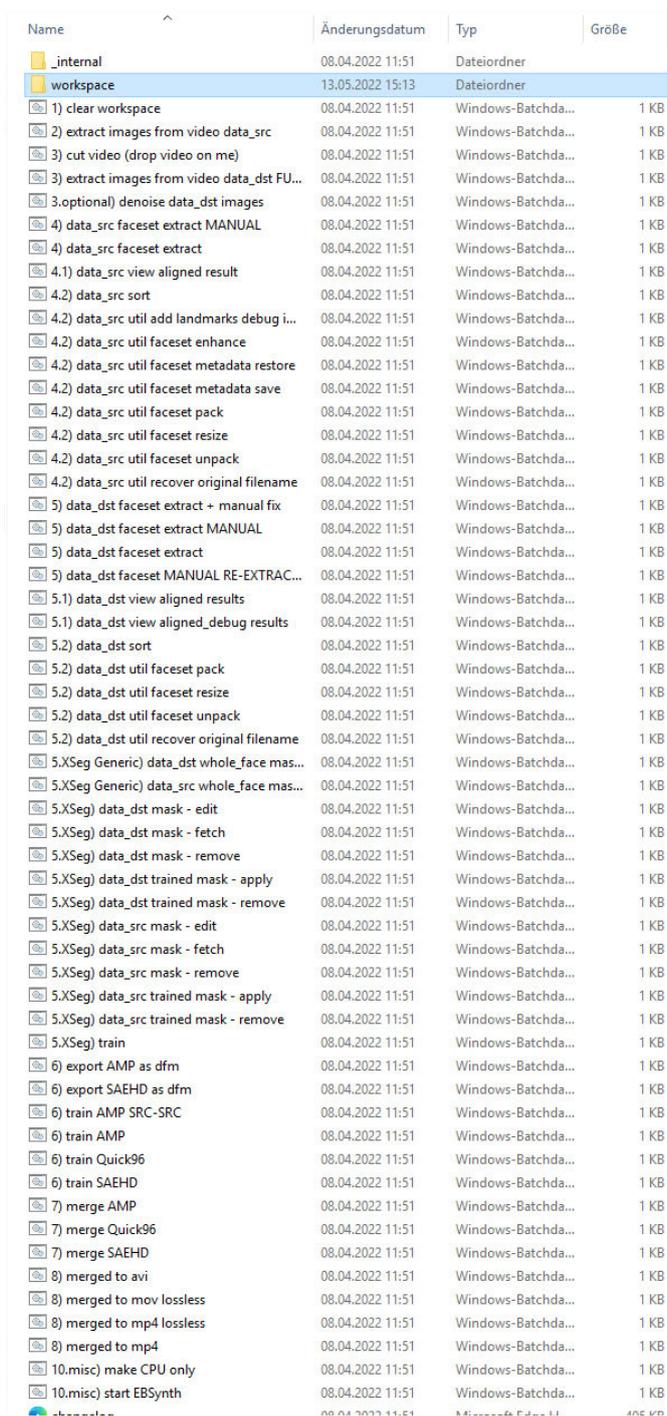
Im ersten Schritt werden die Einzelframes aus dst und src unter Angabe der Framerate und dem Format (Wahl zwischen .jpg und .png). Für qualitätsverlustfreie Einzelbilder wurde das Format .png mit einer Framerate von 25fps gewählt, da diese bereits im Ursprungmaterial vorlag. Die Einzelframes werden in separate Ordner gespeichert. Für die Version „Paulina“ bedeutet dies 3.046 src-Bilder sowie 2.579 dst-Bilder.

Befehle:

```
[2] extract images from video data_src]
[3] extract images from video data_dst
FULL FPS]
```

Schritt 2: Gesichter aus src und dst extrahieren

Die Gesichter werden aus den Einzelframes der beiden Videos separat extrahiert. Dafür wird die GPU ausgewählt. Als nächstes muss sich für eine Methode entschieden werden: FF, WF und HEAD stehen dabei zur Auswahl. FF (Full-Face) eignet sich besonders für niedrige Gesichtstypen, WF (Whole Face) wird als universelle Lösung empfohlen und beim Typ HEAD wird der ganze Kopf verwendet, was für das geplante Vorhaben unnötig ist. Für 1080p Quellvideos wird eine Auflösung von 768 – 1024 empfohlen. Für optimale Ergebnisse wurde



Name	Änderungsdatum	Typ	Größe
workspace	13.05.2022 15:13	Dateiordner	
1) clear workspace	08.04.2022 11:51	Windows-Batchda...	1 KB
2) extract images from video data_src	08.04.2022 11:51	Windows-Batchda...	1 KB
3) cut video (drop video on me)	08.04.2022 11:51	Windows-Batchda...	1 KB
3) extract images from video data_dst FU...	08.04.2022 11:51	Windows-Batchda...	1 KB
3.optional) denoise data_dst images	08.04.2022 11:51	Windows-Batchda...	1 KB
4) data_src faceset extract MANUAL	08.04.2022 11:51	Windows-Batchda...	1 KB
4) data_src faceset extract	08.04.2022 11:51	Windows-Batchda...	1 KB
4.1) data_src view aligned result	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src sort	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src util add landmarks debug i...	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src util facenet enhance	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src util facenet metadata restore	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src util facenet metadata save	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src util facenet pack	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src util facenet resize	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src util facenet unpack	08.04.2022 11:51	Windows-Batchda...	1 KB
4.2) data_src util recover original filename	08.04.2022 11:51	Windows-Batchda...	1 KB
5) data_dst faceset extract + manual fix	08.04.2022 11:51	Windows-Batchda...	1 KB
5) data_dst faceset extract MANUAL	08.04.2022 11:51	Windows-Batchda...	1 KB
5) data_dst faceset extract	08.04.2022 11:51	Windows-Batchda...	1 KB
5) data_dst faceset MANUAL RE-EXTRAC...	08.04.2022 11:51	Windows-Batchda...	1 KB
5.1) data_dst view aligned results	08.04.2022 11:51	Windows-Batchda...	1 KB
5.1) data_dst view aligned_debug results	08.04.2022 11:51	Windows-Batchda...	1 KB
5.2) data_dst sort	08.04.2022 11:51	Windows-Batchda...	1 KB
5.2) data_dst util facenet pack	08.04.2022 11:51	Windows-Batchda...	1 KB
5.2) data_dst util facenet resize	08.04.2022 11:51	Windows-Batchda...	1 KB
5.2) data_dst util facenet unpack	08.04.2022 11:51	Windows-Batchda...	1 KB
5.2) data_dst util recover original filename	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg Generic) data_dst whole_face mas...	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_dst mask - edit	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_dst mask - fetch	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_dst mask - remove	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_dst trained mask - apply	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_dst trained mask - remove	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_src mask - edit	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_src mask - fetch	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_src mask - remove	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_src trained mask - apply	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) data_src trained mask - remove	08.04.2022 11:51	Windows-Batchda...	1 KB
5.XSeg) train	08.04.2022 11:51	Windows-Batchda...	1 KB
6) export AMP as dfm	08.04.2022 11:51	Windows-Batchda...	1 KB
6) export SAEHD as dfm	08.04.2022 11:51	Windows-Batchda...	1 KB
6) train AMP SRC-SRC	08.04.2022 11:51	Windows-Batchda...	1 KB
6) train AMP	08.04.2022 11:51	Windows-Batchda...	1 KB
6) train Quick96	08.04.2022 11:51	Windows-Batchda...	1 KB
6) train SAEHD	08.04.2022 11:51	Windows-Batchda...	1 KB
7) merge AMP	08.04.2022 11:51	Windows-Batchda...	1 KB
7) merge Quick96	08.04.2022 11:51	Windows-Batchda...	1 KB
7) merge SAEHD	08.04.2022 11:51	Windows-Batchda...	1 KB
8) merged to avi	08.04.2022 11:51	Windows-Batchda...	1 KB
8) merged to mov lossless	08.04.2022 11:51	Windows-Batchda...	1 KB
8) merged to mp4 lossless	08.04.2022 11:51	Windows-Batchda...	1 KB
8) merged to mp4	08.04.2022 11:51	Windows-Batchda...	1 KB
10.misc) make CPU only	08.04.2022 11:51	Windows-Batchda...	1 KB
10.misc) start EBSynth	08.04.2022 11:51	Windows-Batchda...	1 KB

Abbildung 21 Übersicht der verfügbaren Skripte in DeepFaceLab

hier der Wert 1024 genutzt und die höchste „jpeg quality“ 100% für beste Qualität gewählt. Nach ca. 40min wurden für die Version „Paulina“ 2.983 Quellgesichter und 2.124 Zielgesichter erkannt. Eine Testversion mit 264 Zielgesichtern nur aus dem Werbeclip lieferte unzureichende Ergebnisse.

Befehle:

```
[4) data_src faceset extract]
[5) data_dst faceset extract]
```

Schritt 3: Gesichtsdatensätze bereinigen

Anschließend werden die Bilddaten sortiert. Hierbei stehen 15 verschiedene Vorgehensweisen zur Verfügung. Für die Deepfakes wurde die Histogramm-Ähnlichkeit gewählt, da dieses Vorgehen ähnlich aussehende Bilder gruppiert. So

können schlechte Bilder zeitlich effizienter gelöscht werden. Schlechte Bilder kennzeichnen sich durch verschwommene Gesichter, unzureichende Belichtung, falsch ausgerichtete Gesichter, abgeschnittene Gesichter, Gesichter fremder Personen oder Gegenstände, die das Gesicht verdecken. Nach den eben genannten Kriterien müssen die Bilder manuell aus dem workspace > data_src > aligned Ordner gelöscht werden. Der bereinigte Datensatz für „Paulina“ besteht nun aus 2.930 Bildern für die Quelle (Testperson) und 2.077 Bilder für das Zielvideo.

Befehle:

```
[4.2) data_src sort]
[4.1) data_src view aligned result]
[5.2) data_dst sort]
[5.1) data_dst view aligned results]
```

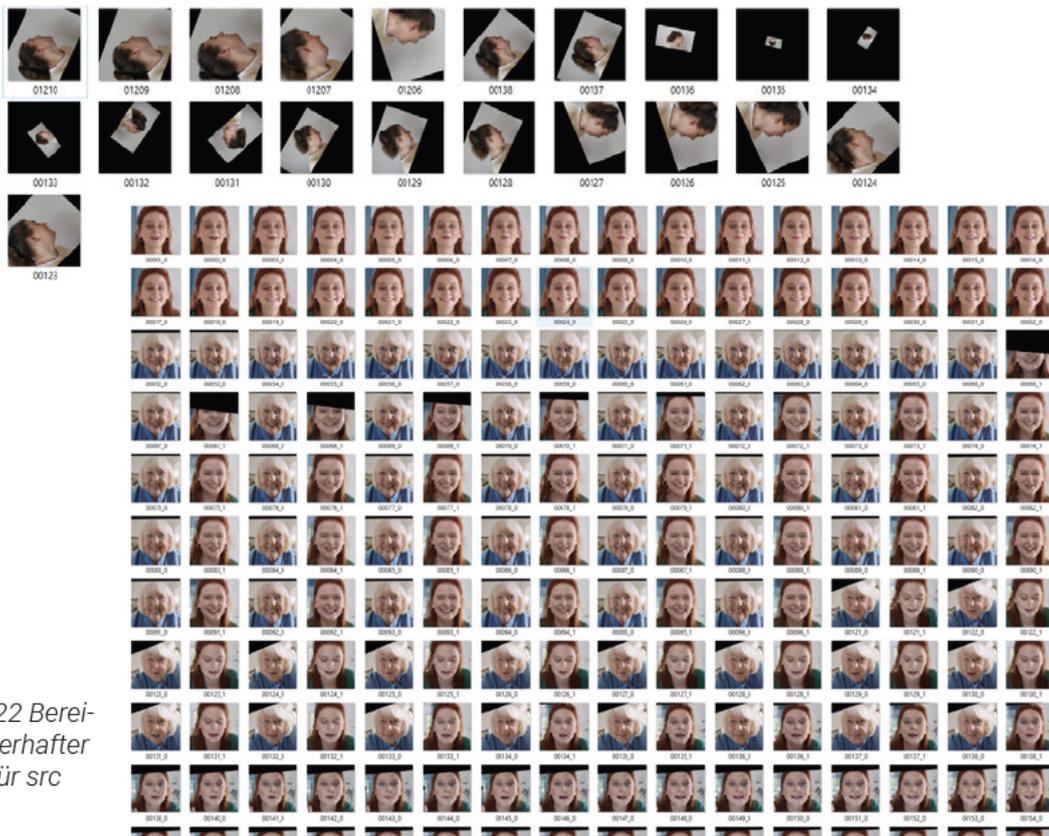


Abbildung 22 Bereinigung fehlerhafter Gesichter für src und dst

Schritt 4: Gesichter maskieren & trainieren
 Sobald die Gesichter extrahiert und bereinigt sind können die Masken für src und dst hinzugefügt werden. Der dafür integrierte XSeg-Editor stellt dafür Polygontools bereit, mit denen man auf den Einzelframes manuell das Gesicht einrahmt. Der DFL-Guide empfiehlt hier nicht alle Gesichter, sondern nur 50 - 200 Gesichter, die wesentliche Veränderungen ausdrücken, manuell zu maskieren. Darunter fallen z.B. Änderungen des Gesichtsausdrucks (offener Mund, Kussmund, Lächeln), Änderungen der Blickrichtung, des Winkels oder der Lichtverhältnisse. Für WF (Whole Face) wird dabei eine Maskierung von der Haarlinie bis zum Kinn empfohlen.

Anschließend werden diese ausgewählten Masken für das Training des gesamten Gesichtsdatsatzes verwendet. Sobald der Loss Value beim Training stagniert und die

Maskenumrisse scharf aussehen, können die Masken auf die Gesichter übertragen werden. Nach 30.000 Iterationen konnten gute Ergebnisse der Gesichtserkennung erzielt werden und die Maskenumrisse der XSeg-Masken waren klar erkennbar und somit bereit für die Synthese der Gesichter.

In der folgenden Tabelle sind die Anzahlen der Gesichter, der maskierten Gesichter, der Iterationen aus dem Trainingsprozess, sowie die Dauer in Stunden und die Loss Values aufgelistet. Bei der Interpretation der Loss Values geht es darum, niedrige Werte von 0,4 bis 0,15 zu erzielen, da diese auf einen hohen Trainingsfortschritt des Modells hinweisen (TMBDF, 2020). In Experimenten zeigte sich jedoch, dass sich nicht allein auf die Loss Values verlassen werden darf, da die visuelle Einschätzung genauso wichtig ist.

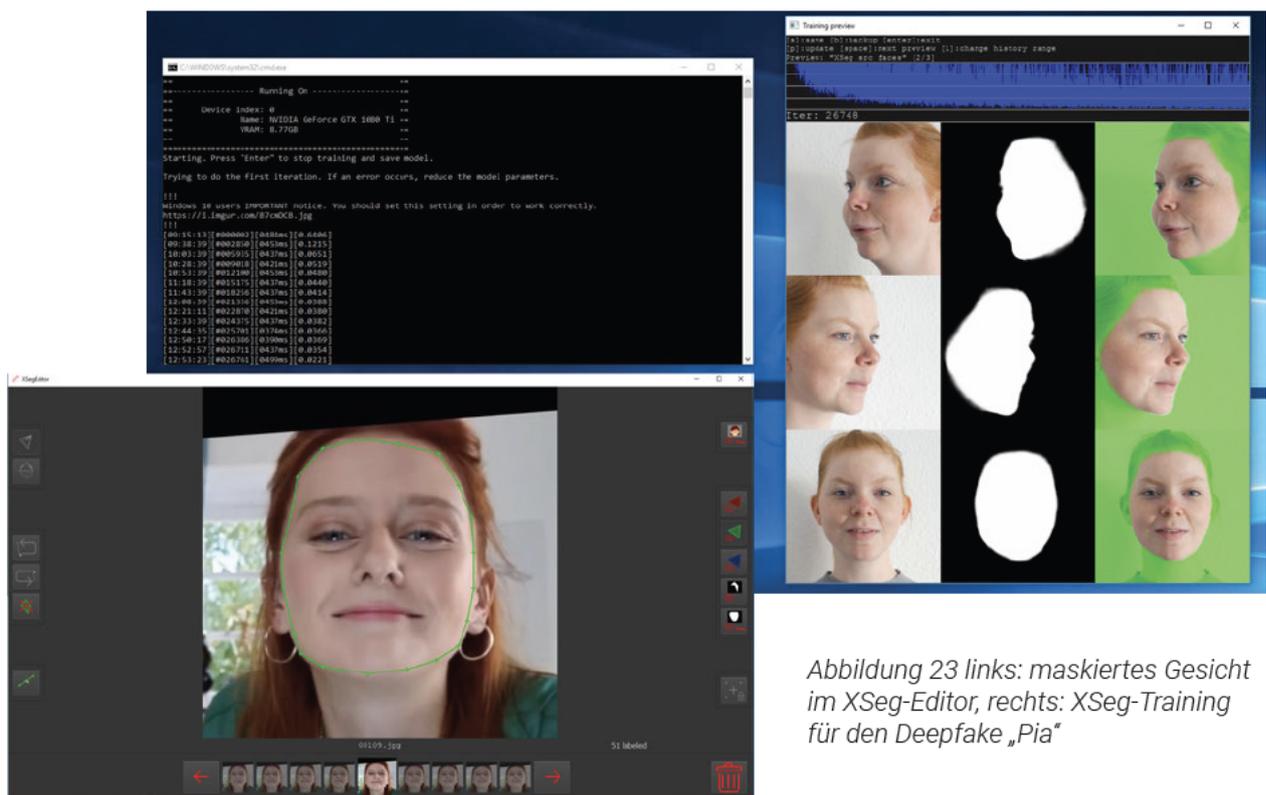


Abbildung 23 links: maskiertes Gesicht im XSeg-Editor, rechts: XSeg-Training für den Deepfake „Pia“

Befehle:

```
[5.XSeg) data_dst mask - edit]
[5.XSeg) data_src mask - edit]
[5.XSeg) train]
[5.XSeg) trained mask data_dst - apply]
[5.XSeg) trained mask data_src - apply]
```

XSeg-Werte	Proband 1 „Paulina“	Proband 2 „Pia“	Proband 3 „Micha“	Proband 4 „Alex“
src: Anzahl Gesichter	2.930	3.228	3.025	2.941
dst: Anzahl Gesichter	2.077	2.077	2.040	2.040
Anzahl maskierte Gesichter in src + dst	50 + 47 = 97	58 + 47 = 105	47 + 89 = 136	52 + 89 = 141
Anzahl Iterationen	30.482	26.748	29.120	33.000
Dauer Training in h	4	4	3	4,5
Loss Values	0.0272	0.0354	0.0362	0.0408

Tabelle 4 Übersicht der XSeg-Werte

Schritt 5: Training des Modells

Um die Gesichter zu trainieren, gibt es drei Methoden: SAEHD, AMP und Quick96. Da sich AMP noch in der Entwicklung befindet und Quick96 nur für schnelle Tests empfohlen wird, wurde sich für den Versuch auf SAEHD konzentriert. Ausgelegt für high-end GPUs ist der „High Definition Styled Auto Encoder“ laut DFL-Guide für die meisten User Standard und erfordert mindestens 6 GB VRAM.

SAEHD erfordert die Einstellung verschiedener Parameter vor dem Start des Trainings. So wurden die in der Tabelle auf der nächsten Seite aufgelisteten Einstellungen vorgenommen.

Die blau markierten Werte sind Standardeinstellungen, die aus dem offiziellen DFL-Guide übernommen wurden. So wurden die Einstellungen zu den Dimensionen des neuronalen Netzwerks bewusst nicht geändert, da dies große Einflüsse auf die Performance und Qualität der Deepfakes haben können [ae_dims etc.].

SAEHD-Werte	Proband 1 „Paulina“	Proband 2 „Pia“	Proband 3 „Micha“	Proband 4 „Alex“
resolution	256	256	256	256
face_type	wf	wf	wf	wf
models_opt_on_gpu	true	true	true	true
archi	df	df	df	df
ae_dims	256	256	256	256
e_dims	64	64	64	64
d_dims	64	64	64	64
d_mask_dims	22	22	22	22
masked_training	true	true	true	true
eyes_mouth_prio	false, ab 100.000 Iter* für 19.000 Iter true	false, ab 105.000 Iter für 18.000 Iter true	false, ab 102.000 Iter für 28.000 Iter true	false, ab 108.000 Iter für 15.000 Iter true
uniform_yaw	true	true	true	true
blur_out_mask	true	true	true	true
adabelief	true	true	true	true
lr_dropout	false	false	false	false
random_warp	true	true, ab 105.000 Iter false	true, ab 102.000 Iter false	true, ab 108.000 Iter false
random_hsv_power	0.0	0.0	0.0	0.0
true_face_power	0.0	0.0	0.0	0.0
face_style_power	0.0	0.0	0.0	0.0
bg_style_power	0.0	0.0	0.0	0.0
ct_mode	none	none	none	none
clipgrad	true	true	true	true
pretrain	false	false	false	false
autobackup_hour	1	1	1	1
write_preview_history	true	true	true	true
target_iter	0	0	0	0
random_src_flip	true	true	true	true

random_dst_flip	true	true	true	true
batch_size	5	5	5	5
gan_power	0.0	0.0	0.0	0.0
gan_patch_size	32	32	32	32
gan_dims	14	14	14	14
Dauer in h für SAEHD	40h regulär, 6h für Eyeprio	43h regulär, 6,5h für Eyeprio	40h regulär, 12h für Eyeprio	43,5h regulär, 6 h für Eyeprio
Loss Values (src/dst)	0,2815/ 0,2389	0,2157/ 0,2287	0,2018/ 0,2365	0,2151/ 0,2343
Trainingsdauer insgesamt in h	46h SAEHD + 6h XSeg = 52h	49,5h SAEHD + 4h XSeg = 53,5h	52h SAEHD + 3h XSeg = 55 h	49,5h SAEHD + 4,5h XSeg = 54 h

* Iter = Iterationen

Tabella 5 Übersicht der SAEHD-Parameter

In einer Testphase (5 Wochen) wurden dafür vorab verschiedene Parameterkombinationen getestet. Die hier genutzten Werte basieren auf eigenen Erfahrungen, sowie Empfehlungen aus dem offiziellen DFL-Guide und den Erfahrungen anderer Nutzer aus dem dazugehörigen Forum.

Für hochwertige Qualität der Gesichter wurde eine Auflösung von 256x256 mit der Methode WF (Whole Face) gewählt. Die Architektur „DF“ eignet sich für die Deepfakes in diesem Fall besser als die zweite Variante „LIAE“, da im Vorfeld verschiedene Winkel im src-Material berücksichtigt wurden. Bei der Variante „DF“ wird hohe Gleichheit der Gesichter geboten. Der Nachteil dieser Variante besteht in schlechteren Beleuchtungs- und Farbergebnissen. Diese können jedoch in der Nachbearbeitung mit Adobe AfterEffects behoben werden.

Auf die Benutzung des pretrain Modus mit der Einspeisung vortrainierter Modelle anderer Nutzer wurde verzichtet, da die Gesichter auf Basis des echten Gesichts trainiert werden sollten und Einflüsse fremder Gesichter nicht erwünscht sind. Der pretrain Modus hätte den Vorteil einer verkürzten Trainingszeit gehabt.

Die Eyes-and-Mouth-Priority wurde zu Beginn des Trainings deaktiviert, da diese nach einigen Testversuchen erst bei bereits trainierten Modellen gute Ergebnisse lieferte. Dadurch konnte beispielsweise die unterschiedliche Blickrichtung der Augen in Testversuchen erheblich verbessert werden. Dies erkennt man an der Unterbrechung der gelb-blauen Loss Value Anzeige zum Ende des Trainings. Jeweils die erste Spalte stellt das Originalmaterial dar. In der zweiten Spalte wird das erzeugte Gesicht der jeweiligen Personen visualisiert. In der letzten Spalte wird das src-Gesicht auf

die Gesichtsausdrücke des dst-Gesichts trainiert und stellt somit den eigentlichen Deepfake dar.

Befehle:

[6] train SAEHD]

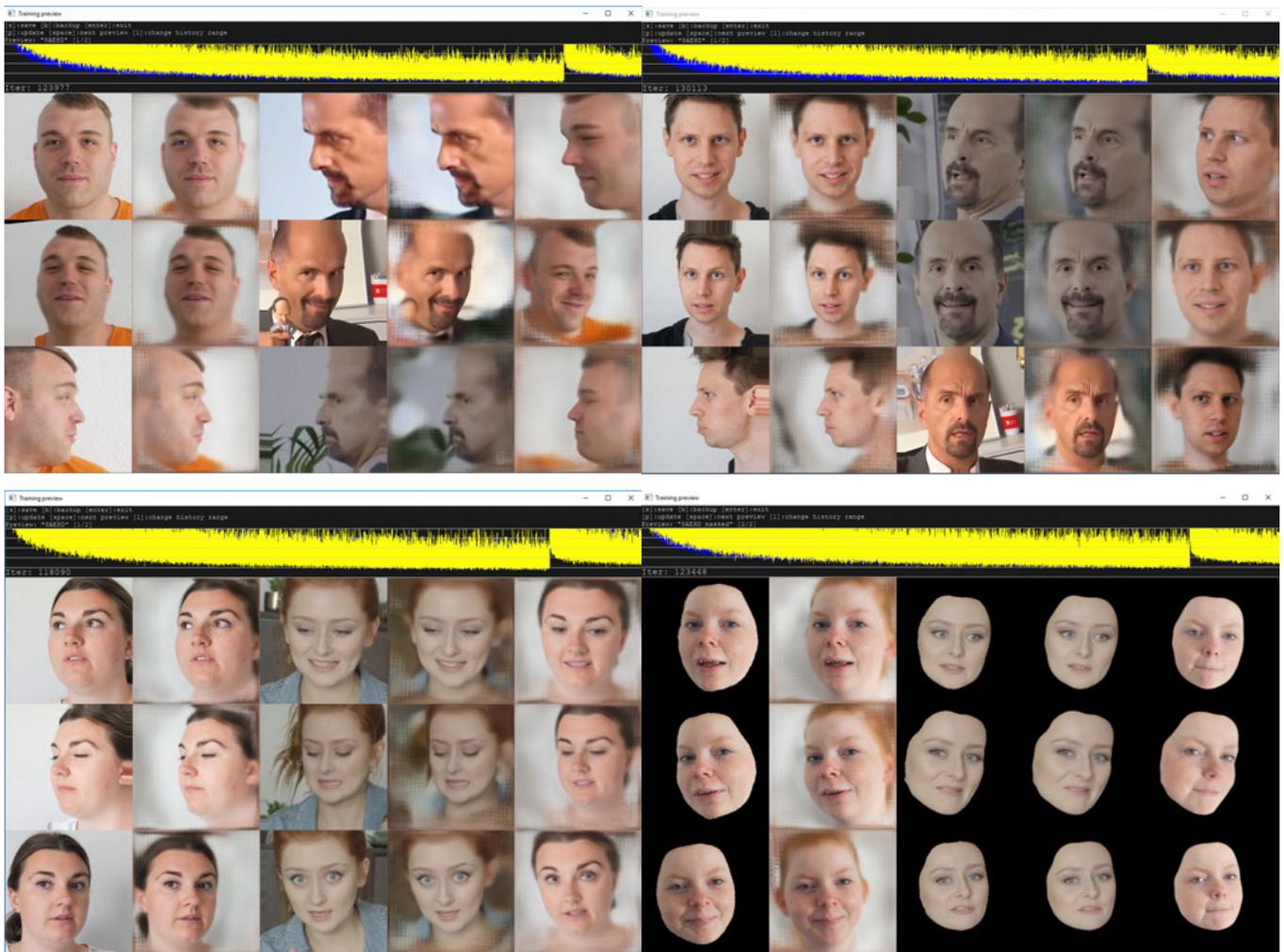


Abbildung 24 SAEHD Trainingsübersichten für die Deepfakes mit Eye-Priority

Schritt 6: Merging

Da für das Training SAEHD als Architektur genutzt wurde, wird in diesem Fall [merge SAEHD] genutzt. Der Merger besteht aus einer Vorschau und einer Bedienung über die Tastatur, welche die nächste Abbildung zeigt. In der Vorschau sieht man Änderungen sofort. Auf die genaue Angabe der einzelnen Parameterwerte für jeden Frame wird verzichtet, da dies zu weit führt. Insbesondere durch die Parameter [erode mask], [blur mask] und [face scale] konnten Verbesserungen der Maske erzielt werden. Ebenso brachte eine Veränderung bei [mask mode] und [color transfer mode] natürlichere Ergebnisse. Mit einer Tastenkombination konnte der Prozess des Mergens abgekürzt werden, da die Konfiguration automatisch auf alle folgenden Frames übertragen wurde.

Befehle:

[7) merge SAEHD]

Schritt 7: Konvertieren in abspielbare Datei

Im letzten Schritt werden die Einzelframes der Deepfake-Werbung wieder mit der originalen Tonspur zusammengefügt. Mit Hilfe des integrierten FFMPEG wird so eine .mp4 Datei erzeugt.

Befehle:

[8) merged to mp4]

Schritt 8: Postproduktion in Adobe AfterEffects

Für die Postproduktion der Deepfakes wurden allerdings die Einzelframes des generierten Gesichts und die erzeugte schwarz-weiß Maske genutzt. So diente die Originaldatei mit den professionellen Schauspielern als Grundlage. Darauf wurde das der Deepfake-Clip gelegt und durch die vorhandene Maske nur das benötigte Gesicht maskiert. Um das Gesicht vom Hintergrund zu trennen wird [Luma] bei [Bewegte Maske] eingestellt. Auf die Maske wird der [Gaußsche Weichzeichner]

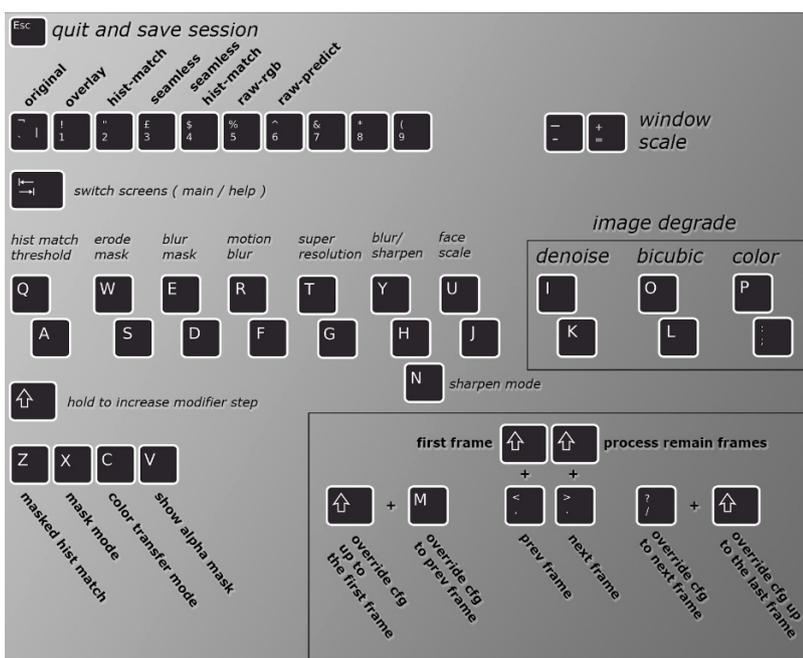


Abbildung 25 Benutzeroberfläche für das Merging in DeepFaceLab

für weiche Maskenkanten, sowie bei Bedarf der Effekt [Matte vergrößern] angewandt. Matte vergrößern bewirkt dabei eine einheitliche Vergrößerung bzw. Verkleinerung der Maske, um z.B. überstehendes Bildmaterial, welches durch den Gaußschen Weichzeichner entstanden ist zu reduzieren. Ebenso wird der [Kanäle tauschen] Effekt auf die Maske angewandt, da die Alpha Maske keine Informationen enthält und für noch bessere Kanteneffekte Informationen von einem der anderen RGB-Kanäle übertragen bekommt.

Problematisch war bei den beiden Stepstone-Deepfakes die Tatsache, dass einzelne Gesichtsframes auch nach wiederholten Versuchen nicht durch DeepFaceLab erzeugt wurden und das Originalgesicht der Schauspielerin im Deepfake zu sehen war. Diese Einzelframes

wurden durch existierende Gesichtsframes gestreckt und manuell, sowie durch Tracking angepasst, weshalb bei genauem Hinsehen verwackelte Sequenzen in den Deepfakes „Paulina“ und „Pia“ erkennbar sind. Beim Deepfake von „Alex“ wurde ebenso der Name „Bernd Stromberg“ als Bauchbinde in Alexander Schier getauscht, um Verwirrungen bei der Umfrage entgegenzuwirken. Im Deepfake von „Micha“ gab es Schwierigkeiten bei der Farbanpassung, da der Original-Clip bereits stark in der Farbe modifiziert ist und erneute Farbanpassungen zu nur noch unnatürlicheren Ergebnissen führten.

In letzter Instanz wurde bei den Deepfakes Farbe und Beleuchtung angepasst, sowie der Text für die Deklaration als Deepfake eingefügt. Folgende Beispiele zeigen den Vorher-Nachher-Effekt.

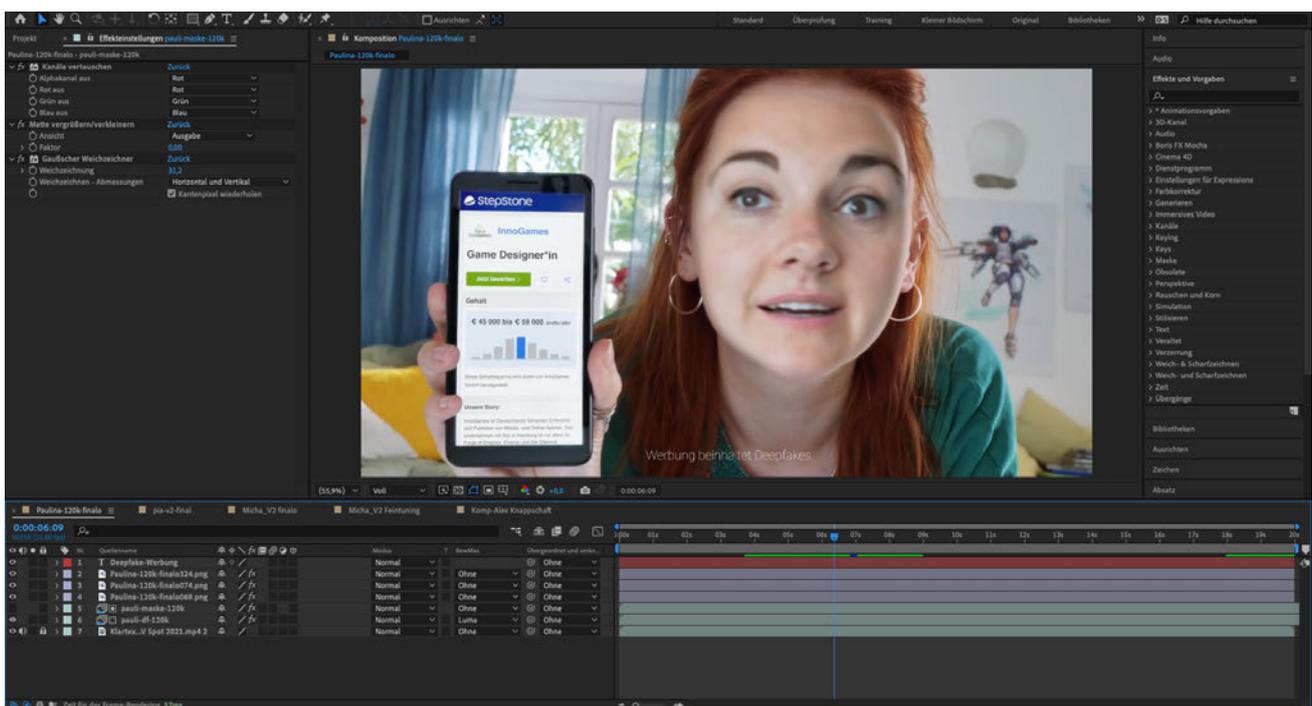


Abbildung 26 Postproduktion in Adobe AfterEffects

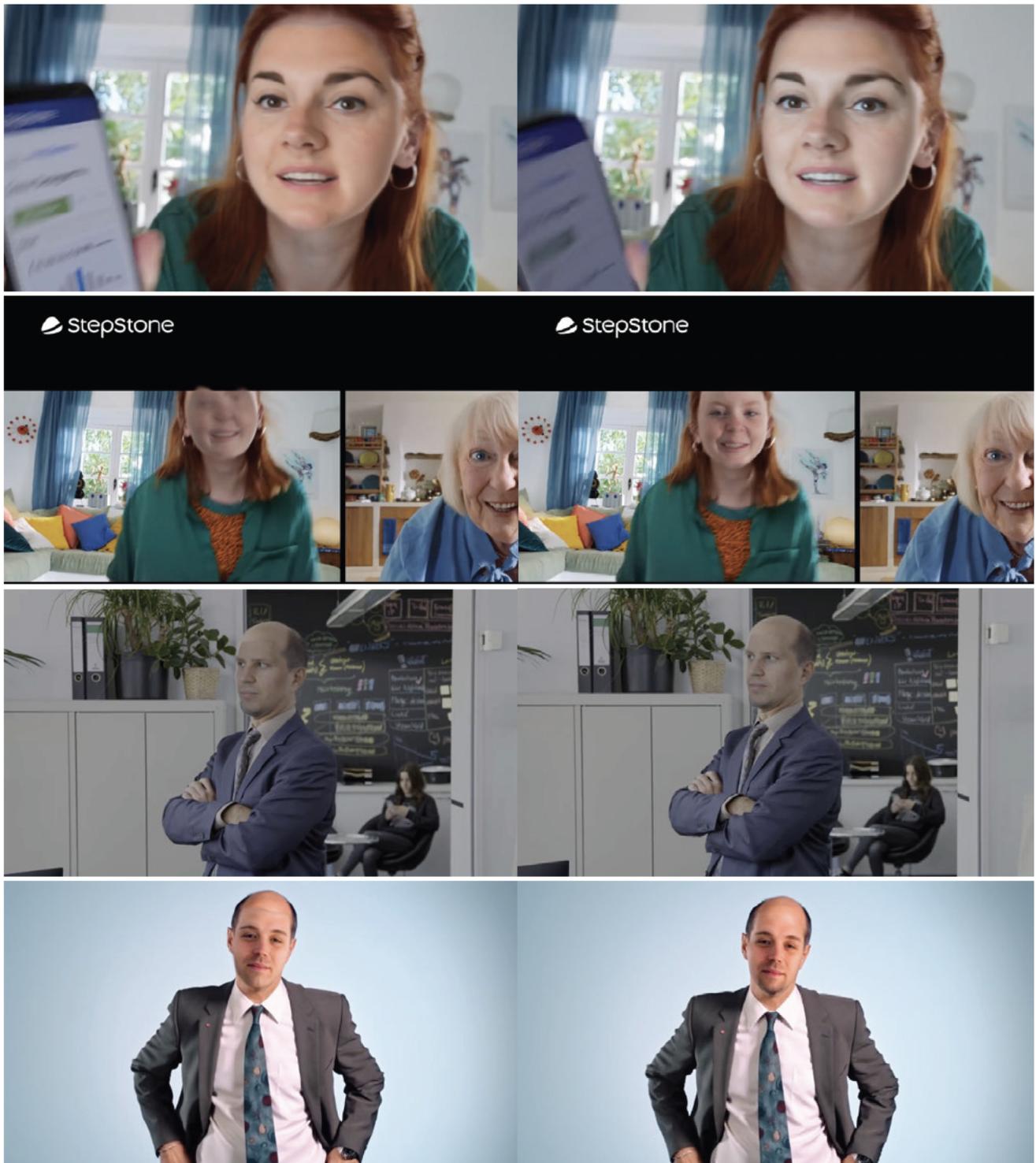


Abbildung 27 Postproduktion Vorher-Nachher-Vergleich (links: Rohdaten aus DFL, rechts: nach Postproduktion)

5.3 Visualisierung der Deepfakes

Mit diesem Vorgehen sind innerhalb 214,5h reiner Trainingszeit in DeepFaceLab die folgenden vier Deepfakes entstanden. Zur Visualisierung werden Einzelframes herangezogen. Der digitale Anhang beinhaltet alle Videodateien und Zeitraffer der Trainings.

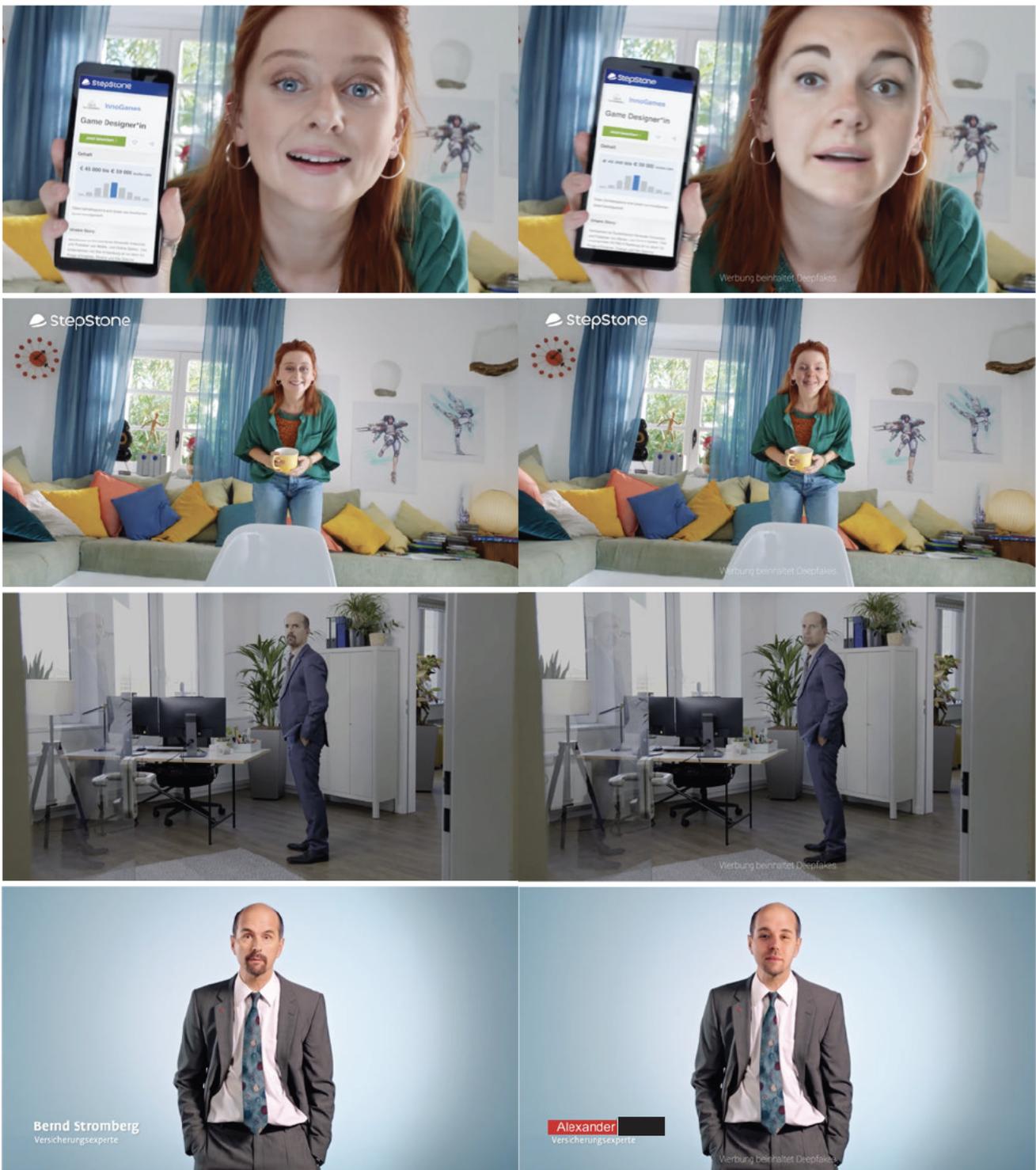


Abbildung 28 Impressionen der Deepfakes (links = Original, rechts = Deepfake)

5.4 Umfrage

Für die Untersuchung möglicher Reaktionen auf diese Deepfakes wird eine Umfrage durchgeführt. Die Umfrage basiert auf den vier selbst erstellten Deepfakes und zwei high-end Deepfakes, die von professionellen Produzenten generiert wurden.

5.4.1 Vorgehen

Die Online-Umfrage ist als Mixed-Methods-Ansatz angelegt und mischt den qualitativen und quantitativen Forschungsansatz. Für die Umfrage wird ein selbst definiertes „Buddy-System“ genutzt. Der Proband, von dem der Deepfake angefertigt wurde, wird als „Person of Interest“ (kurz: POI) gehandelt. Um, die Effekte bezüglich der Hyperpersonalisierung erforschen zu können, werden Probanden aus dem privaten Umfeld (Buddys) der POI für die Umfrage benötigt. Die Probanden sehen den Deepfake der jeweiligen Person of Interest, sowie die beiden high-end Deepfakes und müssen im Anschluss Fragen beantworten, ihre Einstellung sowie Gedanken zu bestimmten Aspekten äußern. Die anonyme Online-Umfrage wird genutzt, um Reproduzierbarkeit, Messgenauigkeit und Neutralität zu gewährleisten.

Für die quantitative Komponente des Forschungsdesigns – der Erhebung statistischer Daten – müssen die folgenden drei Gütekriterien sichergestellt werden (Döring & Bortz, 2016, S. 93ff):

1. Validität: (Wird tatsächlich das gemessen, was gemessen werden soll?)
2. Reliabilität: (Erzielt die Forschung

bei Messwiederholung ähnliche Ergebnisse?)

3. Objektivität: (Verhält sich der Forscher neutral und objektiv während der Untersuchung?)

Die geplante Datenerhebung richtet sich ebenfalls nach den Gütekriterien für qualitative Forschung von Lincoln & Guba (Döring & Bortz, 2016, S. 109f), da Wert auf individuelle und subjektive Meinungsäußerung der kleinen Stichprobe gelegt wird. Dabei gilt es folgende Qualitätskriterien sicherzustellen:

1. Vertrauenswürdigkeit (Sind die Ergebnisse der Studie auf Basis der Daten vertrauenswürdig?)
2. Übertragbarkeit (Sind die Ergebnisse der Studie auf andere Kontexte übertragbar?)
3. Zuverlässigkeit (Wurde der Forschungsprozess nachvollziehbar durchlaufen und dokumentiert?)
4. Bestätigbarkeit (Sind die Ergebnisse durch den Forschenden neutral erhoben worden?)

Die Primärdaten werden durch eine Online-Umfrage erhoben. Aufgrund der heterogenen Verteilung der Probanden erleichtert die ständige Verfügbarkeit und eigene Entscheidung über den Zeitpunkt der Teilnahme das Procedere. Darüber hinaus kann auch eine Anonymität gewährleistet werden, was bestenfalls zu ehrlicheren Antworten im Vergleich zur persönlichen Befragungen führt (Döring & Bortz, 2016, S. 257). In einem Pretest mit 3 Personen wurde der Fragebogen auf Fehler untersucht. Die Umfrage wurde durch das kostenpflichtige Umfragetool Unipark realisiert und war im Zeitraum vom 01. Juni 2022 bis 08. Juni 2022 aktiv.

5.4.2 Stichprobenbeschreibung

Für die Online-Umfrage werden vier „Person of Interest“ kurz POI für die Deepfake-Produktion benötigt. Die POI haben schriftlich zugestimmt Videomaterial für die Deepfakes zur Verfügung zu stellen. Die Einverständniserklärungen befinden sich im digitalen Anhang. Für jeden Deepfake der POI werden 5 „Buddys“ für die Umfrage herangezogen. Diese Buddys werden von den POI ausgewählt und sollen höchstens 30 Jahre alt sein, um in die definierte Altersgrenze der Zielgruppendefinition von Deepfake-Marketing zu passen. Insgesamt wird so eine Stichprobengröße von 22 Probanden erzielt. Die Umfrage ist aufgrund der zu geringen Stichprobengröße somit nicht repräsentativ. Die Ergebnisse der Umfrage liefern nur Tendenzen und müssen für eine Verallgemeinerung der Aussagen mit einer größeren Stichprobe überprüft werden.

5.4.3 Fragenkatalog

Der genaue Fragenkatalog der Umfrage ist im Anhang (Seite 98) einzusehen. Folglich soll nur der grobe Ablauf skizziert werden.

Nach einführenden Worten, der Auswahl der POI und einer Definition zu personalisierten Inhalten wird die persönliche Einstellung zu hyperpersonalisierter Werbung im Allgemeinen ermittelt. Nach einer Definition zu Deepfakes soll auch hier die Einstellung gegenüber Deepfakes ermittelt werden. Im Anschluss wird der Deepfake der POI gezeigt. Dazu sollen dann Fragen zu aktuellen Gedanken und Erkennungsaspekten beantwortet werden. Hiernach werden die professionellen

Deepfakes mit Morgan Freeman und Tom Cruise gezeigt. Die Teilnehmer werden gefragt, welche Videos Deepfakes waren. Diese Frage soll Aufschluss über die Erkennungsfähigkeit von high-end Deepfakes geben. Nach der Bewertung von diversen Aussagen zur Authentizität, Deklaration und Inhalten der Deepfakes werden zum Abschluss demografische Informationen zum Geschlecht, dem Alter und höchstem Schulabschluss ermittelt.

5.5 Deskriptive Auswertung

Die Auswertung der Ergebnisse kann aufgrund der geringen Stichprobe nur deskriptiv erfolgen. Die Auswertung erfolgt durch das kostenpflichtige Statistikprogramm *IBM SPSS Statistics*.

Im digitalen Anhang sind die kompletten Tabellen und Daten der Umfrage aufgeführt. Für die Auswertung der Umfrage wurde eine ungerade Likert Skala herangezogen. Dabei geben die Teilnehmer auf einer mehrstufigen Skala an, wie sehr sie der Aussage zustimmen. Die Skala ist dabei wie folgt definiert: **stimme nicht zu – stimme eher nicht zu – weder noch – stimme eher zu – stimme zu**. Mit der Aussage „weder noch“ kann zum Ausdruck gebracht werden, dass die Teilnehmer der Aussage neutral gegenüberstehen und sich nicht für eine Ausprägung entscheiden müssen. In der Auswertung werden die zwei Ausprägungen für „stimme zu“ und „stimme nicht zu“ für vereinfachte Interpretation als Zustimmung oder Ablehnung gewertet. Die Kontrollfrage wurde lediglich durch eine Person falsch beantwortet. Da jedoch von mehreren Personen während der Umfrage

Nachfragen bezüglich der Kontrollfrage ausgingen, wurde diese eine Person nicht herausgefiltert.

Von den 22 Umfrage-Teilnehmern sind 9 Personen männlich und 13 weiblich. Sie sind zwischen 20 und 31 Jahren alt und spiegeln so die definierte Zielgruppe wider. Der Altersdurchschnitt liegt bei 26 Jahren. Eine Person gibt als höchsten Schulabschluss, den Realschulabschluss an, zwei Personen besitzen das Abitur und sieben eine abgeschlossene Berufsausbildung. Der Großteil der Befragten hat bereits einen akademischen Abschluss (54,5%).

Die Aufteilung der Fragebögen auf die einzelnen Buddys ist ungefähr ausgeglichen. So haben je fünf Personen die Fragebögen für Pia, Alexander und Micha ausgefüllt und sieben Personen für Paulina.

- **Die Einstellung der Probanden zur Personalisierung**

Jeder Proband hat bereits personalisierte Werbung wahrgenommen und weiß, was Personalisierung bedeutet. So empfinden die Personen Werbung als nervig (68,1%) und haben Angst, dass Unternehmen ihre Daten missbräuchlich nutzen (63,7%). Frauen (84,6%) finden Werbung stärker störend als Männer (44,4%). Im Gegensatz dazu lassen sich jedoch 72,7% der Probanden von Werbung in ihren Kaufentscheidungen beeinflussen. 63,7% der Teilnehmer teilen ihre persönlichen Daten nicht im Internet, um sich Vorteile zu verschaffen und würden für einzigartige Vorteile auch nicht zusätzliche Daten preisgeben (63,7%). Gleichzeitig sind die Probanden der Meinung, dass Unternehmen die eigenen Daten nicht besser nutzen

müssen (45,5%), jedoch zu viel über ihre Kunden wissen (54,6%). Fast 40% der Teilnehmer stehen dieser Aussage neutral gegenüber. Ein Ausbau der Aktivitäten für bessere personalisierte Angebote ist aus Kundensicht somit derzeit nicht nötig.

- **Die Einstellung der Probanden zu Deepfakes**

Trotz der in der Umfrage vorangestellten Definition von Deepfakes geben 31,8% an nicht zu wissen, was Deepfakes sind. Fast die Hälfte der Befragten hat bereits Deepfakes vor der Umfrage gesehen. Je 6 Personen wissen es nicht oder haben noch nie einen Deepfake gesehen.

Der Großteil (77,3%) empfindet Deepfakes als Gefahr. Fast alle (91%) Befragten sind sich einig, dass man in Zukunft noch stärker auf die Glaubhaftigkeit im Internet achten muss (Stichwort: Fake News!). Passend dazu behaupten fast 70% der Teilnehmer, Deepfakes nicht auf den ersten Blick zu erkennen. Zustimmung erhält dabei jedoch der zeitnahe Einsatz von Deepfakes im Fernsehen oder den sozialen Medien (77,3%). Uneinigkeit herrscht bei der Frage, ob Deepfakes die Werbung für den Einzelnen noch attraktiver gestalten können. 40,9% der Teilnehmer stimmen für „nein“ und eine ebenso hohe Teilnehmerzahl (31,8%) ist sich unschlüssig. Wiederum finden über die Hälfte der Probanden (54,6%), dass Deepfakes die Personalisierung von Werbung noch präziser gestalten kann. Hierbei entsteht eine Diskrepanz zwischen Personalisierung im Allgemeinen und Personalisierung für den Einzelnen.

Nachdem die Probanden die Deepfakes der Person of Interest gesehen haben, zeichnete sich folgende Erkennungsreihenfolge ab:

1. Unechte Mundbewegungen (63,6%)
2. Unnatürliche Gesichtsübergänge (59,1%)
3. Unnatürliche Augen (45,5%)
4. Kennzeichnung (27,3%)
5. Flackern des Bildes (13,6%)

Daraus lässt sich schließen, dass diese Elemente in der Deepfake-Produktion besser gelöst werden müssten. Auffällig ist hier, dass die Probanden den Deepfake nur zu 27% an der eigentlichen Deklaration erkannt haben. Interessant wäre an dieser Stelle die Einbeziehung einer zusätzlichen Eyetracking-Analyse, um nachzuvollziehen, auf welche Bereiche des Deepfake-Videos besonders geachtet wurde.

Als andere Erkennungsgründe des Deepfakes wurde angegeben:

- die Werbung war bekannt,
- das Video lief nicht flüssig,
- Gesichtsfarbe/-form,
- der Mund ist verschwommen,
- und die Stimme ist falsch.

Bei der Analyse der drei offenen Antwortboxen zu den Deepfakes konnten detaillierte Meinungen gesammelt werden. Da die Probanden unterschiedliche Deepfakes sahen, wurde die Auswertung im Anhang zusätzlich nach Unterscheidung der vier Personen aufgeführt. Im Folgenden wird nur eine vereinfachte Auswahl der Aussagen aufgeführt.

- **Was denkst Du über die technische Umsetzung des Deepfakes?**

„Wahrscheinlich hätte ich gedacht, die Person existiert wirklich, wenn ich vorher nicht gewusst hätte, wer das ist.“
(Proband für den Deepfake von Micha)

Ein oft genannter Punkt waren optische Mängel. So wurde ein verschwindendes Kinn, unförmiger Kussmund oder nicht übereinstimmender Hautton bemängelt. Ebenso ist die Spiegelung von Stromberg zu sehen und die Stimmen sind falsch. Man erkennt ebenso deutlich das ersetzte Gesicht, so die Probanden. Ein Teilnehmer kritisiert die mäßige Synchronisation bei der jüngeren Person [Paulina]. Die Teilnehmer sind erschrocken, was technisch bereits jetzt möglich ist und geben an, dass die Deepfakes ohne Vorkenntnisse oft nicht zu erkennen sind. Ein Teilnehmer wagt eine Aufwandsprognose und schätzt die Deepfake-Produktion als aufwändig zu programmieren ein. Ein weiterer Teilnehmer kann die technische Umsetzung nicht beurteilen. Insgesamt reicht das subjektive Stimmungsbild von täuschend echt, erstaunlich gut, hervorragend, genial gelungen realistisch bis hin zu ausbaufähig, mittelmäßig und nicht sehr echt.

- **Was denkst Du über den Einsatz von Deepfakes in den Medien?**

„Führt zu noch mehr Confirmation Bias und Misstrauen gegenüber Aussagen und Meinungen.“

(Proband für den Deepfake von Paulina)

Als Confirmation Bias oder Bestätigungs-Irrtum bezeichnet man in der Psychologie die Bestätigung der eigenen Erwartungen, indem man Informationen gezielt sucht und so interpretiert, dass die eigene Erwartung rechtfertigt (Stangl, 2022). Als Teil der selektiven Wahrnehmung erliegt man dabei einem Selbstbetrug. Der Gefahrenaspekt hat einen hohen Stellenwert bei den Teilnehmern. Die Möglichkeiten der Täuschung beunruhigen und wirken angsteinflößend auf die Teilnehmer. Es wird schwieriger Fake News zu entlarven, da man damit so tagtäglich konfrontiert wäre. Die Teilnehmer sind besorgt, dass sie selbst Opfer von Deepfakes mit negativen Folgen werden könnten und stufen Deepfakes in bestimmten Thematiken/ Kreisen als gefährlich ein. Ein Teilnehmer findet Deepfakes in den Medien nur für Propagandazwecke sinnvoll. Ebenso sind sich die Teilnehmer sicher, dass Deepfakes irgendwann nicht mehr wegzudenken sind. Für Unternehmen sehen die Probanden große Vorteile. So könne man mit Deepfakes in den Medien Kunden auf ein Produkt aufmerksamer machen. Es ist sinnvoll für den Aufbau einer persönlichen Ebene und bringt Nähe zum Kunden. Die Teilnehmer gehen davon aus, dass ein besserer Verkauf erzielt wird, wenn ein Produkt durch Deepfakes mit Freunden vermarktet wird. Obendrein schätzen die Teilnehmer Deepfake-Produktion als kosten- und

ressourcensparend ein. Insgesamt reicht das subjektive Stimmungsbild von witzig, gut für Social Media, spannend bis hin zu unnötig, erschreckend und gefährlich.

- **Was denkst Du über die Personalisierung von Werbung durch Deepfakes?**

„Die Personalisierung von Werbung kann damit auf ein neues Level gehoben werden.“
(Proband für den Deepfake von Paulina)

Die Teilnehmer gehen davon aus, dass die Werbung nachhaltiger im Gedächtnis bleibt, wenn man das Gesicht des Werbenden kennt und somit auch zu einer stärkeren Identifikation mit der Werbung führen. Eine Person findet, personalisierte Werbung mit Deepfakes suggeriert einen größeren persönlichen Bezug zu den Produkten und verstärkt somit psychologisch manipulierend den Wunsch zu noch mehr Konsum. Ein anderer Teilnehmer erwähnt die höhere Kontrolle über Inhalte auf der Unternehmensseite. Und je nach Zielgruppe kann der Deepfake andere Inhalte vermitteln. Für Kunden sei es außerdem einfacher Vertrauen in Produkte zu fassen, da man Freunden vertraut. So wird die Kaufaktivität durch bekannte Gesichter angeregt. Insgesamt reicht auch hier das subjektive Stimmungsbild von irreführend, macht mir Angst, gefährlich, überflüssig bis hin zu spannend und witzig. Diese zwiespaltenen Meinungsbilder bestätigen sich auch im Rest der Umfrage.

- **Waren die Videos Deepfakes?**

Die Beurteilung der high-end Deepfakes von Morgan Freeman (Youtube (7), 2021) und Tom Cruise (Youtube (8), 2021) geht weit auseinander. Nur sechs der Teilnehmer erkannten richtig, dass beide Videos Deepfakes sind. Dabei lagen 5 Frauen und nur ein Mann richtig. Dies geht einher mit der Beobachtung von Köbis et al. (2021), der eine Überschätzung der eigenen Fähigkeiten in Bezug auf die Erkennung von Deepfakes erkannte. Erkennen die Befragten Deepfakes nicht sofort, so sind sie in erster Instanz überrascht vom Fortschritt der Technik (84,4%) und eher weniger misstrauisch (54,5%).

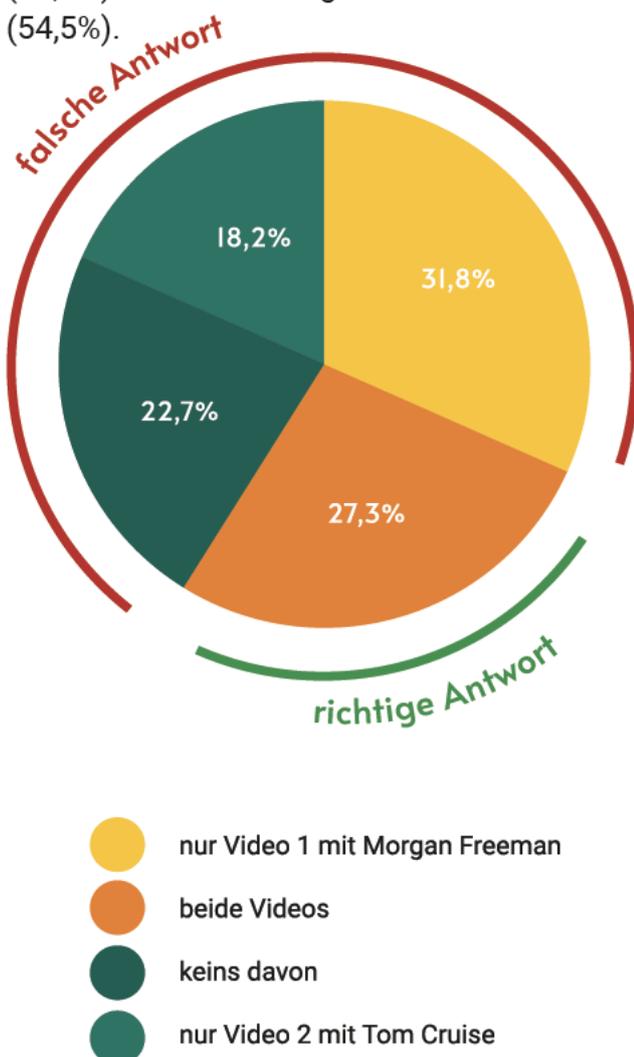


Abbildung 29 Antwortverteilung zur Frage „Waren die Videos Deepfakes?“

- **Die Einstellung der Probanden zur Authentizität von Deepfake-Werbung**

Die Authentizität der Werbung spielt für die Probanden eine große Rolle. So finden 72,7% der Teilnehmer Werbung mit einem hohem Realitätsgrad ansprechend. Und ausnahmslos alle Teilnehmer (stimme zu: 72,7%, stimme eher zu: 27,3%) sprechen der Glaubwürdigkeit einen hohen Stellenwert für die Werbung im Allgemeinen zu. Fehlerhafte Werbung wirkt für die Probanden unprofessionell (77,2%) und insbesondere fehlerhafte Deepfake-Werbung überzeugt sie nicht (54,5%). Unprofessionalität und fehlende Überzeugung können Auslöser für Zweifel und Misstrauen sein, was die Skepsis gegenüber Deepfakes sogar steigern könnte. Dadurch, dass die Probanden glaubwürdige Werbung wichtig und ansprechend empfinden, kann *Hypothese H2 „Je authentischer der Deepfake ist, desto niedriger ist die Gefahr von Skepsis gegenüber der Werbemaßnahme.“* für die Stichprobe dieser Umfrage als wahr eingestuft werden.

- **Die Einstellung der Probanden zur Deklaration von Deepfake-Werbung**

Die Notwendigkeit der offenen Deklaration von Videomaterial als Deepfake wurde von den Teilnehmern bestätigt. So würden sich die Teilnehmer nicht getäuscht fühlen, wenn Deepfake-Werbung als Deepfake gekennzeichnet wäre (72,7%). Ebenso bedeutet eine Kennzeichnung die Steigerung des Vertrauens in das Unternehmen (68,2%). Den Befragten fiel es leichter diese Form von Werbung zu akzeptieren, wenn sie dementsprechend gekennzeichnet ist (72,8%).

Folglich lässt sich **Hypothese 1 „Die offensichtliche Deklaration als Deepfake wirkt sich positiv auf die Akzeptanz hyperpersonalisierter Werbung aus“** bestätigen. Die offensichtliche Deklaration von Werbung als Deepfake-Werbung steigert das Vertrauen und erhöht die Akzeptanz dieser Werbeform. Auffällig ist außerdem, dass es dem Großteil der Personen (11 von 16 Probanden), die den Deepfake nicht an der Kennzeichnung erkannt haben leichter fallen würde, die Deepfake-Werbung zu akzeptieren.

- **Die Einstellung der Probanden zu Deepfakes von Familie, Freunden und Bekannten**

Im Bereich der Deepfake-Werbung mit Familie, Freunden und Bekannten sind sich die Befragten uneinig. So finden zwar 45,4% der Teilnehmer, dass Deepfake-Werbung mit Freunden und Familie das eigene Interesse am beworbenen Inhalt steigert – fast genauso viele (40,9%) lehnen die Aussagen jedoch auch ab. Der gleiche Effekt zeigt sich auch bei der Aussage: „Wofür Freunde/Familie werben kann nur gut sein, deshalb fällt meine Meinung über das Unternehmen positiv aus.“. Die Hälfte aller Teilnehmer verneint die Aussage und für 40,9% der Befragten trifft die Aussage zu. Zugleich sehen die Teilnehmer die authentische Wirkung von Deepfake-Werbung mit ihnen bekannten Personen zweigeteilt: genau 50% lehnen ab, 50% wollen sich nicht einordnen oder stimmen zu.

Aufgrund der Uneinigkeit werden für die Überprüfung der **Hypothese 3a „Deepfakes mit Material aus dem privaten Umfeld haben einen positiven Effekt auf die Akzeptanz hyperpersonalisierter Werbung.“** die

Antworten aus den offenen Antwortboxen herangezogen. Die Antworten der Teilnehmer im Stichpunkt „Was denkst Du über die Personalisierung von Werbung durch Deepfakes?“ suggerieren, dass die Werbung nachhaltiger im Gedächtnis bleibt, wenn man das Gesicht des Werbenden kennt und somit auch zu einer stärkeren Identifikation mit der Werbung führen. Weiterhin sei es einfacher Vertrauen in Produkte zu fassen, wenn man die Personen, die dafür werben kennt. Die Teilnehmer sprechen von „gibt der Werbung eine persönliche Note“, „die Personalisierung [...] kann damit auf ein neues Level gehoben werden“ oder „ist die Zukunft“. So lässt sich **Hypothese 3a** durch die offenen Antworten besser einschätzen und kann für diese Stichprobe als wahr eingestuft werden.

- **Die Einstellung der Probanden zu Deepfakes mit dem eigenen Gesicht**

Der Trend der Uneinigkeit ist auch bei der Meinung über Deepfakes mit dem eigenen Gesicht zu erkennen. So würden zwar 40,9% der Befragten ihre persönlichen Daten für die Produktion von Deepfakes zur Verfügung stellen, wenn sie an der Entscheidung beteiligt werden, für welche Projekte das eigene Gesicht genutzt werden – der Großteil (59,1%) jedoch nicht. Nimmt man eine Einteilung in die Innovationsakzeptanztypen nach Rogers (2003, S. 248ff) vor, so zählen ganze 40,9% der Befragten zu den unternehmungslustigen Vorantreibern und/oder Meinungsführern – „Innovatoren“ und „Early Adopters“.

Die Teilnehmer haben Angst, dass Deepfakes mit dem eigenen Gesicht persönlichen Schaden anrichten können (68,2%). Die Angst ist besonders groß bei Personen, die angaben, Deepfakes zu kennen. Auf dieser Grundlage dieses Stimmungsbildes erfolgt die Verifizierung der **Hypothese 3b „Deepfake-Werbung mit dem eigenen Gesicht schürt Bedenken zum Datenschutz und missbräuchlicher Nutzung.“**. Deepfake-Werbung mit dem eigenen Gesicht steigert die Skepsis gegenüber Deepfakes nicht. Ebenso sehen alle Teilnehmer Potential für Deepfakes in der Werbung. Deepfakes mit dem eigenen Gesicht können Aufmerksamkeit erregen (54,5%) und zum Teilen z.B. in sozialen Medien einladen (54,6%).

Insgesamt ließen sich mit dieser Umfrage detaillierte, neue Erkenntnisse zur Hyperpersonalisierung durch Deepfake-Inhalte generieren. Es zeichneten sich zweigeteilte Lager der Befürworter und Gegner von Deepfakes im Marketing ab. Da das Deepfake-Marketing jedoch noch in den Kinderschuhen steckt, muss die Forschung für generalisierbare Ergebnisse diesbezüglich noch weiter ausgebaut werden.



6. Fazit

In dieser Masterarbeit wird der Einsatz von Deepfakes im Marketing analysiert. Es geht darum, Deepfake-Marketing zu analysieren und ein eigenes Deepfake-Modell zu generieren, um Effekte bei Rezipienten festzustellen. Dafür wird zunächst die Personalisierung von Werbung, sowie Deepfakes im Allgemeinen thematisiert. Im Anschluss daran werden die Vor- und Nachteile der Hyperpersonalisierung durch die Integration von Deepfakes in die Werbung analysiert und vier eigene Deepfakes auf Basis realer Video-Werbungen generiert. Zur Überprüfung von Effekten bei Rezipienten wird eine Umfrage erarbeitet, bei der die Probanden die vier selbst produzierten Deepfakes sehen und Fragen zu ihrer persönlichen Einstellung zur Personalisierung durch Deepfakes mitteilen.

Mit Hilfe der Umfrage soll eine Antwort auf die zentrale Forschungsfrage „*Welche Effekte lösen hyperpersonalisierte Werbebotschaften durch Deepfakes bei Probanden aus?*“ gefunden werden. So lassen sich die Auswirkungen bei hyperpersonalisierten Werbebotschaften durch Deepfakes primär in die folgenden zwei Kategorien einordnen: **positive Überraschung** und **Misstrauen**. Aus der Datenerhebung geht im Detail hervor, dass die Teilnehmer von Werbung genervt sind. Es besteht die Angst, dass persönliche Daten missbräuchlich genutzt werden. Zeitgleich lassen sich die Befragten durch Werbung in ihren Kaufentscheidungen beeinflussen, was für eine Wirkung der Werbeangebote spricht. Beim Thema Deepfakes gehen die Meinungen auseinander. Sie sind gefährlich und wirken durch die Möglichkeiten der Täuschung

beängstigend. Aus den offenen Antworten geht sogar die Prognose eines Teilnehmers hervor: „*Irgendwann wird es [Deepfakes] nicht mehr wegzudenken sein*“. Auf der anderen Seite kann die Personalisierung von Werbung durch Deepfakes noch präziser gestaltet werden. Ein Proband spricht von: „*Die Personalisierung von Werbung kann damit [Deepfakes] auf ein neues Level gehoben werden*“. Nur authentische, fehlerfreie Deepfake-Werbung wirkt auf die Teilnehmer ansprechend. Beim Erkennen von Deepfakes besteht allerdings noch Verbesserungsbedarf. Eine offene Deklaration als Deepfake erzeugt Vertrauen, sowie Akzeptanz bei dieser Werbeform und baut Skepsis ab. Deepfake-Werbung mit Freunden bleibt bei den Probanden länger im Kopf und gestaltet den Vertrauensaufbau zu Produkten einfacher, wenn man die Person im Video kennt. Hierbei lassen sich Parallelen zum Influencer-Marketing erkennen. Schlussendlich würden bereits 4 von 10 Personen ihre Daten für die Produktion von Deepfake-Werbung zur Verfügung stellen.

Durch die Auswertung der nicht repräsentativen Online-Umfrage lassen sich neue Erkenntnisse zur Eignung von Deepfakes in der Werbebranche herausstellen. Die Ergebnisse knüpfen an die aktuelle Forschung zum Thema Deepfakes im Marketing (Campbell et al., 2021; Kietzmann et al., 2021; Whittaker et al., 2021) an und bestätigen Teilaspekte und Anregungen deren Forschung.

Im Hinblick auf den erleichterten Vertrauensaufbau bei Werbung von privat bekannten Personen, kann es sinnvoll sein im Bereich Produktvermarktung mit Deepfake-Werbung weiterzuforschen und in erster Instanz eine umfangreichere Datenerhebung anzugehen. Außerdem müsste für eine erfolgreiche Integration die gesetzliche Grundlage für das Deepfake-Marketing präzise definiert werden. Hinzu kommt die Verbesserung der Erkennungsmechanismen bei Deepfakes, wenn diese für negative Zwecke missbraucht werden, sowie die bessere Auswertung und Nutzung der Kundendaten für eine echte, datengesteuerte Hyperpersonalisierung auf Unternehmensseite.

Ob Deepfake-Marketing die hier geweckten Erwartungen erfüllt, in Zukunft weiter viral gehen wird, schlicht Neugier auf Unbekanntes erweckt oder nur zu schwerwiegenden Problemen führt bleibt abzuwarten. Ob Deepfake-Marketing sich im Online-Marketing als eigene Disziplin etablieren wird bleibt ebenso abzuwarten, denn der Erfolg von hyperpersonalisierter Werbung liegt nicht zuletzt in der subjektiven Wahrnehmung jeder einzelnen Person.

Literaturverzeichnis

- Accenture. (2018). *Making it Personal: Why brands must move from communication to conversation for greater personalization*. Pulse Check 2018. Accenture Interactive. https://www.accenture.com/t20161011T222718_w_/us-en/_acnmedia/PDF-34/Accenture-Pulse-Check-Dive-Key-Findings-Personalized-Experiences.pdf. Zugriff am: 23.03.2022
- Adobe. (2020). *Experience Index: Digitale Trends 2020*. https://business.adobe.com/content/dam/dx/de/de/resources/reports/digital-trends-2020/digital-trends-2020-full-report_de.pdf. Zugriff am: 07.03.2022
- Adobe, & Econsultancy. (2020, März). *Wofür setzt Ihr Unternehmen KI derzeit im Hinblick auf die Automatisierung bestimmter Aktivitäten im Bereich Marketing ein?* Statista. <https://de-statista-com.ezproxy.th-brandenburg.de/statistik/daten/studie/984521/umfrage/einsatzbereiche-von-ki-im-marketing-weltweit/>. Zugriff am: 07.03.2022
- Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). *MesoNet: A Compact Facial Video Forgery Detection Network*. 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 1–7. <https://doi.org/10.1109/WIFS.2018.8630761>
- Agarwal, S., El-Gaaly, T., Farid, H., & Lim, S.-N. (2020). *Detecting Deep-Fake Videos from Appearance and Behavior*. arXiv:2004.14491 [cs, eess]. <http://arxiv.org/abs/2004.14491>
- Agarwal, S., Farid, H., Fried, O., & Agrawala, M. (2020). *Detecting Deep-Fake Videos from Phoneme-Viseme Mismatches*. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2814–2822. <https://doi.org/10.1109/CVPRW50498.2020.00338>
- Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2019). *Protecting World Leaders Against Deep Fakes*. 38–45. https://openaccess.thecvf.com/content_CVPRW_2019/papers/Media%20Forensics/Agarwal_Protecting_World_Leaders_Against_Deep_Fakes_CVPRW_2019_paper.pdf. Zugriff am: 09.03.2022
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice-Hall.
- Albahar, M., & Almalki, J. (2019). *DEEPPAKES: THREATS AND COUNTERMEASURES SYSTEMATIC REVIEW*. *Journal of Theoretical and Applied Information Technology*, 97(22), 3242–3250.
- Allcott, H., & Gentzkow, M. (2017). *Social Media and Fake News in the 2016 Election*. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Amerini, I., Galteri, L., Caldelli, R., & Del Bimbo, A. (2019). *Deepfake Video Detection through Optical Flow Based CNN*. 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), 1205–1207. <https://doi.org/10.1109/ICCVW.2019.00152>
- Anand, A., Chakraborty, T., & Park, N. (2019). *We used Neural Networks to Detect Clickbaits: You won't believe what happened Next!* arXiv:1612.01340 [cs]. <http://arxiv.org/abs/1612.01340>
- Antova, G., Dumont, A., & Scheffler, M. (2021). *1.3 Sicherheitskultur*. *Digitale Welt*, 5(1), 50–55. <https://doi.org/10.1007/s42354-020-0321-5>
- Appel, M. (Hrsg.). (2020). *Die Psychologie des Postfaktischen: Über Fake News, „Lügenpresse“, Clickbait & Co*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-58695-2>

- Beisch, N., Egger, A., & Schäfer, C. (2021). *Bewegtbildmarkt in Bewegung: Videonutzung habitualisiert sich in mittlerer Altersgruppe. Ergebnisse der ARD/ZDF-Massenkommunikation Trends und der ARD/ZDF-Onlinestudie 2021*. https://www.ard-zdf-onlinestudie.de/files/2021/Beisch_Egger_Schaefer.pdf. Zugriff am: 14.02.2022
- Belch, G. E., & Belch, M. A. (2018). *Advertising and promotion: An integrated marketing communications perspective (11th edition)*. McGraw-Hill Education.
- Bendel, O. (2019, Juli 25). *Deepfake Definition*. Gabler Wirtschaftslexikon. <https://wirtschaftslexikon.gabler.de/definition/deepfake-120960>. Zugriff am: 28.01.2022
- Bendel, O. (2021). *Gamification Definition*. Gabler Wirtschaftslexikon. <https://wirtschaftslexikon.gabler.de/definition/gamification-53874/version-384606>. Zugriff am: 28.01.2022
- Benrath, B., & Theile, G. (2020, Januar 7). *Facebook geht gegen manipulierte Videos vor*. Frankfurter Allgemeine. <https://www.faz.net/aktuell/wirtschaft/digitec/deepfakes-facebook-geht-gegen-manipulierte-videos-vor-16570055.html>. Zugriff am: 30.03.2022
- Bernecker, Prof. Dr. M. (2021, August 20). *Was ist Online-Marketing? – Online-Marketing-Definition*. Deutsches Institut für Marketing. <https://www.marketinginstitut.biz/blog/was-ist-online-marketing-definition/>. Zugriff am: 28.01.2022
- Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P., & Nayar, S. K. (2008). *Face swapping: Automatically replacing faces in photographs*. ACM SIGGRAPH 2008 Papers on - SIGGRAPH '08, 1. <https://doi.org/10.1145/1399504.1360638>
- Boermann, S. C., van Noort, G., Helberger, N., & Hoofnagle, C. J. (2018). *Sponsored Blog Content: What do the Regulations Say? And what do Bloggers Say?* JIPITEC, 9(2018), 146–159.
- Brock, A., Donahue, J., & Simonyan, K. (2019). *Large Scale GAN Training for High Fidelity Natural Image Synthesis*. <http://arxiv.org/abs/1809.11096>
- BSI. (2022). *Deepfakes – Gefahren und Gegenmaßnahmen*. https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html. Zugriff am: 26.03.2022
- Buo, S. A. (2020). *The Emerging Threats of Deepfake Attacks and Countermeasures*. arXiv:2012.07989 [cs]. <https://doi.org/10.13140/RG.2.2.23089.81762>
- Bursic, S., D'Amelio, A., Granato, M., Grossi, G., & Lanzarotti, R. (2021). *A Quantitative Evaluation Framework of Video De-Identification Methods*. 2020 25th International Conference on Pattern Recognition (ICPR), 6089–6095. <https://doi.org/10.1109/ICPR48806.2021.9412186>
- Butcher, M. (2019, April 25). *The startup behind that deep-fake David Beckham video just raised \$3M*. TechCrunch.Com. <https://tcrn.ch/2XMT8bL>. Zugriff am: 14.03.2022
- Buxmann, P., & Schmidt, H. (Hrsg.). (2019). *Künstliche Intelligenz: Mit Algorithmen zum wirtschaftlichen Erfolg*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-57568-0>
- BuzzFeedVideo. (2018, April 17). *You Won't Believe What Obama Says In This Video!* youtube.de. <https://www.youtube.com/watch?v=cQ54GDm1eL0&t=29s>. Zugriff am: 11.03.2022
- California Assembly Bill. (2019, Oktober 3). *AB 730 – Elections: Deceptive audio or visual media*. openstates. <https://openstates.org/ca/bills/20192020/AB730/>. Zugriff am: 30.03.2022
- Campbell, C., Plangger, K., Sands, S., & Kietzmann, J. (2021). *Preparing for an Era of Deepfakes and AI-Generated Ads: A Framework for Understanding Responses to Manipulated Advertising*. Journal of Advertising, 1–17. <https://doi.org/10.1080/00913367.2021.1909515>

- Capgemini Consulting, & ESSEC. (2016). Hyper-personalization vs. Segmentation: Has Big Data made customer segmentation redundant? https://www.capgemini.com/consulting-fr/wp-content/uploads/sites/31/2017/08/hyperpersonnalisation_vs_segmentation_english_05-01-2017.pdf. Zugriff am: 23.03.2022
- Carrel, D. (2022). Digitale Trends 2022. <https://business.adobe.com/de/resources/digital-trends-report.html#download-digital-trend-2022>. Zugriff am: 31.01.2022
- Chen, Y., Conroy, N. J., & Rubin, V. L. (2015). Misleading Online Content: Recognizing Clickbait as „False News“. Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection, 15–19. <https://doi.org/10.1145/2823465.2823467>
- Chesney, R., & Citron, D. (2019, Januar). Deepfakes and the New Disinformation War. The Coming Age of Post-Truth Geopolitics. *Foreign Affairs*, 98:147.
- Chitrakorn, K. (2021, Januar 11). How deepfakes could change fashion advertising. *Vogue Business*. https://www.voguebusiness.com/companies/how-deepfakes-could-change-fashion-advertising-influencer-marketing?fbclid=IwAR0zCLaH4e79UPB4DyzdpHvI9zP5VQ8OnIOc2yqqv_D26SyUGz2mxwjJ614. Zugriff am: 29.03.2022
- Choi, Y., Uh, Y., Yoo, J., & Ha, J.-W. (2020). StarGAN v2: Diverse Image Synthesis for Multiple Domains. arXiv:1912.01865 [cs]. <http://arxiv.org/abs/1912.01865>
- Cinecom.net. (2019, Dezember 10). DEEPFAKE Tutorial: A Beginners Guide (using DeepFace Lab). youtube.de. <https://www.youtube.com/watch?v=t59gRbpYMiY>. Zugriff am: 11.03.2022
- Coelho, P. S., & Henseler, J. (2012). Creating customer loyalty through service customization. *European Journal of Marketing*, 46(3/4), 331–356. <https://doi.org/10.1108/03090561211202503>
- Coker, K. K., Flight, R. L., & Baima, D. M. (2017). SKIP IT OR VIEW IT: THE ROLE OF VIDEO STORYTELLING IN SOCIAL MEDIA MARKETING. *The Marketing Management Journal*, 27(2), 75–87.
- Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, 35(1), 53–65. <https://doi.org/10.1109/MSP.2017.2765202>
- Daly, R. (2019, Juli 17). Meet Lil Miquela, the real life Ashley O. Nme.Com. <https://www.nme.com/features/interview-lil-miquela-real-life-ashley-o-2530049>. Zugriff am: 12.05.2022
- Davenport, T. H., Harris, J. G., & Kohli, A. K. (2001). How do they know their customers so well? *MIT Sloan Management Review*, 42(2).
- Davies, V. (2021, Dezember 6). Deepfakes to become a growing trend in 2022 says IntSights. cybermagazine. <https://cybermagazine.com/cyber-security/deepfakes-become-growing-trend-2022-says-intsights>. Zugriff am: 03.06.2022
- deeptomcruise. (2022, Mai 12). Deeptomcruise. TikTok. <https://www.tiktok.com/@deeptomcruise>. Zugriff am: 12.05.2022
- Deges, F. (2020). *Grundlagen des E-Commerce: Strategien, Modelle, Instrumente*. Springer Gabler. <https://doi.org/10.1007/978-3-658-26320-1>
- Der Tagesspiegel. (2016, September 21). Angela Merkel im Wortlaut „Wenn wir nicht gerade aus Stein sind“. [tagesspiegel.de. https://www.tagesspiegel.de/politik/angela-merkel-im-wortlaut-wenn-wir-nicht-gerade-aus-stein-sind/14576252.html](https://www.tagesspiegel.de/politik/angela-merkel-im-wortlaut-wenn-wir-nicht-gerade-aus-stein-sind/14576252.html). Zugriff am: 16.02.2022

- Deutsche Post. (2019, September 24). Dmexco 2019: Helmut Briggel – Das Print Mailing hat eine Renaissance erlebt. Youtube. <https://youtu.be/6s4wQ5izHEI>. Zugriff am: 07.03.2022
- Deutschland sicher im Netz. (2021). DsiN SicherheitsIndex 2021: Digitale Sicherheitslage von Verbraucher:innen in Deutschland. <https://www.sicher-im-netz.de/file/13161/download?token=se3us1Mq>. Zugriff am: 26.03.2022
- DiPietro, R., & Hager, G. D. (2020). Deep learning: RNNs and LSTM. In *Handbook of Medical Image Computing and Computer Assisted Intervention* (S. 503–519). Elsevier. <https://doi.org/10.1016/B978-0-12-816176-0.00026-0>
- DIVSI. (2016, Juni 20). DIVSI Internet-Milieus 2016: Die digitalisierte Gesellschaft in Bewegung. <https://www.divsi.de/publikationen/studien/divsi-internet-milieus-2016-die-digitalisierte-gesellschaft-bewegung/4-die-divsi-internet-milieus-2016/4-1-entwicklung-der-divsi-internet-milieus/index.html>
- Dobber, T., Metoui, N., Trilling, D., Helberger, N., & de Vreese, C. (2021). Do (Microtargeted) Deepfakes Have Real Effects on Political Attitudes? *The International Journal of Press/Politics*, 26(1), 69–91. <https://doi.org/10.1177/1940161220944364>
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-41089-5>
- dotCMS. (2018). *Hyper-Personalization – It’s Time to Truly Meet Your Customers*. <https://cdn2.hubspot.net/hubfs/2389934/eBooks/Hyper-Personalization-whitepaper-dotCMS.pdf>. Zugriff am: 23.03.2022
- Drees, C. (2020, Juli 1). Deepfakes: Disney stößt das Tor nach Hollywood auf. <https://www.mobilegeeks.de/news/deepfakes-disney-stoesst-das-tor-nach-hollywood-auf/>. Zugriff am: 23.03.2022
- Duden (a). (2022, Januar 12). Individualisieren. <https://www.duden.de/node/147013/revision/534236>. Zugriff am: 28.01.2022
- Duden (b). (2022, Januar 12). Personalisieren. <https://www.duden.de/node/155848/revision/642799>. Zugriff am: 28.01.2022
- Duden (c). (2022, Januar 14). Fake News. <https://www.duden.de/node/44637/revision/506546>. Zugriff am: 14.02.2022
- FaceSwap. (2022, März 11). FaceSwap. FaceSwap. <https://faceswap.dev>. Zugriff am: 11.03.2022
- Fan, H., & Poole, M. S. (2006). What Is Personalization? Perspectives on the Design and Implementation of Personalization in Information Systems. *Journal of Organizational Computing and Electronic Commerce*, 16(3–4), 179–202. <https://doi.org/10.1080/10919392.2006.9681199>
- Fei, J., Xia, Z., Yu, P., & Xiao, F. (2021). Exposing AI-generated videos with motion magnification. *Multimedia Tools and Applications*, 80(20), 30789–30802. <https://doi.org/10.1007/s11042-020-09147-3>
- Fish, T. (2019, April 4). Deep fakes: AI-manipulated media will be ‘WEAPONISED’ to trick military. *EXPRESS*. <https://www.express.co.uk/news/science/1109783/deep-fakes-ai-artificial-intelligence-photos-video-weaponised-china>. Zugriff am: 15.03.2022

- Floridi, L. (2018). *Artificial Intelligence, Deepfakes and a Future of Ectypes*. *Philosophy & Technology*, 31(3), 317–321. <https://doi.org/10.1007/s13347-018-0325-3>
- Fontana, L. (2018, August 10). *VFX: Kino aus dem Computer*. *Galaxus*. <https://www.galaxus.de/de/page/vfx-kino-aus-dem-computer-8422>. Zugriff am: 30.03.2022
- Gentsch, P. (2019). *Künstliche Intelligenz für Sales, Marketing und Service: Mit AI und Bots zu einem Algorithmic Business – Konzepte und Best Practices*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-25376-9>
- Gesellschaft für deutsche Sprache e. V. (2016, Dezember 9). *Pressemitteilung: GfdS wählt »postfaktisch« zum Wort des Jahres 2016*. <https://gfds.de/wort-des-jahres-2016/>. Zugriff am: 16.02.2022
- GitHub DeepFaceLab. (2022, März 11). *DeepFaceLab*. *DeepFaceLab*. <https://github.com/iperov/DeepFaceLab>. Zugriff am: 11.03.2022
- GitHub FaceSwap. (2022, März 11). *Deepfakes/faceswap*. *GitHub*. <https://github.com/deepfakes/faceswap>. Zugriff am: 11.03.2022
- Glynn, F. (2015, April 16). *Why it Takes 6-8 Marketing Touches To Generate a Viable Sales Lead*. *salesforce.com*. <https://www.salesforce.com/blog/takes-6-8-touches-generate-viable-sales-lead-heres-why-gp/>. Zugriff am: 01.02.2022
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative Adversarial Networks*. *arXiv:1406.2661 [cs, stat]*. <http://arxiv.org/abs/1406.2661>
- Gouthier, M. H. J., & Kern, N. (2021). *Hyperpersonalisierung – Hochpersonalisierte Kundenansprache durch den Einsatz Künstlicher Intelligenz*. In M. Bruhn & K. Hadwich (Hrsg.), *Künstliche Intelligenz im Dienstleistungsmanagement* (S. 127–154). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-34326-2_5
- Graf, D. (2007). *Die Theorie des geplanten Verhaltens*. In D. Krüger & H. Vogt (Hrsg.), *Theorien in der biomedizinischen Forschung* (S. 33–43). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-68166-3_4
- Ham, C.-D. (2017). *Exploring how consumers cope with online behavioral advertising*. *International Journal of Advertising*, 36(4), 632–658. <https://doi.org/10.1080/02650487.2016.1239878>
- Hancock, J. T., & Bailenson, J. N. (2021). *The Social Impact of Deepfakes*. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 149–152. <https://doi.org/10.1089/cyber.2021.29208.jth>
- Hao, K. (2018, November 17). *What is machine learning?* *MIT Technology Review*. <https://www.technologyreview.com/2018/11/17/103781/what-is-machine-learning-we-drew-you-another-flowchart/>. Zugriff am: 05.03.2022
- Haridy, R. (2021, August 16). *Deepfake tech puts you in the trailer of new Hugh Jackman movie*. <https://newatlas.com/computers/deepfake-tech-trailer-hugh-jackman-movie-reminiscence/>. Zugriff am: 07.02.2022

- Harwell, D. (2018, Dezember 31). *Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image'*. Washington Post. <https://www.washingtonpost.com/technology/2018/12/31/scarlett-johansson-fake-ai-generated-sex-videos-nothing-can-stop-someone-cutting-pasting-my-image/>. Zugriff am: 15.03.2022
- Hauser, A. (2018, November 22). *DEEPFAKES ANALYSIS - AMOUNT OF IMAGES, LIGHTING AND ANGLES*. <https://www.scip.ch/en/?labs.20181122>. Zugriff am: 30.05.2022
- Heinemann, G. (2017). *Personalisierung im Onlinehandel: Bestandsaufnahme und Ausblick*. In E. Stüber & K. Hudetz (Hrsg.), *Praxis der Personalisierung im Handel: Mit zeitgemäßen E-Commerce-Konzepten Umsatz und Kundenwert steigern* (S. 1–23). Springer Gabler. <https://doi.org/10.1007/978-3-658-16244-3>
- Hendricks, V. F., & Vestergaard, M. (2018). *Postfaktisch: Die neue Wirklichkeit in Zeiten von Bullshit, Fake News und Verschwörungstheorien* (T. Borchert, Übers.; 1. Auflage). Blessing.
- Hippner, H., & Wilde, K. D. (Hrsg.). (2006). *Grundlagen des CRM: Konzepte und Gestaltung* (2., überarb. und erw. Aufl., Nachdr. Juli 2007). Gabler.
- Hour One. (2022). *Hour One AI*. Hour One. <https://www.hourone.ai>. Zugriff am: 07.03.2022
- Hsu, C.-C., Zhuang, Y.-X., & Lee, C.-Y. (2020). *Deep Fake Image Detection Based on Pairwise Learning*. *Applied Sciences*, 10(1), 370. <https://doi.org/10.3390/app10010370>
- Initiative D21 e.V. (2021). *D21 Digital Index 2020/2021 (Jährliches Lagebild zur Digitalen Gesellschaft)*. https://initiatived21.de/app/uploads/2021/02/d21-digital-index-2020_2021.pdf. Zugriff am: 31.01.2022
- Jaster, R., & Lanius, D. (2019). *Die Wahrheit schafft sich ab: Wie Fake News Politik machen*. Reclam.
- Joecks-Laß, D. (2020). *Kundenintegration im Zeitalter der Digitalisierung – Eine Dienstleistungstypologie automatisierter Personalisierung zur Ableitung von Kundenrollen*. In M. Bruhn & K. Hadwich (Hrsg.), *Automatisierung und Personalisierung von Dienstleistungen: Konzepte – Kundeninteraktionen – Geschäftsmodelle* (Bd. 1, S. 73–103). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-30166-8_3
- Kamps, I., & Schetter, D. (2018). *Performance Marketing: Der Wegweiser zu einem mess- und steuerbaren Marketing – Einführung in Instrumente, Methoden und Technik*. Springer Fachmedien. <https://doi.org/10.1007/978-3-658-18453-7>
- Kantar Millward Brown. (2017). *Make A Lasting Impression*. https://www.millwardbrown.ru/library/Kantar_MillwardBrown_Make_a_Meaningful_Impression.pdf. Zugriff am: 03.04.2022
- Karnouskos, S. (2020). *Artificial Intelligence in Digital Media: The Era of Deepfakes*. *IEEE Transactions on Technology and Society*, 1(3), 138–147. <https://doi.org/10.1109/TTS.2020.3001312>
- Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). *Progressive Growing of GANs for Improved Quality, Stability, and Variation*. arXiv:1710.10196 [cs, stat]. <http://arxiv.org/abs/1710.10196>
- Karras, T., Laine, S., & Aila, T. (2019). *A Style-Based Generator Architecture for Generative Adversarial Networks*. ArXiv:1812.04948 [Cs, Stat]. <http://arxiv.org/abs/1812.04948>

- Kerner, T. (2021, März 24). *Hyperpersonalisierung verbessert die Customer Experience im Mittelstand*. *The Future of Commerce*. <https://www.the-future-of-commerce.com/2021/03/24/hyperpersonalisierung-mittelstand-customer-experience/>. Zugriff am: 23.03.2022
- Keuper, F., Schomann, M., & Sikora, L. I. (Hrsg.). (2018). *Homo Connectus*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-19133-7>
- Khalil, H. A., & Maged, S. A. (2021). *Deepfakes Creation and Detection Using Deep Learning*. 2021 *International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 1–4. <https://doi.org/10.1109/MIUCC52538.2021.9447642>
- Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2020). *Deepfakes: Trick or treat?* *Business Horizons*, 63(2), 135–146. <https://doi.org/10.1016/j.bushor.2019.11.006>
- Kietzmann, J., Mills, A. J., & Plangger, K. (2021). *Deepfakes: Perspectives on the future "reality" of advertising and branding*. *International Journal of Advertising*, 40(3), 473–485. <https://doi.org/10.1080/02650487.2020.1834211>
- Köbis, N. C., Doležalová, B., & Soraperra, I. (2021). *Fooled twice: People cannot detect deepfakes but think they can*. *IScience*, 24(11), 103364. <https://doi.org/10.1016/j.isci.2021.103364>
- Kohlstedt, J. (2022, Mai 13). *30 JAHRE GZSZ: JO GERNER KNUTSCHT BEIM JUBILÄUM FREMD UND DIE FANS FLIPPEN AUS*. *TAG24*. <https://www.tag24.de/unterhaltung/tv/gzsz/30-jahre-gzsz-jo-gerner-knutscht-beim-jubilaeum-fremd-und-die-fans-flippen-aus-2454780>. Zugriff am: 03.06.2022
- Kollmann, T. (1998). *Akzeptanz innovativer Nutzungsgüter und -systeme: Konsequenzen für die Einführung von Telekommunikations- und Multimediasystemen (Bd. 239)*. Springer Fachmedien.
- Kollmann, T. (2019). *E-Business: Grundlagen elektronischer Geschäftsprozesse in der Digitalen Wirtschaft (7., überarbeitete und erweiterte Auflage)*. Springer Gabler. <https://doi.org/10.1007/978-3-658-26143-6>
- Kolowich Cox, L. (2017, Juni 28). *16 Video Marketing Statistics to Inform Your 2020 Strategy [Infographic]*. *Hubspot.Com*. <https://blog.hubspot.com/marketing/video-marketing-statistics#sm.0000f7ujhkwse8sqa62aq63w23fi>. Zugriff am: 14.02.2022
- Koopman, M., Macarulla Rodriguez, A., & Geradts, Z. (2018, August). *Detection of Deepfake Video Manipulation*. *Proceedings of the 20th Irish Machine Vision and Image Processing conference, Belfast*. https://www.researchgate.net/publication/329814168_Detection_of_Deepfake_Video_Manipulation. Zugriff am: 22.02.2022
- Kötke, J.-T. (2021). *Deepfake -Eine kurze Einleitung*. https://www.academia.edu/45125668/DEEPFAKE_EINE_KURZE_EINLEITUNG_Deepfake_Eine_kurze_Einleitung. Zugriff am: 24.02.2022
- Kreutzer (a), R. T. (2021). *Praxisorientiertes Online-Marketing: Konzepte – Instrumente – Checklisten*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-31990-8>
- Kreutzer, R. T. (2021). *Online-Marketing*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-33937-1>

- Kreutzer, R. T., & Sirrenberg, M. (2019). *Künstliche Intelligenz verstehen: Grundlagen – Use-Cases – unternehmenseigene KI-Journey*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-25561-9>
- Kröger, J., & Marx, S. (2020). *Agile Marketing: Marketing in dynamischen Zeiten: Mindset – Methoden – Tools*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-29548-6>
- Kruse Brandão, T., & Wolfram, G. (2018). *Digital Connection: Die bessere Customer Journey mit smarten Technologien – Strategie und Praxisbeispiele*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-18759-0>
- Langhart, N. (2021, Mai 17). *Deepfakes: Wie die AI-Technologie eindruckliche Marketing Massnahmen ermöglicht*. Das Schweizer Marketing Portal. <https://marketing.ch/deepfakes-wie-die-ai-technologie-eindrueckliche-marketing-massnahmen-ermoeglicht/>. Zugriff am: 29.03.2022
- Lebo, T. (2019). *Data-Driven Personalization: What It Is and Why You Need to Use It in Your 2020 Marketing*. <https://research.ascend2.com/2019-data-driven-personalization/>. Zugriff am: 28.01.2022
- Lee, E.-J., & Park, J. K. (2009). *Online service personalization for apparel shopping*. *Journal of Retailing and Consumer Services*, 16(2), 83–91. <https://doi.org/10.1016/j.jretconser.2008.10.003>
- Li, L., Bao, J., Yang, H., Chen, D., & Wen, F. (2020). *FaceShifter: Towards High Fidelity And Occlusion Aware Face Swapping*. arXiv:1912.13457 [cs]. <http://arxiv.org/abs/1912.13457>
- Li, Y., Chang, M.-C., & Lyu, S. (2018). *In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking*. 2018 IEEE International Workshop on Information Forensics and Security (WIFS), 1–7. <https://doi.org/10.1109/WIFS.2018.8630787>
- Li, Y., & Lyu, S. (2019). *Exposing DeepFake Videos By Detecting Face Warping Artifacts*. arXiv:1811.00656 [cs]. <http://arxiv.org/abs/1811.00656>
- Li, Y., Sun, P., Qi, H., & Lyu, S. (2022). *Toward the Creation and Obstruction of DeepFakes*. In C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, & C. Busch (Hrsg.), *Handbook of Digital Face Manipulation and Detection* (S. 71–96). Springer International Publishing. https://doi.org/10.1007/978-3-030-87664-7_4
- Liu, M.-Y., & Tuzel, O. (2016). *Coupled Generative Adversarial Networks*. <http://arxiv.org/abs/1606.07536>
- Loewenstein, G. (1994). *The psychology of curiosity: A review and reinterpretation*. *Psychological Bulletin*, 116(1), 75–98. <https://doi.org/10.1037/0033-2909.116.1.75>
- Luber, S., & Litzel, N. (2017, April 26). *Definition: Was ist Deep Learning?* Bigdata Insider. <https://www.bigdata-insider.de/was-ist-deep-learning-a-603129/>. Zugriff am: 05.03.2022
- Luber, S., & Litzel, N. (2021, Februar 19). *Definition Was ist ein Generative Adversarial Network (GAN)?* Bigdata Insider. <https://www.bigdata-insider.de/was-ist-ein-generative-adversarial-network-gan-a-999817/>. Zugriff am: 18.02.2022
- Lundqvist, A., Liljander, V., Gummerus, J., & van Riel, A. (2013). *The impact of storytelling on the consumer brand experience: The case of a firm-originated story*. *Journal of Brand Management*, 20(4), 283–297. <https://doi.org/10.1057/bm.2012.15>

- Maras, M.-H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *The International Journal of Evidence & Proof*, 23(3), 255–262. <https://doi.org/10.1177/1365712718807226>
- Mari, A. (2019). *The Rise of Machine Learning in Marketing: Goal, Process, and Benefit of AI-Driven Marketing*. <https://doi.org/10.13140/RG.2.2.16328.16649>
- Markgraf, Prof. Dr. D. (2018, Februar 19). Akzeptanz Definition. *Gabler Wirtschaftslexikon*. <https://wirtschaftslexikon.gabler.de/definition/akzeptanz-26995/version-250658>. Zugriff am: 09.02.2022
- Masood, M., Nawaz, M., Malik, K. M., Javed, A., & Irtaza, A. (2021). *Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward*. arXiv:2103.00484 [cs, eess]. <http://arxiv.org/abs/2103.00484>
- Matern, F., Riess, C., & Stamminger, M. (2019). *Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations*. 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), 83–92. <https://doi.org/10.1109/WACVW.2019.00020>
- Max-Planck-Gesellschaft. (2020, Februar 25). *Wie Deutsche Personalisierung im Internet sehen: Repräsentative Studie über Einstellung zu Künstlicher Intelligenz und Datenerfassung*. <https://www.mpg.de/14497641/0221-bild-134137-pm-2020-february>. Zugriff am: 25.03.2022
- Mayer, F. (2020). *Wie viel wissen Sie wirklich über Clickbait? – 7 überraschende Fakten, von denen Sie so noch nie gehört haben!* In M. Appel (Hrsg.), *Die Psychologie des Postfaktischen: Über Fake News, „Lügenpresse“, Clickbait & Co.* (S. 67–79). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-58695-2_7
- Medienstaatsvertrag. (2020, November 7). *Staatsvertrag zur Modernisierung der Medienordnung in Deutschland (Medienstaatsvertrag) vom 14.-28.04.2020 in Kraft getreten am 07.11.2020*. <https://www.daserste.de/ard/die-ard/Medienstaatsvertrag-100.pdf>. Zugriff am: 02.04.2022
- Meffert, H., Burmann, C., Kirchgeorg, M., & Eisenbeiß, M. (2019). *Marketing: Grundlagen marktorientierter Unternehmensführung Konzepte – Instrumente – Praxisbeispiele*. Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-21196-7>
- Melzer, B. (2022, Februar 10). *Hyperpersonalisierung im Marketing*. Cronos. <https://www.cronos.de/news/hyperpersonalisierung-im-marketing>
- Meßmer, A.-K., Sänglerlaub, A., & Schulz, L. (2021). *„Quelle: Internet“? Digitale Nachrichten- und Informationskompetenzen der deutschen Bevölkerung im Test*. Stiftung Neue Verantwortung.
- Mirsky, Y., & Lee, W. (2020). *The Creation and Detection of Deepfakes: A Survey*. *ACM Computing Surveys*, 54(1), 1–41. <https://doi.org/10.1145/3425780>
- Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
- Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020). *Emotions Don't Lie: An Audio-Visual Deepfake Detection Method Using Affective Cues*. arXiv:2003.06711 [cs]. <http://arxiv.org/abs/2003.06711>

- Moon, M. (2021, August 14). Warner Bros. „Reminiscence“ promo uses deepfake tech to put you in the trailer. *engadget*. https://www.engadget.com/warner-bros-remembrance-promo-deepfake-trailer-052114554.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAJVZLJVwHjUrCeKevlzdAGY4eeSEpPQroOziK0Y7htPTWLLm75iXTKGVTFtLcn7QRpIBfgatOwsNSohOSfUTIZNrKw4QHV1dLo8Jc6nJR5cbSREB7DATNEm9Smwc-mnFmAlaSWTDMQcenfkTdlFPWgg-aXEcdpFeBpJ017HAdIZ. Zugriff am: 02.04.2022
- Müller, U. (2005). *Kundenbindung im E-Commerce: Personalisierung als Instrument des Customer Relationship Marketing* (1. Aufl.). Deutscher Universitätsverlag.
- Murthi, B. P. S., & Sarkar, S. (2003). The Role of the Management Sciences in Research on Personalization. *Management Science*, 49(10), 1344–1362. <https://doi.org/10.1287/mnsc.49.10.1344.17313>
- Naderer, B., Matthes, J., & Schäfer, S. (2021). Effects of disclosing ads on Instagram: The moderating impact of similarity to the influencer. *International Journal of Advertising*, 40(5), 686–707. <https://doi.org/10.1080/02650487.2021.1930939>
- Naruniec, J., Helminger, L., Schroers, C., & Weber, R. M. (2020). High-Resolution Neural Face Swapping for Visual Effects. *Computer Graphics Forum*, 39(4), 173–184. <https://doi.org/10.1111/cgf.14062>
- Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyen, T. T., Pham, Q.-V., & Nguyen, C. M. (2022). Deep Learning for Deepfakes Creation and Detection: A Survey. *ArXiv:1909.11573 [Cs, Eess]*. <http://arxiv.org/abs/1909.11573>
- Nielsen. (2015, September). *Global Trust In Advertising: Winning Strategies For An Evolving Media Landscape*. <https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/global-trust-in-advertising-report-sept-2015-1.pdf>. Zugriff am: 03.04.2022
- Nirkin, Y., Masi, I., Tran, A. T., Hassner, T., & Medioni, G. (2017). On Face Segmentation, Face Swapping, and Face Perception. *arXiv:1704.06729 [cs]*. <http://arxiv.org/abs/1704.06729>
- Perov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Umé, C., Dpfks, M., Facenheim, C. S., RP, L., Jiang, J., Zhang, S., Wu, P., Zhou, B., & Zhang, W. (2021). DeepFaceLab: Integrated, flexible and extensible face-swapping framework. *arXiv:2005.05535 [cs, eess]*. <http://arxiv.org/abs/2005.05535>
- Peteranderl, S. (2019, September 3). Deepfake-App Zao China im Gesichtstausch-Rausch. *SPIEGEL Netzwelt*. <https://www.spiegel.de/netzwelt/apps/deepfake-app-zao-sekundenschnell-aus-einem-selfie-deepfakes-erstellen-a-1284951.html>. Zugriff am: 11.03.2022
- Pressmar, D. B. (1982). Zur Akzeptanz von Computergestützten Planungssystemen. In H. Krallmann (Hrsg.), *Unternehmensplanung und -steuerung in den 80er Jahren*. vol 3. Springer, Berlin. https://doi.org/10.1007/978-3-642-68601-6_17
- Pritzkau, A., & Schade, U. (2021). Vorsicht: Mögliche „Fake News“ – ein technischer Ansatz zur frühen Erkennung. In P. Klimczak & T. Zoglauer (Hrsg.), *Wahrheit und Fake im postfaktisch-digitalen Zeitalter* (S. 115–131). Springer Fachmedien Wiesbaden. DOI: 10.1007/978-3-658-32957-0_6

- Pu, J., Mangaokar, N., Kelly, L., Bhattacharya, P., Sundaram, K., Javed, M., Wang, B., & Viswanath, B. (2021). Deepfake Videos in the Wild: Analysis and Detection. *Proceedings of the Web Conference 2021*, 981–992. <https://doi.org/10.1145/3442381.3449978>
- Pulizzi, J. (2012). The Rise of Storytelling as the New Marketing. *Publishing Research Quarterly*, 28(2), 116–123. <https://doi.org/10.1007/s12109-012-9264-5>
- PWC. (2019). *Ergebnisse: Bevölkerungsbefragung zu personalisierter Werbung. Personalisierte Werbung und E-Privacy 2019*. <https://www.pwc.de/de/technologie-medien-und-telekommunikation/bevoelkerungsbefragung-personalisierte-werbung.pdf>. Zugriff am: 10.02.2022
- PWC (a). (2019, März 31). *Pressemitteilung: Personalisierte Online-Werbung erreicht die Nutzer*. PWC. <https://www.pwc.de/de/pressemitteilungen/2019/personalisierte-online-werbung-erreicht-die-nutzer.html>. Zugriff am: 28.03.2022
- Radford, A., Metz, L., & Chintala, S. (2016). *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. <http://arxiv.org/abs/1511.06434>
- Rahman, A., Islam, Md. M., Moon, M. J., Tasnim, T., Siddique, N., MD. Shahiduzzaman, & Ahmed, S. (2022). A Qualitative Survey on Deep Learning Based Deep fake Video Creation and Detection Method. *Australian Journal of Engineering and Innovative Technology*, 4(1), 13–26. <https://doi.org/10.34104/ajeit.022.013026>
- Reichelt, F. (2021). What makes deepfakes believable? A categorization attempt. <https://doi.org/10.13140/RG.2.2.24357.63204>
- Reinsel, D., Gantz, J., & Rydning, J. (2018). *The digitization of the world*. <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>. Zugriff am: 10.02.2022
- Reisig, D., Runge, F., & Greve, Prof. Dr. G. (2011). Targeting mit InVideo Advertising. In C. Bauer, Prof. Dr. G. Greve, & G. Hopf (Hrsg.), *Online Targeting und Controlling: Grundlagen—Anwendungsfelder—Praxisbeispiele* (1. Aufl, S. 243–260). Gabler.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed). Free Press.
- Sardarizadeh, S., & Robinson, O. (2022, März 8). *Ukraine invasion: False claims the war is a hoax go viral*. BBC. <https://www.bbc.com/news/60589965>. Zugriff am: 15.03.2022
- Sassy Justice. (2020, Dezember 12). *SASSY JUSTICE - Cheyenne 9's Coverage of the Official White House Address*. youtube.de. https://youtu.be/h_jrebvmPlk. Zugriff am: 11.03.2022
- Schlohmann, K. (2012). *Innovatorenorientierte Akzeptanzforschung bei innovativen Medientechnologien*. Gabler Verlag.
- Schneider, B. (2018, Januar 8). *Individualisierung und Personalisierung: Marketing für die Generation Y*. ABTasty. <https://www.abtasty.com/de/blog/individualisierung-personalisierung/>. Zugriff am: 28.01.2022
- Schreiner, M. (2021, Februar 21). *Deepfakes: Wie alles begann – und wohin es führen könnte*. mixed.de. <https://mixed.de/geschichte-der-deepfakes-so-rasant-geht-es-mit-ki-fakes-voran/>. Zugriff am: 22.02.2022
- Schürmann, K. (2019, Mai 9). *Standortbezogenes Marketing – das ist erlaubt*. etailment.de. <https://etailment.de/news/stories/Location-Based-Services-Recht-Smartphone-Marketing-22262>. Zugriff am: 14.02.2022

- Shukla, P. S., & Nigam, Dr. P. V. (2018, April). E-Shopping using Mobile Apps and the Emerging Consumer in the Digital Age of Retail Hyper personalization: An Insight. *Pacific Business Review International*, 10(10), 131–139. http://www.pbr.co.in/2018/2018_month/April/16.pdf. Zugriff am: 23.03.2022
- Simonson, I. (2003). Determinants of Customers' Responses to Customized Offers: Conceptual Framework and Research Propositions. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.405060>
- Sodhi, A. (2018). HYPER-PERSONALIZATION: A MARKETING PRIMER FOR ENHANCED CUSTOMER EXPERIENCE. WNS Global Services. https://wnscom-bucket.s3.amazonaws.com/S3_5/Documents/Articles/PDFFiles/7064/102/Hyper-personalization-A-Marketing-Primer-for-Enhanced-Customer-Experience.pdf. Zugriff am: 23.03.2022
- spiegel.de. (2022, Juni 25). Giffey telefoniert mit falschem Klitschko. *SPIEGEL Politik*. <https://www.spiegel.de/politik/deutschland/vermeintlicher-buergermeister-talk-franziska-giffey-telefoniert-mit-falschem-vitali-klitschko-a-cfb365ad-a335-45d3-9652-949f12be84c4>. Zugriff am: 27.06.2022
- Stangl, W. (2022, Juni 9). Confirmation bias. *Online Lexikon für Psychologie und Pädagogik*. <https://lexikon.stangl.eu/10640/confirmation-bias-bestaetigungsfehler-bestaetigungstendenz>. Zugriff am: 09.06.2022
- Strong, E. K., Jr. (1925). Theories of selling. *Journal of Applied Psychology*, 9(1), 75–86. <https://doi.org/10.1037/h0070123>
- Stüber, E. (2013). Personalisierung im Internethandel. Gabler Verlag. <https://doi.org/10.1007/978-3-8349-4559-4>
- Stupp, C. (2019, August 30). Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case. *The Wall Street Journal*. <https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402>. Zugriff am: 15.03.2022
- Suwajanakorn, S., Seitz, S. M., & Kemelmacher-Shlizerman, I. (2017). Synthesizing Obama: Learning lip sync from audio. *ACM Transactions on Graphics*, 36(4), 1–13. <https://doi.org/10.1145/3072959.3073640>
- syntheticfutures.org. (2022, Mai 12). Synthetic Futures. <https://syntheticfutures.org>. Zugriff am: 12.05.2022
- Temir, E. (2020). Deepfake: New Era in The Age of Disinformation & End of Reliable Journalism. *SELÇUK İLETİŞİM DERGİSİ*, 13(2), 1009–10024. <https://doi.org/10.18094/JOSC.685338>
- Tencé, F., Buche, C., De Loor, P., & Marc, O. (2010). The Challenge of Believability in Video Games: Definitions, Agents Models and Imitation Learning. arXiv:1009.0451 [cs]. <http://arxiv.org/abs/1009.0451>
- The Dali Museum. (2019, Januar 23). dalí lives: Museum brings artist back to life with ai. [thedali.org](https://thedali.org/press-room/dali-lives-museum-brings-artists-back-to-life-with-ai/). <https://thedali.org/press-room/dali-lives-museum-brings-artists-back-to-life-with-ai/>. Zugriff am: 11.03.2022
- Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Niessner, M. (2016). Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2387–2395. <https://doi.org/10.1109/CVPR.2016.262>

- Thies, J., Zollhöfer, M., Theobalt, C., Stamminger, M., & Niessner, M. (2018). *Headon: Real-time reenactment of human portrait videos*. *ACM Transactions on Graphics*, 37(4), 1–13. <https://doi.org/10.1145/3197517.3201350>
- Think with Google. (2018, Oktober). *L'Oréal dare to trust data as they bring the power of customisation to scalable branding campaigns*. Think with Google. <https://www.thinkwithgoogle.com/intl/en-154/marketing-strategies/video/loreal-dare-trust-data-they-bring-power-customisation-scalable-branding-campaigns/>. Zugriff am: 29.03.2022
- Tibber Norge. (2021, Mai 14). *Tibber | Es folgt kein Werbespot zur Bundestagswahl (Merkel Deepfake)*. Youtube. <https://www.youtube.com/watch?v=dLAjWtITpnY>. Zugriff am: 07.03.2022
- TMBDF. (2020, Februar 5). *DeepFaceLab Guide*. DeepFaceLab Guide. <https://mrdeepfakes.com/forums/thread-guide-deepfacelab-2-0-guide>. Zugriff am: 11.03.2022
- Tong, X., Wang, L., Pan, X., & Wang, J. gya. (2020). *An Overview of Deepfake: The Sword of Damocles in AI*. 2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL), 265–273. <https://doi.org/10.1109/CVIDL51233.2020.00-88>
- Tractica. (2016, September). *Prognose zum Umsatz mit Unternehmensanwendungen im Bereich künstliche Intelligenz weltweit von 2016 bis 2025*. Statista. <https://de-statista-com.ezproxy.th-brandenburg.de/statistik/daten/studie/620443/umfrage/umsatz-mit-unternehmensanwendungen-im-bereich-kuenstliche-intelligenz-weltweit/>. Zugriff am: 07.03.2022
- Treder, M., & Eter, N. (2018). *„Deep Learning“ und neuronale Netzwerke in der Augenheilkunde: Einsatzmöglichkeiten im Bereich der optischen Kohärenztomographie*. *Der Ophthalmologe*, 115(9), 714–721. <https://doi.org/10.1007/s00347-018-0706-0>
- Tucker, P. (2019, März 31). *The Newest AI-Enabled Weapon: 'Deep-Faking' Photos of the Earth*. *Defense One*. <https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/>. Zugriff am: 15.03.2022
- Uyenco, B., & Kaplan, D. (2010). *Variations in Video Advertising Impact*. *The ARF 56th Annual Convention + Expo*. <https://docplayer.net/12689254-Variations-in-video-advertising-impact.html>. Zugriff am: 14.02.2022
- Vaccari, C., & Chadwick, A. (2020, Mai 28). *'Deepfakes' are here. These deceptive videos erode trust in all news media*. *Washington Post*. <https://www.washingtonpost.com/politics/2020/05/28/deepfakes-are-here-these-deceptive-videos-erode-trust-all-news-media/>. Zugriff am: 30.03.2022
- Vasisht, A. (2020, April 16). *Einfach erklärt: Wie Personalisierung die Customer Journey optimiert*. *onlinemarketing.de*. <https://onlinemarketing.de/performance-marketing/personalisierung-customer-journey-optimieren-erklaerung>. Zugriff am: 31.01.2022
- Volz, P., & Griep, A. (2020). *Personalisierung im digitalen Content Marketing*. In M. Wesselmann (Hrsg.), *Content gekonnt: Strategie, Organisation, Umsetzung, ROI-Messung und Fallbeispiele aus der Praxis* (S. 142–158). Springer Fachmedien Wiesbaden. <https://doi.org/10.1007/978-3-658-24620-4>

- von Piechowski, N. (2021, Dezember 16). 42 Prozent der Verbraucher:innen enttäuscht: Unternehmen nutzen persönliche Daten nicht ausreichend zur Personalisierung. *onlinemarketing.de*. <https://onlinemarketing.de/cases/studie-verbraucher-enttaeuscht-von-unternehmenfehlende-personalisierung-trust>. Zugriff am: 31.02.2022
- Wardle, C. (2017). Fake news. It's complicated. <https://firstdraftnews.org/articles/fake-news-complicated/>. Zugriff am: 16.02.2022
- Weerawardana, M., & Fernando, T. (2021). Deepfakes Detection Methods: A Literature Survey. 2021 10th International Conference on Information and Automation for Sustainability (ICIAfS), 76–81. <https://doi.org/10.1109/ICIAfS52090.2021.9606067>
- Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*, 9(11), 39–52. <https://doi.org/10.22215/timreview/1282>
- Whittaker, L., Kietzmann, T. C., Kietzmann, J., & Dabirian, A. (2020). "All Around Me Are Synthetic Faces": The Mad World of AI-Generated Media. *IT Professional*, 22(5), 90–99. <https://doi.org/10.1109/MITP.2020.2985492>
- Whittaker, L., Letheren, K., & Mulcahy, R. (2021). The Rise of Deepfakes: A Conceptual Framework and Research Agenda for Marketing. *Australasian Marketing Journal*, 29(3), 204–214. <https://doi.org/10.1177/1839334921999479>
- Wuttke, L. (2022). Personalisierung im Marketing: Definition, Vorteile und Beispiele. *datasolut.com*. <https://datasolut.com/personalisierung/>. Zugriff am: 14.02.2022
- Xinwei, L., Jinlin, G., & Junnan, C. (2021). An Overview of Face Deep Forgery. 2021 International Conference on Computer Engineering and Application (ICCEA), 366–370. <https://doi.org/10.1109/ICCEA53728.2021.00078>
- Yang, X., Li, Y., & Lyu, S. (2019). Exposing Deep Fakes Using Inconsistent Head Poses. ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 8261–8265. <https://doi.org/10.1109/ICASSP.2019.8683164>
- YouGov, & Global Witness. (2021, April 15). Do people really want personalised ads online? *Global Witness*. <https://www.globalwitness.org/en/blog/do-people-really-want-personalised-ads-online/>. Zugriff am: 10.02.2022
- Youtube (1). (2021, März 12). Compilation: Werbungen mit Bernd Stromberg. <https://youtu.be/ljT52pQDfRQ>. Zugriff am: 30.05.2022
- Youtube (2). (2021, Dezember 16). Besser jetzt als später—Stromberg reagiert | Zusammen gegen Corona. <https://youtu.be/DQH2lFn2FqE>. Zugriff am: 30.05.2022
- Youtube (3). (2022, Januar 17). Der Gehaltsplaner—Damit du verdienst, was du verdienst! | StepStone TV Spot 2022. <https://youtu.be/H7bxu2brk4k>. Zugriff am: 30.05.2022
- Youtube (4). (2021, August 16). Klartext bei Geld | StepStone TV Spot 2021. <https://youtu.be/EYm8syuKlZs>. Zugriff am: 30.05.2022
- Youtube (5). (2017, November 23). Malin Steffen stellt den YouTube-Kanal vor | Rote Rosen. <https://youtu.be/LZxyarf1GsM>. Zugriff am: 30.05.2022
- Youtube (6). (2016, Januar 22). TV Spot Knappschaft Krankenkasse „Napoleon“ Reiseschutzimpfung. <https://youtu.be/acTDnGKrKoU>. Zugriff am: 30.05.2022
- Youtube (7). (2021, Juli 7). This is not Morgan Freeman—A Deepfake Singularity. <https://youtu.be/oxXpB9pSETo>. Zugriff am: 30.05.2022

- Youtube (8). (2021, Februar 28). *Very realistic Tom Cruise Deepfake | AI Tom Cruise*. <https://youtu.be/iyiOVUbsPcM>. Zugriff am: 30.05.2022
- Yu, N., Davis, L., & Fritz, M. (2019). *Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints*. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 7555–7565. <https://doi.org/10.1109/ICCV.2019.00765>
- Zakharov, E., Shysheya, A., Burkov, E., & Lempitsky, V. (2019). *Few-Shot Adversarial Learning of Realistic Neural Talking Head Models*. arXiv:1905.08233 [cs]. <http://arxiv.org/abs/1905.08233>
- Zhang, C., Zhao, Y., Huang, Y., Zeng, M., Ni, S., Budagavi, M., & Guo, X. (2021). *FACIAL: Synthesizing Dynamic Talking Face with Implicit Attribute Learning*. arXiv:2108.07938 [cs]. <http://arxiv.org/abs/2108.07938>
- Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). *Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks*. *IEEE Signal Processing Letters*, 23(10), 1499–1503. <https://doi.org/10.1109/LSP.2016.2603342>
- Zhang, T., Deng, L., Zhang, L., & Dang, X. (2020). *Deep Learning in Face Synthesis: A Survey on Deepfakes*. 2020 IEEE 3rd International Conference on Computer and Communication Engineering Technology (CCET), 67–70. <https://doi.org/10.1109/CCET50901.2020.9213159>
- Zhou, X., & Zafarani, R. (2020). *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities*. *ACM Computing Surveys*, 53(5), 1–40. <https://doi.org/10.1145/3395046>

Fotos der Personas:

- Tom: Photo by Andrea Piacquadio from Pexels: <https://www.pexels.com/photo/happy-young-man-using-smartphone-on-street-3799835/>
- Luna: Photo by Ron Lach from Pexels: <https://www.pexels.com/photo/a-screen-of-a-smartphone-woman-smiling-while-cooking-8357681/>



Anhang

Online-Umfrage in Druckversion

Umfrage für die Masterarbeit

Der folgende Fragebogen dient einer Studie im Rahmen meiner Masterarbeit im Studium "Digitale Medien" an der Technischen Hochschule Brandenburg. Mit der Teilnahme an der Umfrage leistest Du einen wertvollen Beitrag zur Erforschung von Deepfakes.

Vielen Dank bereits jetzt für Dein Engagement! Die Umfrage wird etwa **10 Minuten** Deiner Zeit in Anspruch nehmen. Bitte lies Dir alles in Ruhe durch. Deine subjektive Meinung ist hierbei gefragt.

Die Umfrage ist anonym und Deine Daten werden nicht an Dritte weitergegeben.

Wenn Du Fragen zu dieser Studie hast, kannst Du Dich gerne an mich wenden:
Paulina Selent
paulina.selent@th-brandenburg.de

Und jetzt geht es los!

2 Testseite

Bitte wähle zuerst Deine bekannte Person aus.

Bitte wähle die Person aus, von der Du später den Deepfake sehen möchtest.

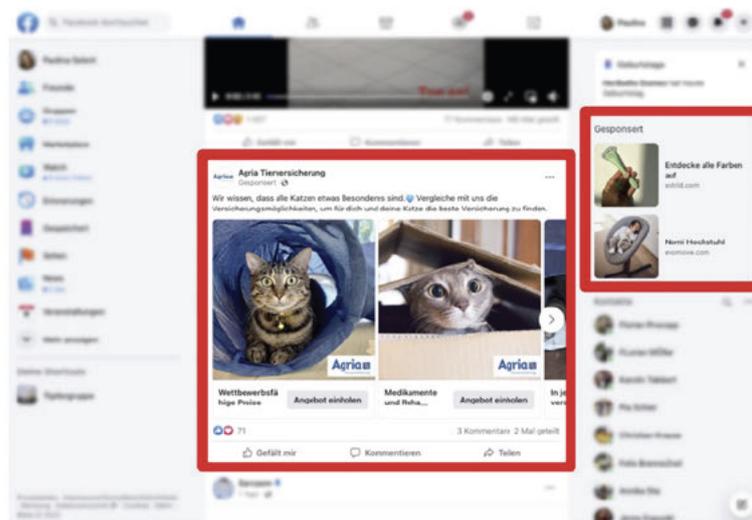
- Paulina
- Pia
- Alexander
- Micha

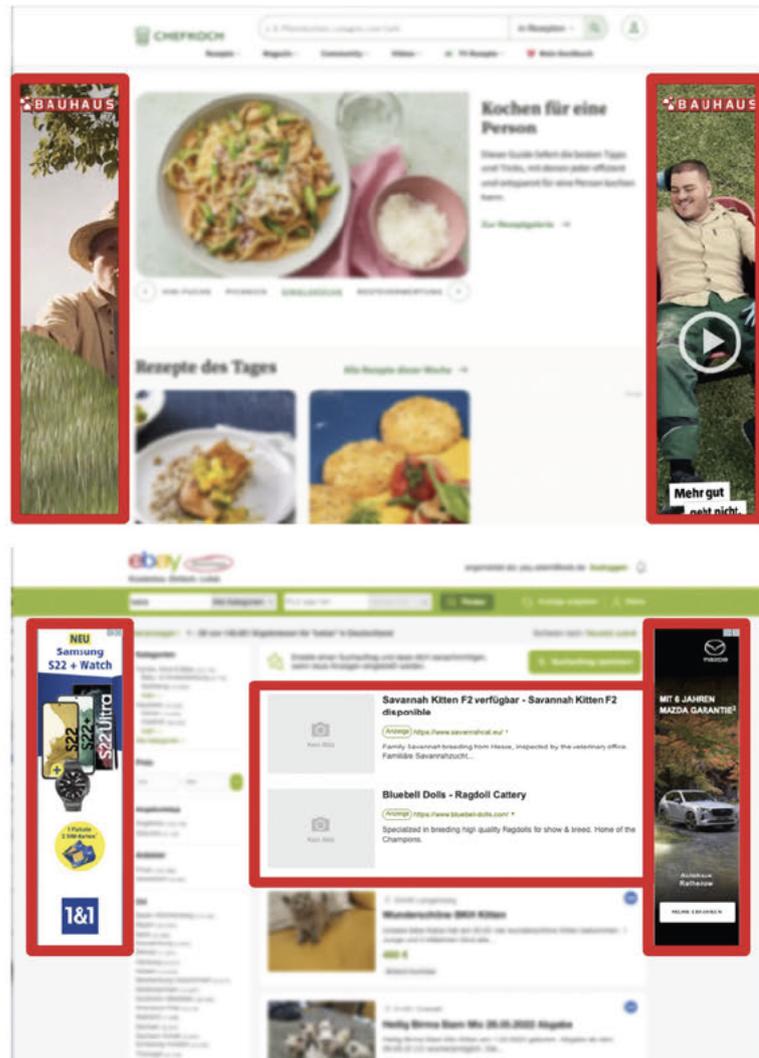
3 Standardseite

Was ist personalisierte Werbung?

Unternehmen sammeln Daten über ihre Kunden, um daraus Empfehlungen abzuleiten oder ihnen besondere Angebote zu bieten. Diese persönlichen Daten werden genutzt, um Werbung auf Deine Interessen zuzuschneiden.

Hier siehst du einige Beispiele, in welcher Form Dir personalisierte Werbung begegnen könnte:





Bitte beantworte nun die folgenden Aussagen zum Thema Personalisierung.

4 Standardseite

Personalisierte Werbung ist mir bereits im Internet aufgefallen.

- ja
- nein
- ich weiß nicht

Deine Einstellung zu personalisierter Werbung:

Bitte lies die Fragen sorgfältig durch. Es gibt kein richtig oder falsch. Es geht um Deine persönliche Meinung.

stimme nicht zu
 stimme eher nicht zu
 weder noch
 stimme eher zu
 stimme zu

Ich finde Werbung nervig.

Ich teile meine persönlichen Daten

im Internet, um mir Vorteile zu verschaffen.

Ich würde mehr Daten preisgeben, wenn ich dafür Vorteile erhalte.

Ich habe Angst, dass Unternehmen meine Daten missbrauchen.

Unternehmen wissen zu viel über ihre Kunden.

Unternehmen müssen meine Daten besser nutzen.

5 Standardseite

Bei Deepfakes werden Gesichter oder Stimmen einer Person mit dem Gesicht oder der Stimme eines anderen digital getauscht. So kann es aussehen, als ob Du etwas in einer bestimmten Umgebung sagst, was so nie im echten Leben passiert ist.

Bitte beantworte nun die folgenden Aussagen zu Deepfakes.

Ich weiß, was ein Deepfake ist.

- ja
 nein

Ich habe vor dieser Umfrage schon mal einen Deepfake gesehen.

- ja
 nein
 weiß nicht

6 Einstellung zu Deepfakes

Deine Einstellung zu Deepfakes

	stimme nicht zu	stimme eher nicht zu	weder noch	stimme eher zu	stimme zu
Ich denke, Deepfakes sind gefährlich.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Mit der Verbreitung von Deepfakes muss stärker aufgepasst werden, was man im Internet glauben kann.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich erkenne Deepfakes auf den ersten Blick.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich denke, Deepfakes können bereits im Fernsehen oder den	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

sozialen Medien eingesetzt werden.

Ich denke, Deepfakes können die Werbung von Unternehmen für mich noch attraktiver machen.

Ich denke, mit Deepfakes kann die Personalisierung von Werbung noch präziser werden.

Ich wähle als Antwortmöglichkeit "stimme eher zu" aus.

7 Deepfake Videos

Du siehst nun das Werbevideo mit Pia.

Bitte schau es Dir aufmerksam an.

Du siehst nun das Werbevideo mit Paulina.

Bitte schau es Dir aufmerksam an.

Du siehst nun das Werbevideo mit Alexander.

Bitte schau es Dir aufmerksam an.

Du siehst nun das Werbevideo mit Micha.

Bitte schau es Dir aufmerksam an.

8 Standardseite

Was denkst Du, nachdem Du den Deepfake gesehen hast, über

Antworte bitte in 2-3 Stichpunkten. Bitte antworte spontan und schreibe auf, was Dir als Erstes einfällt.

die Technische Umsetzung des Deepfakes:

den Einsatz von Deepfakes in den Medien:

die Personalisierung von Werbung durch Deepfakes:

9 DF erkannt an XY

Ich habe den Deepfake an folgenden Punkten erkannt:

- Flackern des Bildes
- Unechte Mundbewegungen
- Unnatürliche Augen
- Unnatürliche Gesichtsübergänge
- Kennzeichnung

Falls Dir der Deepfake anders aufgefallen ist, benenne es bitte:

10 Standardseite

Du siehst jetzt zwei weitere Videos nacheinander.

Bitte schau sie Dir genau an.

11 Standardseite

Waren die Videos Deepfakes?

- nur Video 1 mit Morgan Freeman
- nur Video 2 mit Tom Cruise
- beide Videos
- keins davon

12 Auflösung: beide Deepfakes

Fast richtig! Beide Videos sind Deepfakes.

Richtig! Beide Videos sind Deepfakes.

13 Authentizität

Deine Meinung zu folgenden Aussagen:

	stimme nicht zu	stimme eher nicht zu	weder noch	stimme eher zu	stimme zu
Werbung mit hohem Realitätsgrad spricht mich an.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Glaubwürdigkeit spielt für die Werbung eine große Rolle.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Werbung, bei der man Fehler sieht, wirkt unprofessionell.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deepfake-Werbung, bei der man Fehler sieht, überzeugt mich nicht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich lasse mich nicht von Werbung in meinen Kaufentscheidungen beeinflussen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Erkenne ich einen Deepfake nicht sofort, bin ich
Mehrfachauswahl möglich.

- überrascht von Fortschritt der Technik.

misstrauisch.

14 Deklarierung

Deine Einstellung zu:

	stimme nicht zu	stimme eher nicht zu	weder noch	stimme eher zu	stimme zu
Wird eine Deepfake Werbung als „Deepfake“ angekündigt, fühle ich mich nicht vom Unternehmen getäuscht.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich kann mich darauf einstellen, dass jetzt individuell auf mich abgestimmte Werbung kommt, wenn ich den Zusatz „Deepfake“ dabei sehe.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deepfakes sollten gekennzeichnet sein, da mein Vertrauen in das Unternehmen gestärkt wird.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deepfakes sollten gekennzeichnet sein, da es mir leichter fällt, diese Form von Werbung zu akzeptieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

15 DF Bekannte/eigenes Gesicht

Jetzt denke bitte nur an den Deepfake von Deiner/Deinem Bekannten.

Deepfakes mit dem Gesicht von Freunden und Familie

	stimme nicht zu	stimme eher nicht zu	weder noch	stimme eher zu	stimme zu
Deepfake Werbung mit Freunden/Familie erhöhen mein Interesse am beworbenen Inhalt.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Wofür Freunde/Familie werben kann nur gut sein, deshalb fällt meine Meinung über das Unternehmen positiv aus.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deepfake Werbung mit mir bekannten Personen wirkt authentisch.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

16 DF eigenes Gesicht

Stell Dir vor, Deepfakes werden in Zukunft statt mit dem Gesichts Deiner/Deines Bekannten mit Deinem eigenen Gesicht angefertigt.

Deepfakes mit dem eigenen Gesicht

	stimme nicht zu	stimme eher nicht zu	weder noch	stimme eher zu	stimme zu
Deepfakes mit meinem Gesicht erregen Aufmerksamkeit.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deepfakes mit meinem Gesicht laden zum Teilen z.B. in sozialen Medien ein.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Deepfakes mit meinem Gesicht könnten persönlichen Schaden anrichten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich stelle meine Daten für die Produktion von Deepfakes zur Verfügung, wenn ich mitentscheiden kann für welche Projekte sie genutzt werden.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

17 Informationen zur Person

Fast geschafft! Als letztes werden noch allgemeine Informationen zu Deiner Person benötigt:

Zu welchem Geschlecht fühlst Du Dich zugehörig?

- männlich
- weiblich
- divers

18 Standardseite

Wie alt bist du?

Alter:

19 Schulabschluss

Mein höchster Schulabschluss ist:

- kein Schulabschluss
- Grund-/Hauptschulabschluss
- Realschulabschluss (mittlere Reife)
- Gymnasium (Abitur)

- abgeschlossene Berufsausbildung
- Hochschulabschluss

20 Endseite



Vielen Dank für Deine Meinung und Zeit!

Wenn Du Fragen zu dieser Studie hast, kannst Du Dich gerne an mich wenden:
paulina.selent@th-brandenburg.de
Du kannst die Umfrage jetzt schließen.

Herzlichst,
Paulina Selent

Auflistung der offene Antworten aufgeteilt nach Person-Of-Interest

POI: Paulina

Technische Umsetzung	Deepfakes in den Medien	Personalisierung von Werbung durch Deepfakes
<p>das Kinn verschwindet am Anfang kurz / beim Kuss berühren die Lippen sich nicht / Hautton zwischen Gesicht und Händen stimmt nicht überein / Man kann sehen, dass es ein Deepfake ist / Die Umsetzung ist mittelmäßig. / Man erkennt deutlich das ersetzte Gesicht und die mäßige Synchronisation bei der jüngeren Person. / K.a. / ist je nach verfügbaren technischen Mitteln sehr gut / auf der Basis von Sozialen Medien (Apps) meist aber gut erkennbar / ausbaufähig, aber man versteht wohin die technischen Möglichkeiten sich entwickeln / Sicherlich sehr aufwendig zu programmieren</p>	<p>Die Möglichkeiten zur Täuschung beunruhigen mich / Man weiß nie was wirklich echt ist / sicherlich Kosten- und Ressourcensparend für Unternehmen / gefährlich, da es einen evtl. selber treffen könnte und man sich in einer Deepfake Situation vorfindet mit der man nichts zu tun hat / Deepfakes bergen ein hohes Potential, aber auch Risiko. / K.a. / kann ganz witzig sein, kann aber auch für falsche Nachrichten verwendet werden (kann gefährlich werden) / führt zu noch mehr confirmation bias und misstrauen gegenüber Aussagen und Meinungen die der eigenen Überzeugung nicht entsprechen / Schwierig, man kann kaum die Unterschiede erkennen / Gefährlich, wenn reale Leute nachgestellt werden</p>	<p>Spannend und auch irritierend / wenn man das Gesicht erkennt bleibt die Werbung nachhaltiger im Gedächtnis / mehr Identifikation mit der Werbung / Marken erhalten noch mehr Kontrolle über ihre Inhalte / Die Personalisierung von Werbung kann damit auf ein neues Level gehoben werden. / StepStone / sollte meiner Meinung nicht stattfinden (Irreführung) / suggerieren noch mehr persönlichen Bezug zu Produkten und verstärken damit psychologisch manipulierend den Wunsch, zu noch mehr Konsum / je nach Zielperson könnte der Deepfake andere Inhalte vermitteln. / Werbeaussagen könnten als nach Zielgruppe stark variieren</p>

POI: Pia

Technische Umsetzung	Deepfakes in den Medien	Personalisierung von Werbung durch Deepfakes
Kann ich nicht beurteilen / Hervorragend / Sah nicht sehr echt aus / verbesserungswürdig / scheint plausibel, ohne Vorkenntnisse oft nicht zu erkennen	Wird immer mehr vorhanden sein in Zukunft / Irgendwann wird es nicht mehr wegzudenken sein / Viel verbreitet / könnte Funktionieren / bringt das produkt näher zum verbraucher / besserer verkauf weil es durch freunde vermarktet wird / kann bald eingesetzt werden / wirkt natürlich	Ist die Zukunft / Gefährlich, Hoaxes / Funktioniert / macht es für den kunden leichter sachen zu kaufen weil man freunden mehr vertraut / ist eine mögliche Zukunft

POI: Alexander

Technische Umsetzung	Deepfakes in den Medien	Personalisierung von Werbung durch Deepfakes
Gut nur sollte man nicht die Stimme von Christopher Maria Herbst nehmen. / kann täuschend echt aussehen / Erstaunlich gut / sieht realistisch aus, man erkennt die Person ohne Probleme / gut gelungen / erschreckend was technisch möglich ist	Finde ich gut / viele Möglichkeiten prominete Gesichter für Werbung zu benutzen, den Kunden aufmerksamer auf das Produkt zu machen / Nur für Propagandazwecke gut / gut für Social Media, erregt Aufmerksamkeit weil man es noch nicht so oft gesehen hat / unnötig / Menschen sehen dann tagtäglich Fake News	Muss mit bedachte erarbeitet werden. / kann die Kaufaktivität anregen wegen der bekannten Gesichter / Finde ich überflüssig / kann bestimmt nützlich sein / könnte helfen verschiedenste Vorlieben leichter zu verbinden

POI: Micha

Technische Umsetzung	Deepfakes in den Medien	Personalisierung von Werbung durch Deepfakes
An diesem Beispiel sehr gut / Wahrscheinlich hätte ich gedacht, die Person existiert wirklich, wenn ich vorher nicht gewusst hätte, wer das ist / Ist genial gelungen / man sieht in der Spiegelung noch Stromberg / schon gut gemacht , aber man sieht trotzdem noch das irgendwas nicht stimmt mit dem Video	Kommt auf Thema an, grundsätzlich in bestimmten Kreisen/Themen vielleicht etwas gefährlich / Ist bestimmt sinnvoll, wenn man mehr Nähe zum Kunden erreichen möchte, und eine persönlichere Ebene / Beunruhigt mich, es wird schwieriger fakenews zu entlarven. / erschreckend was heutzutage möglich ist / angsteinflößend / spannend zu sehen / regen die Aufmerksamkeit an	Bei kommerzieller Werbung unnötig, zur Aufklärung gut / Die Mischung von 2 Personen im ersten Moment seltsam und komisch / Macht mir Angst. / gibt der Werbung eine persönliche Note / witzig wenn man die Person kennt die man dort sieht / große Auswahl an Möglichkeiten für Personalisierung durch Deepfakes

Die SD-Karte mit dem digitalen Anhang dieser Masterarbeit beinhaltet folgende Dateien:

- digitale Version der Masterarbeit
- 4 selbst produzierte Deepfakes
- Screenshots aus DeepFaceLab während der Deepfake-Produktion
- Test-Deepfakes
- Zeitraffer-Videos der SAEHD-Trainings
- Deepfake Assets (src-Videos der POI, dst-Videos der Original-Werbungen)
- alle Skripte aus dem genutzten DFL *build*
- Online-Fragebogen als Druckversion
- Rohdaten als Excel-Tabelle
- Codebuch
- Auflistung der offene Antworten aufgeteilt nach Person-Of-Interest
- SPSS Auswertung (Deskriptive Statistik, Häufigkeiten, Kreuztabellen)
- SPSS Projekt
- Einverständniserklärungen der POI
- Leitfaden für Videoaufnahmen der POI