



**Hochschule  
Bonn-Rhein-Sieg**  
University of Applied Sciences

Fachbereich: Elektrotechnik, Maschinenbau und Technikjournalismus (EMT)  
Studiengang: Technik- und Innovationskommunikation (M.Sc.)

MASTER-THESIS

# Die Verifikation audiovisuellen Materials im digitalen Zeitalter

---

Technische Möglichkeiten und journalistische Praxis

Vorgelegt von:

Martha Rebekka Peters



email@martha-peters.de

Matrikel-Nr: 

Erstprüferin: Prof. Dr. Susanne Keil

Zweitprüferin: Prof. Dr. Tanja Köhler

Sankt Augustin, 14. Juli 2023<sup>1</sup>

---

<sup>1</sup> Für die Veröffentlichung wurde die Arbeit nachträglich noch an einzelnen Stellen angepasst.

## INHALTSVERZEICHNIS

|  |           |
|--|-----------|
| <b>ABBILDUNGSVERZEICHNIS .....</b>   | <b>IV</b> |
| <b>TABELLENVERZEICHNIS .....</b>   | <b>V</b>  |
| <b>ABKÜRZUNGSVERZEICHNIS .....</b>   | <b>VI</b> |
| <b>1 EINLEITUNG .....</b>  | <b>1</b>  |
| <b>2 FORSCHUNGS- UND SACHSTAND .....</b>   | <b>3</b>  |
| 2.1 ÜBERBLICK ÜBER DEN FORSCHUNGSSTAND.....  | 3         |
| 2.2 DIE JOURNALISTISCHE BEDEUTUNG AUDIOVISUELLER MEDIEN IM DIGITALEN ZEITALTER .....     | 4         |
| 2.2.1 <i>Audiovisuelle Medien</i> .....  | 4         |
| 2.2.2 <i>Die Bedeutung audiovisueller Medien für den Journalismus</i> .....              | 5         |
| 2.2.3 <i>Audiovisuelle Medien im digitalen Zeitalter</i> .....                           | 6         |
| 2.2.4 <i>Audiovisuelle Medien als Träger von Desinformation</i> .....                    | 8         |
| 2.3 DIE MANIPULATION AUDIOVISUELLEN MATERIALS IM DIGITALEN ZEITALTER .....               | 11        |
| 2.3.1 <i>Shallowfakes / Cheapfakes</i> .....   | 13        |
| 2.3.2 <i>Synthetische Medien / Deepfakes</i> .....                                       | 15        |
| 2.4 DER TECHNISCHE RAHMEN FÜR D. VERIFIKATION AUDIOVISUELLEN MATERIALS .....             | 22        |
| 2.4.1 <i>Der Weg multimedia-forensischer Methoden in den Journalismus</i> .....          | 23        |
| 2.4.2 <i>Überprüfung des Publikationskontextes u. Prüfung früherer Verbreitung</i> ..... | 25        |
| 2.4.3 <i>Inhaltliche Überprüfung</i> .....   | 29        |
| 2.4.4 <i>Methoden auf Pixelebene</i> .....   | 32        |
| 2.4.5 <i>Identifizierung Synthetischer Medien / Deepfakes</i> .....                      | 39        |
| 2.4.6 <i>Automatisierte Manipulationserkennung</i> .....                                 | 41        |
| 2.4.7 <i>Aktive Forensik</i> .....   | 44        |
| 2.5 DIE VERIFIZIERUNGSPRAXIS IN REDAKTIONEN .....  | 46        |
| <b>3 METHODIK DER EMPIRISCHEN ERHEBUNG .....</b>   | <b>53</b> |
| 3.1 LEITFRAGEN .....   | 53        |
| 3.2 QUALITATIVE EXPERTENINTERVIEWS .....   | 54        |
| 3.2.1 <i>Vorstellung der Interviewpartner:innen</i> .....                                | 56        |
| 3.2.2 <i>Interviewleitfäden als Erhebungsinstrument</i> .....                            | 61        |
| 3.2.3 <i>Datenerhebung und -erfassung</i> .....  | 65        |
| 3.2.4 <i>Transkription und Interviewbericht</i> .....                                    | 66        |
| 3.3 EINE QUALITATIVE INHALTSANALYSE NACH UDO KUCKARTZ .....                              | 66        |
| 3.4 PROBLEME UND ANMERKUNGEN .....   | 70        |

## INHALTSVERZEICHNIS

---

|          |   |            |
|----------|---|------------|
| <b>4</b> | <b>ERGEBNISSE: DIE EINSCHÄTZUNGEN DER EXPERT:INNEN.....</b>                                       | <b>71</b>  |
| 4.1      | DIE TECHNISCHE ENTWICKLUNG AUDIOVISUELLER MANIPULATIONEN .....                                    | 71         |
| 4.2      | MÖGLICHKEITEN UND GRENZEN DER MULTIMEDIA-FORENSIK .....   | 76         |
| 4.2.1    | <i>Definition und Systematisierung multimedia-forensischer Bereiche.....</i>                      | <i>76</i>  |
| 4.2.2    | <i>Die ML-basierte Detektion von Deepfakes .....</i>  | <i>78</i>  |
| 4.2.3    | <i>Möglichkeiten und Hürden aktiver Forensik.....</i>   | <i>79</i>  |
| 4.3      | DIE PRAXIS .....  | 81         |
| 4.3.1    | <i>Vorkommende Manipulationsformen .....</i>  | <i>81</i>  |
| 4.3.2    | <i>Praxis und Probleme in Redaktionen und Recherchenetzwerken.....</i>                            | <i>84</i>  |
| 4.4      | EINSCHÄTZUNG DER TATSÄCHLICHEN DETEKTIERBARKEIT AUDIOVISUELLER MEDIEN – HEUTE UND ZUKÜNFTIG ..... | 91         |
| 4.5      | EINSCHÄTZUNGEN ZU EINFLUSS U. RISIKEN AUDIOVISUELLER MANIPULATIONEN.....                          | 94         |
| 4.6      | WAS MUSS GETAN WERDEN? MÖGLICHE HANDLUNGSFELDER .....   | 96         |
| <b>5</b> | <b>FAZIT: ZUSAMMENFASSUNG UND AUSBLICK .....</b>  | <b>100</b> |
| 5.1      | ZUSAMMENFASSUNG.....  | 100        |
| 5.2      | FAZIT.....  | 101        |
| 5.3      | WEITERE MÖGLICHE FORSCHUNGSFELDER .....   | 102        |
|          | <b>LITERATURVERZEICHNIS.....</b>  | <b>103</b> |
|          | <b>INDEX.....</b>   | <b>114</b> |
|          | <b>ANHANG .....</b>   | <b>116</b> |

## ABBILDUNGSVERZEICHNIS

|   |    |
|---|----|
| ABBILDUNG 1: PROJEKTE RUND UM DIE VERIFIZIERUNG AUDIOVISUELLER INHALTE UNTER BETEILIGUNG DER<br>DW-INNOVATIONSABTEILUNG (GRAFIK: MARTHA PETERS) ..... | 23 |
| ABBILDUNG 2: DAS KATEGORIENSYSTEM (GRAFIK: MARTHA PETERS).....  | 69 |

## TABELLENVERZEICHNIS

|   |    |
|---|----|
| TABELLE 1: ÜBERBLICK ÜBER MANIPULATIONSFORMEN (GRAFIK: MARTHA PETERS) .....                               | 13 |
| TABELLE 2: ÜBERBLICK ÜBER DIE IN DEN EXPERTENINTERVIEWS ANGESPROCHENEN THEMEN (GRAFIK: MARTHA PETERS).... | 64 |
| TABELLE 3: ÜBERBLICK ÜBER VERIFIZIERUNGSVERFAHREN (GRAFIK: MARTHA PETERS) .....                           | 77 |

### **ABKÜRZUNGSVERZEICHNIS** (für Erklärungen siehe Index)

|                |   |
|----------------|---|
| AFP            | Agence France-Presse  |
| AI             | Artificial Intelligence   |
| ARD            | Arbeitsgemeinschaft der öffentlich-rechtlichen Rundfunkanstalten der Bundesrepublik Deutschland                               |
| AV-Medien      | audiovisuelle Medien  |
| BBC            | British Broadcasting Corporation  |
| BR             | Bayerischer Rundfunk  |
| BSI            | Bundesamt für Sicherheit in der Informationstechnik   |
| CAA            | Context Aggregation and Analysis (= Software-Eigenname)   |
| CAI            | Content Authenticity Initiative   |
| C2PA           | Coalition for Content Provenance and Authenticity   |
| CGI            | Computer Generated Imagery  |
| DL             | Deep Learning   |
| dpa            | Deutsche Presse-Agentur   |
| DW             | Deutsche Welle  |
| DW ReCo        | Deutsche Welle Research and Cooperations Projects   |
| EDMO           | European Digital Media Observatory  |
| ENF            | Electrical Network Frequency  |
| FAU            | Friedrich-Alexander-Universität Erlangen-Nürnberg   |
| FAZ            | Frankfurter Allgemeine Zeitung  |
| fps            | frames per second   |
| Fraunhofer SIT | Fraunhofer-Institut für Sichere Informationstechnologie   |
| GAN            | Generative Adversarial Networks   |
| GADMO          | German-Austrian Digital Media Observatory   |
| HDR            | High Dynamic Range  |
| IFCN           | International Fact-Checking Network   |
| IKSŽ           | Institut komunikačních studií a žurnalistiky (= Institut für Kommunikationswissenschaft u. Journalismus d. Karls-Universität) |
| iOS            | Internetwork Operating System (= Betriebssystem v. Apple)   |
| IT             | Informationstechnologie   |

## ABKÜRZUNGSVERZEICHNIS

---

|           |   |
|-----------|---|
| ITAS      | Institut für Technikfolgenabschätzung und Systemanalyse des KIT                         |
| JPEG-AI   | KI-basierter Kompressionsstandard für Fotos der Joint Photographic Experts Group (JPEG) |
| KI        | Künstliche Intelligenz  |
| KIT       | Karlsruher Institut für Technologie   |
| ML        | Machine Learning / Maschinelles Lernen  |
| NGO       | Non-Governmental Organization / Nichtregierungsorganisation                             |
| OSINT     | Open Source Intelligence  |
| PRNU      | Photo Response Non-Uniformity<br>(= charakteristisches Bildsensorrauschen)              |
| RGB-Werte | Rot-Grün-Blau-Werte   |
| RND       | RedaktionsNetzwerk Deutschland  |
| RTL       | Radio Télévision Luxembourg   |
| SZ        | Süddeutsche Zeitung   |
| T2I / T2V | Text-to-Image-/Text-to-Video-Verfahren  |
| UGC / UGV | user-generated content / user-generated video   |
| URL       | Uniform Resource Locator (= Internetadresse)  |
| WHO       | World Health Organization / Weltgesundheitsorganisation                                 |
| ZDF       | Zweites Deutsches Fernsehen   |

## 1 Einleitung

"But we're not just fighting an epidemic; we're fighting an infodemic. Fake news spreads faster and more easily than this virus, and is just as dangerous."

(Tedros Adhanom Ghebreyesus 2020)

Das Zitat des WHO-Generaldirektors zu Beginn der Covid-19-Pandemie auf der Münchner Sicherheitskonferenz im Februar 2020 unterstreicht, dass Desinformation nicht nur den (Wissenschafts-)Journalismus vor Herausforderungen stellt, sondern auch das Potenzial hat, unmittelbaren Einfluss auf jedes einzelne Individuum auszuüben, in diesem Fall mit gesundheitlichen Folgen.

Gerade die Manipulation audiovisuellen Materials stellt dabei ein zunehmendes Risiko dar. Auf der einen Seite wird audiovisuellen Medien die Möglichkeit zugeschrieben, ein auf physikalischen Gegebenheiten beruhendes Abbild der Wirklichkeit zu schaffen, wodurch sie für den Journalismus von zentraler Bedeutung sind (vgl. Kapitel 2.2.2). Auf der anderen Seite ermöglichen die technologischen Entwicklungen der letzten Jahre immer einfacher, kostengünstiger und schneller authentisch wirkende Manipulationen zu erstellen. Noch vor zehn Jahren war die Manipulation von Videomaterial, abgesehen von trivialen Operationen auf Bildebene, nur im Rahmen von Filmproduktionen möglich. Das ist inzwischen anders (vgl. Kapitel 2.3.2). Besonders der rasche Fortschritt bei Anwendungen maschinellen Lernens spielt dabei eine Rolle und führt, zusammen mit den veränderten Verbreitungsmöglichkeiten durch das Internet und dem Aufstieg populistischer Akteur:innen (vgl. Riess 2021, 344) zu einer Gefahr für die demokratische Meinungsbildung. Audiovisuelle Manipulationen werden damit zu einem zentralen Problem für die Gesellschaft (vgl. Verdoliva 2020, 910).

Gerade in Krisen- und Kriegssituationen wie zu Beginn der Covid-19-Pandemie oder während des Krieges in der Ukraine ist es besonders schwierig, Material einzuordnen. Zumal gerade dann nutzergenerierte Inhalte mitunter die einzig verfügbare Quelle sind (vgl. Kapitel 2.2.3). Als gesellschaftliches Teilsystem, das einen Gesamtüberblick über das relevante und reale Geschehen verschaffen und in Demokratien auch als Moderator des öffentlichen Diskurses fungieren soll, ist der Journalismus von den Auswirkun-



gen authentisch wirkender Manipulationen besonders betroffen. Schon heute stellen audiovisuelle Manipulationen Redaktionen vor eine Herausforderung und schaffen es mitunter auch als vermeintlich authentischer Inhalt in die Berichterstattung. Um seiner Funktion gerecht zu werden und eine demokratische Meinungsbildung zu ermöglichen, muss im Journalismus authentisches Material von Manipulationen unterschieden werden können (siehe Kapitel 2.2.4 u. 2.3). Sollte das heute oder in Zukunft nicht mehr möglich sein, wird es notwendig, die Bedeutung audiovisuellen Materials im Journalismus zu hinterfragen und neue Kriterien für dessen Verwendung zu erarbeiten. Es stellt sich daher die folgende Forschungsfrage:

**Inwiefern ist und bleibt es möglich, die Authentizität audiovisuellen Materials in Redaktionen sicherzustellen?**

Um sich dieser Frage zu nähern, liefert die Arbeit zunächst einen Überblick über den zugehörigen Forschungs- und Sachstand. Ergänzt, vertieft und eingeordnet werden die zusammengetragenen Informationen im zweiten Teil der Arbeit durch eine qualitative Inhaltsanalyse von insgesamt sieben Experteninterviews mit Akteur:innen aus Wissenschaft und Praxis. Die Experteninterviews erlauben zum einen den aktuellen Stand der Technik einzufangen, der sich dynamisch verhält, als dass entsprechende Aussagen in wissenschaftlichen Veröffentlichungen längere Zeit ihre Gültigkeit behalten. Vor allem aber geht es um die Beurteilung existierender Probleme und potenzieller Lösungen, sowie eine Einschätzung der zukünftigen Entwicklung relevanter Bereiche durch die Interviewten. Aus ihrer Perspektive sollen mögliche Antworten auf die Forschungsfrage, aber auch Ansätze für eine weitere Forschung in diesem Bereich entwickelt werden. Durch die Auswahl der Interviewten wurde versucht, ein möglichst großes Spektrum an Perspektiven auf die Forschungsfrage zusammenzutragen, eine vollumfängliche Betrachtung ist jedoch nicht möglich. Zudem kann die Arbeit aufgrund des dynamischen Forschungsbereichs nur eine Momentaufnahme darstellen:

"The pace of technological developments suggest that, by the end of the decade, synthetic media may radically change the information ecosystem and demand new architectures and systems to aid sensemaking" (Hendrix u. Morozoff 2022, 9)

## 2 Forschungs- und Sachstand

### 2.1 Überblick über den Forschungsstand

Die wissenschaftliche Literatur zu den technischen Herausforderungen und Lösungsansätzen im Rahmen der Verifizierung audiovisuellen Materials ist bisher fast ausschließlich englischsprachig und besteht zum großen Teil aus Momentaufnahmen einzelner IT-forensischer Lösungsansätze. Ein Symptom, das genauso auf das vergleichsweise junge Alter der digitalen Multimedia-Forensik hinweist, wie auf die starke Dynamik des Forschungsbereichs. Die Entwicklung der digitalen Multimedia-Forensik wird vorwiegend in den letzten 15 Jahren gesehen<sup>2</sup> (vgl. Verdoliva 2020, 910), das Potenzial KI-basierter Manipulationen zeigt sich überhaupt erst seit der öffentlichkeitswirksamen Verbreitung synthetischer Medien Ende 2017.

Für die technischen Grundlagen stützt sich die vorliegende Arbeit vor allem auf den Sammelband *Multimedia Forensics* (2022), in dem die Professor:innen Sencar, Verdoliva und Memonder einen aktuellen Überblick über die grundlegend verschiedenen Ansätze geben, mit denen man die Authentizität audiovisuellen Materials prüfen oder sicherstellen kann. Ergänzt wird der Überblick in dieser Arbeit um einen stärkeren Bezug zur journalistischen Praxis durch den Sammelband *Video Verification in the Fake News Era* von 2019, der anhand der für Journalist:innen entwickelten Softwarelösungen im Rahmen des Projekts InVID technologische Fortschritte und praktische Werkzeuge für das Auffinden, Überprüfen und Visualisieren von Videoinhalten in sozialen Medien zusammenfasst (vgl. Mezaris et al. 2019, V). Darüber hinaus greift die Arbeit im Theorieteil auf zahlreiche Beiträge aus der Informatik, den Medienwissenschaften und der journalistischen Praxis zurück, die sich mit Teilbereichen des Themas befassen.

---

<sup>2</sup> Diese zeitliche Einordnung spiegelt sich auch in der Entwicklung der Institutionen und Forschungsgruppen wieder, aus denen die Interviewten für diese Arbeit stammen. Besonders lange beschäftigt sich das Fraunhofer SIT mit der Sicherheit von Bild-, Audio- und Videodaten – seit etwas über 20 Jahren. Die offizielle Arbeitsgruppe der FAU Erlangen-Nürnberg für Multimedia-Sicherheit gibt es seit Ende 2016, auch wenn ihr Leiter, Dr. Christian Riess, sich bereits seit 2008 mit dem Thema beschäftigt. Die DW-Innovationsabteilung verfolgt das Thema seit 2012. In diese Zeit fällt auch die Gründung von Bellingcat durch Higgins 2014, nachdem er selbst bereits seit 2011 mit der Verifizierung audiovisuellen Materials beschäftigt war. Die Beschäftigung mit Deepfakes ist entsprechend der Technologienentwicklung noch neuer und findet im BSI seit 2019/20 statt. Ähnlich sieht es beim KIT aus (vgl. Interviews).

## 2.2 Die journalistische Bedeutung audiovisueller Medien im digitalen Zeitalter

### 2.2.1 Audiovisuelle Medien

Eine einheitliche Definition des Begriffs der audiovisuellen Medien existiert nicht. Orientiert man sich an der eng gefassten Definition des Medienwissenschaftlers Hans J. Kleinsteuber (2013, 21) im *Lexikon Kommunikations- und Medienwissenschaft*, handelt es sich bei audiovisuellen Medien ausschließlich um „Medienangebote, die gleichzeitig akustisch, also auf das Ohr, und optisch, also auf das Auge, wirken“. Gleichzeitig weist Kleinsteuber bereits darauf hin, dass der Begriff inzwischen auch „auch auf das multimediale Angebot des Internet angewandt“ (ebd.) werde. In einem anderen kommunikationswissenschaftlichen Lehrbuch wird auch das Radio als audiovisuelles Medium benannt – mit dem Hinweis, „dass durch die Aufsplitterung von Geräten, Diensten und Institutionen nur noch schwer von *einem Medium* gesprochen werden könne“ (Paus-Hasebrink et al. 2006, 2; Bezug auf Goertz 1995, 485).

Die aktuelle Mediennutzungsanalyse des Verbandes der audiovisuellen Medienunternehmen in Deutschland VAUNET unterstreicht die uneinheitliche Verwendung des Begriffs in der Öffentlichkeit. Während in der Überschrift noch in „Audio- und audiovisuelle Mediennutzung“ (VAUNET 03.02.22) differenziert wird, ist anschließend nur noch von „Fernsehen, Radio und anderen audiovisuellen Medien“ (ebd.) die Rede und im „audiovisuellen Medienzeitbudget“ (ebd.) sind Radio, Podcasts/Hörbücher und CDs genauso aufgeführt wie Videospiele, Kino, DVD/Blu-Ray, Onlinevideos und Fernsehen.<sup>3</sup> Unabhängig von der begrifflichen Unschärfe zeigt das erhobene Medienzeitbudget in jedem Fall die (zunehmende) Relevanz audiovisuellen Materials für die deutsche Medienlandschaft: Zehn Stunden verbrachten über-14-Jährige laut der Studie 2021 durchschnittlich pro Tag mit dem Konsum audiovisueller Medien, auch vor der Covid-19-Pandemie waren es bereits über neun Stunden. Die Zahlen sind lediglich als Näherungswert zu betrachten, da die medienübergreifenden Angaben zur täglichen Medi-

---

<sup>3</sup> Die uneinheitliche Verwendung des Begriffs findet sich genauso in der Pressemitteilung wie in der publizierten Langfassung der Studie.

ennutzungsdauer aus unterschiedlichen Quellen zusammengetragen werden (vgl. ebd.), machen aber den Stellenwert audiovisueller Medien in der Gesellschaft deutlich. In der vorliegenden Arbeit wird eine entsprechend weit gefasste Definition angewandt, die der heutigen Medienlandschaft und ihren schwer voneinander abzugrenzenden Mischformen gerecht wird. Gerade im Internet zeigt sich inzwischen eine Vielzahl an Medieninhalten, die (Bewegt-)Bild, Ton und teilweise sogar Text ganz unterschiedlich miteinander kombinieren. Als Beispiel sind Videos zu nennen, die aus einer Aneinanderreihung von Fotos bestehen, überlagert mit Text, untermalt mit Musik. Eine Manipulation kann dabei auf mehreren dieser Ebenen oder auch nur auf einer einzelnen Ebene stattfinden, weshalb die auditiven und visuellen Elemente des Materials in jedem Fall einzeln betrachtet werden müssen. Unter dem Begriff des audiovisuellen Materials werden daher sowohl auditive als auch visuelle Inhalte zusammengefasst.

### *2.2.2 Die Bedeutung audiovisueller Medien für den Journalismus*

Bildmaterial, Fotos wie Videos, haben im Journalismus eine zentrale Bedeutung. Der Stellenwert erklärt sich zum einen durch die vermeintliche Möglichkeit, ein auf physikalischen Gegebenheiten beruhendes Abbild der Wirklichkeit zu schaffen und weiterzugeben. In der Folge wird Bildmaterial als Beweis eingesetzt und auch von Rezipient:innen als solcher verstanden (vgl. Peters 2020, 16ff.). Noch immer wird die Glaubwürdigkeit von Bildmaterial besonders im alltäglichen Gebrauch kaum infrage gestellt. Selbst wenn etwas unwahrscheinlich ist, glaube man, was man sieht (vgl. Gerling 2022, 298; Kietzmann et al. 2020, 136). Hinzu kommt, dass Bildinhalte schneller erfasst und nachhaltiger behalten werden als Textinhalte und dadurch in der Lage sind, uns in besonderem Maß zu beeinflussen. Das ist vermutlich eine Ursache dafür, dass der Stellenwert von Bild- und Videomaterial im Zuge der Digitalisierung weiter zugenommen hat:<sup>4</sup>

„Wenngleich ‚Show, don’t tell!‘ vor allem in der fiktionalen Welt bekannt ist, lässt sich das Prinzip ohne viel Fantasie auch in der generellen Medien- und Nachrichtenlandschaft des 21. Jahrhunderts wiederentdecken: Die einstige Dominanz des Formats Text ist hier sukzessive einem Kaleidoskop aus statischen und bewegten Bildern gewichen.“ (Stein et al. 2020, 178)

---

<sup>4</sup> Siehe Pictorial Superiority Effect, die negativen Auswirkungen zeigen sich u. a. als „Visualisierungszwang“ (vgl. Peters 2020, 24ff.)

### 2.2.3 Audiovisuelle Medien im digitalen Zeitalter

Die technischen Möglichkeiten der Erstellung und Verbreitung audiovisueller Medien haben sich in den letzten 40 Jahren grundlegend verändert – und damit auch das journalistische Berufsfeld (vgl. Koltermann u. Grittmann 2022, 14ff.). Insgesamt habe sich die Art und Weise, wie wir Informationen erstellen, verbreiten und konsumieren, im Zeitalter der Digitalisierung so stark gewandelt, wie zuletzt mit dem Buchdruck (vgl. Sangerlaub 2020, 106f.). Dabei ist es nicht nur deutlich einfacher und kostengunstiger geworden, audiovisuelles Material zu erstellen, sondern auch, es weltweit zu verbreiten – fur Journalist:innen, aber vor allem fur Privatpersonen:

"Nowadays, it is possible (at least technically) for anyone to publish and distribute digital content with ease and speed to a potentially worldwide audience, and reach, potentially, millions of people. In turn, the so-called 'legacy media' (news agencies, traditional newspapers, magazines, broadcasters, etc.) are no longer the exclusive gatekeepers who decide what is being circulated to a wider audience, and what is not." (Spangenberg u. Teyssou 2019, 4)

In der Folge ist von einer „Inflation der Bilder“ (Koltermann u. Grittmann 2022, 9) die Rede. Alleine auf der Videoplattform Youtube werden nach den dortigen Angaben jeden Tag mehrere Milliarden Stunden Videomaterial geschaut und der Plattform jede Minute mehr als 500 Stunden Videomaterial hinzugefugt (vgl. YouTube 2023). Mit der Folge, dass gerade durch Privatpersonen auch Material von Situationen existiert, die unvorhersehbar und/oder fur Journalist:innen unerreichbar sind. So ermogliche die fortgeschrittene Digitalisierung eine Dokumentation des Krieges in der Ukraine, wie sie bei bisherigen Konflikten kaum moglich gewesen sei (vgl. Bischoff 2022a, 71):

„Was die Informationsdichte angeht, unterscheiden sich die heutigen Konflikte deutlich von jenen in der Vergangenheit. Denn der Arabische Fruhling hat eine Zeitenwende in der Berichterstattung eingeleitet [...]. Einerseits eroffnete das ungeahnte Moglichkeiten fur Journalistinnen und Journalisten, um die Lage vor Ort einzuschatzen, ohne sich selbst in Gefahr zu begeben. Andererseits stellte das die Medien, die fundiert und unabhangig berichten mochten, vor eine Herausforderung: Welchen Informationen kann man glauben? Wie vergewissert man sich, dass die Aufnahmen nicht gestellt oder gefalscht sind [...]?“ (Bischoff 2022b, 64)

In einigen Situationen ist nutzergeneriertes Material nicht nur als erstes, sondern auch als einziges verfugbar. Daher sind nutzergenerierte Inhalte (UGC = user-generated content) inzwischen ein wichtiger Bestandteil journalistischer Berichterstattung (vgl. Khan

et al. 2023, 6748; Apostolidis et al. 2019, 54; Fricke u. Thomsen 2019, 301; Spangenberg u. Teyssou 2019, 3f.). Während bei Inhalten professioneller Pressefotograf:innen oder Journalist:innen jedoch gemeinhin die Autorenschaft als Authentizitätsbeweis akzeptiert wird (vgl. Gerling 2022, 304), müssen nutzergenerierte Inhalte zunächst auf ihre Authentizität geprüft werden, um die Verbreitung falscher Informationen zu verhindern (vgl. Spangenberg u. Teyssou 2019, 6; Apostolidis et al. 2019, 87):

„Die Verifikation von Bildern und Videos aus dem Netz im Rahmen der aktuellen Berichterstattung gehört zur journalistischen Sorgfaltspflicht.“  
(Sell u. Oswald 2021, 258)

Schließlich unterscheide die Praxis des Verifizierens Journalismus von anderen Formen öffentlicher Kommunikation wie Public Relations, Werbung, Propaganda, Gerüchten oder Fiktionen (vgl. Sell u. Oswald 2021, 244). Gleichwohl gab es auch bei professionellen Quellen bereits Fälle, welche im Nachhinein als unzulässige Manipulation beurteilt wurden und zu strikten Regeln für die Nachbearbeitung von Material geführt haben (vgl. Gerling 2022, 310f.). An diesen Regeln kann man sich orientieren, um erlaubte Bildanpassungen von Manipulationen zu unterscheiden.

Zugleich erhöhen das Internet und insbesondere Social Media den Druck auf Journalist:innen, möglichst in Echtzeit über Nachrichtenlagen zu berichten. So stellt u. a. der AFP-Journalist Denis Teyssou fest:

"In nineteenth-century empires, it took months or weeks to get international news until the telegraph submarine cables reduced that delay to a few hours. In the twentyfirst century, news now spreads in milliseconds on social media and major events are broadcasted live on connected screens and devices. Time to react, understand, and verify has shrunk dramatically while complexity of events, interactions, strategies, and technology is increasing endlessly."  
(Teyssou 2019, 265)

Auf der einen Seite setzt die schnelle Verbreitung von Informationen über Social Media also journalistische Akteure zeitlich unter Druck, auf der anderen Seite verleiht die Verwendung nutzergenerierter Inhalte der Notwendigkeit, audiovisuelles Material zu verifizieren, eine neue Brisanz.

#### 2.2.4 Audiovisuelle Medien als Träger von Desinformation

Relevant für den Journalismus werden audiovisuelle Manipulationen insbesondere in dem Moment, wo sie zu *Fake News* werden und somit entweder selbst eine „in manipulativer Absicht verbreitete Falschmeldung“ (Duden 2020) darstellen oder zu einer solchen beitragen. Der Begriff *Fake News* wird jedoch als politisch motivierter Begriff gegen die Presse von Wissenschaftler:innen und Journalist:innen nicht gerne verwendet. Eine Übersetzung von *Fake News* durch die deutsche *Falschmeldung* greife nach dem Kommunikationswissenschaftler und Journalisten Alexander Sänglerlaub auch zu kurz, da es so nicht möglich sei, zwischen absichtlichen und unbeabsichtigten Falschmeldungen zu unterscheiden. Stattdessen wird die gezielt verbreitete Falschinformation als Desinformation bezeichnet und damit von der Verwendung des Begriffs mit politischen Absichten aber auch der unbeabsichtigten Falschmeldung abgegrenzt (vgl. Sänglerlaub 2020, 100; Stern 2020, 119f.).

Gerade während der Covid-19-Pandemie und des Krieges in der Ukraine habe die Menge an Desinformation nochmal deutlich zugenommen, meint der Leiter des Faktencheck-Teams der Deutschen Welle (DW), Joscha Weber (vgl. Pörzgen 19.08.22). Vom RedaktionsNetzwerk Deutschland (RND) heißt es, noch nie habe ein moderner Krieg eine solche Flut an nicht nachprüfbar Bildern produziert (vgl. Grimm 16.03.22) und auch das ZDF beschreibt, dass gerade die ersten Wochen nach dem Angriff von einer unklaren Nachrichtenlage, als auch von einer Videoflut aus den Sozialen Netzwerken geprägt gewesen sei, die unter Zeitdruck verifiziert werden musste (vgl. Rating 2022, 360). Der Gründer des Recherchenetzwerks Bellingcat, Elliot Higgins (2021, 149, vgl. 59), sieht Desinformation inzwischen als wichtigen Bestandteil der Außenpolitik autoritärer Staaten und spricht von einem „Zeitalter der Informationskriege“.

So werden tatsächlich nicht nur Einzelpersonen, sondern auch staatliche Akteure mit der Erstellung und Verbreitung von Desinformation in Verbindung gebracht, u. a. Saudi-Arabien, die Vereinigten Arabischen Emirate und Russland. Gerade in den letzten zwei Jahren habe es eine beträchtliche Menge an audiovisuellen Manipulationen aus China gegeben, die bisher allerdings nicht in direkten Zusammenhang mit der chinesischen Regierung gebracht werden könnten. Fakt ist jedoch bspw., dass der

chinesische Außenminister im November 2020 ein gefälschtes Bild eines australischen Soldaten in Afghanistan postete, dem die Intention nachgesagt wird, Spannungen in Afghanistan zu befeuern. Für Staaten, die nicht selbstständig in der Lage seien, ein solches Manipulationspotenzial zu entfalten, zum Beispiel durch einen Mangel an technologischen Ressourcen, gebe es inzwischen eine völlig neue Branche an Unternehmen (vgl. Hendrix u. Morozoff 2022, 15f.).<sup>5</sup>

Als gesellschaftliches Teilsystem, das einen Gesamtüberblick über das relevante und reale Geschehen verschaffen, wie auch in Demokratien als Moderator des öffentlichen Diskurses fungieren soll (vgl. Neuberger u. Kapern 2013, 26ff.), ist der Journalismus von den Auswirkungen authentisch wirkender audiovisueller Manipulationen potenziell besonders betroffen. Werden fälschlicherweise manipulierte Inhalte für authentisch gehalten und weitergegeben, schadet das nicht nur dem Ruf der Journalist:innen und des Medienunternehmens, sondern beeinträchtigt auch das Vertrauen in den Journalismus allgemein (vgl. Khan et al. 2023, 6749; Spangenberg u. Teyssou 2019, 10; Apostolidis et al. 2019, 55).

Zudem kann die Verbreitung von Desinformation durch Journalist:innen auch großen gesellschaftlichen Schaden anrichten: Eine Studie des Berliner Think Tanks *Stiftung neue Verantwortung*, zu dem Sänglerlaub gehört, weist darauf hin, dass Desinformation erst dann eine große Reichweite erziele, wenn sich klassische Medien an deren Verbreitung beteiligen. Das könne versehentlich, absichtlich oder auch im Rahmen einer Richtigstellung passieren (vgl. Sänglerlaub 2020, 102f.). Die gleiche Studie kam zu dem Ergebnis, dass Desinformationen in der Regel eine deutlich höhere Reichweite erreichen als entsprechende Richtigstellungen durch einen Faktencheck (vgl. ebd., 105f.). Daraus lässt sich schließen, dass die Folgen einer voreilig verbreiteten Desinformation kaum mehr zu korrigieren sind. Als Konsequenz aus den Folgen, die mit der Verbreitung von Desinformationen durch klassische Medien verbunden sind, sieht Sänglerlaub (vgl. ebd., 103) eine große Verantwortung für den Journalismus und Lernbedarf, wie Themen weiterverbreitet werden. Diese Verantwortung lässt sich auf den Umgang mit

---

<sup>5</sup> Einen ausführlicheren Einblick in die aktuelle Nutzung audiovisueller Manipulationen zur politischen Einflussnahme bietet der Beitrag „Media Forensics in the Age of Disinformation“ von Justin Hendrix und Dan Morozoff (2022, 14ff.) der New York University, aber auch ein Blick auf die Ergebnisse des internationalen Rechercheprojekts Storykillers (vgl. Stark u. Zimmermann 20.02.23).



audiovisuellem Material übertragen, welches zum Teil erhebliche gesellschaftliche und politische Wirkung entfaltet. So wird beispielsweise dem Video im Zusammenhang der sogenannten Ibiza-Affäre eine „politische Sprengkraft“ zugeschrieben, welche in Österreich „die größte Staatskrise der Zweiten Republik“ (Krake 2020, 102) samt Regierungswechsel ausgelöst habe.<sup>6</sup>

Technologische Entwicklungen im Zuge der Digitalisierung haben die Bedeutung audiovisuellen Materials im Journalismus also auf der einen Seite noch verstärkt, stellen aber gleichzeitig deren Funktion als authentische Abbildung der Wirklichkeit infrage aufgrund der zunehmenden Schwierigkeit, ihre Authentizität sicherzustellen:

"Technology - once regarded as a boon to democracy - is now regarded by some as a threat to it." (Hendrix u. Morozoff 2022, 9)

---

<sup>6</sup> Beispiele dafür, dass manipulierte Bild- und Videoinhalte ihren Weg in die Medien finden, gibt es immer wieder. Bspw. zeigte der Nachrichtensender WELT im März 2022 ein Video, das vermeintlich den Abschuss von Hubschraubern in der Ukraine zeigt, tatsächlich aber aus einem Computerspiel stammt (vgl. Nowotny 2022).

### 2.3 Die Manipulation audiovisuellen Materials im digitalen Zeitalter

Technologische Entwicklungen vereinfachen nicht nur den Einfluss audiovisueller Manipulationen in der Gesellschaft, sondern erschweren teilweise auch ihre Verifizierung. So sind die Primärquellen oftmals nicht ersichtlich und mit der Verbreitung im Internet und besonders auf Social-Media-Plattformen geht in aller Regel ein Qualitätsverlust des Materials durch Komprimierungsschritte und eine verringerte Auflösung einher, wodurch Hinweise auf den Materialursprung und Spuren einer nachträglichen Bearbeitung stark in Mitleidenschaft gezogen werden (vgl. Cozzolino et al. 2022, 305f.; Lyu 2022, 325; Riess 2021, 344ff.; Mercier et al. 2019, 162f.). Natürlich kann die Verifizierung von Inhalten auch bewusst durch eine geringe Qualität des manipulierten Materials erschwert werden. Zudem haben veränderte Verfahren im Rahmen der Materialerstellung Einfluss auf die Möglichkeiten der Verifizierung. Daher gibt dieses Kapitel einen Überblick über die verschiedenen Möglichkeiten, audiovisuelles Material zu manipulativen Zwecken einzusetzen.

Anders als früher nehmen digitale Geräte inzwischen bereits während der Aufnahme erste Anpassungen des audiovisuellen Materials vor, besonders bei Fotos. Computergestützte Bildgebungsverfahren (engl. *Computational Imaging*) optimieren, was und wie der Sensor Licht aufnimmt und wie dieses Signal in der Software verarbeitet wird. Beispiele dafür sind Funktionen wie Panorama-Stitching, HDR-Aufnahmen mit mehreren Bildern oder ein Porträtmodus für geringe Schärfentiefe (vgl. McCloskey 2022, 41). Folgt man der Auslegung des Medienwissenschaftlers Winfried Gerling, gibt es daher kaum mehr unbearbeitetes Bildmaterial:

„Als unbearbeitet gilt, was die Optik herstellt, als Bearbeitung, was Software tut.“  
(Gerling 2022, 312)

Und er steht mit seiner Meinung nicht alleine da. Auch der Informatiker Dr. Christian Riess (vgl. 2021, 346f.) bemängelt eine technische Unschärfe des Begriffs Wahrheit durch den unterschiedlichen Umfang kamerainterner Nachbearbeitung. Genauso kann bei Videos argumentiert werden und dort kommt noch eine weitere Dimension dazu, die unter anderem vom Journalisten und Kommunikationswissenschaftler Prof. Dr. Klaus Meier angesprochen wird: „Die Manipulation in Bewegtbild beginnt schon vor

dem Deepfake. Letztlich werden Eindrücke durch Bewegtbild immer durch die Schnittabfolge erzeugt“ (Bayerische Akademie der Wissenschaften 2020, 5:44 min).

Die vorliegende Arbeit beschäftigt sich jedoch nur am Rande mit der Frage, was als Bearbeitung zu verstehen ist, inwiefern Material bearbeitet sein darf, um journalistisch verwendet zu werden und ab wann es als solches gekennzeichnet werden muss. Vielmehr geht es um die grundlegende Möglichkeit, besonders für Journalist:innen, Formen intentionaler Manipulationsformen zu erkennen.

Um die **Manipulation** audiovisuellen Materials zu definieren, wird auf die Auslegung des Kommunikationswissenschaftlers Klaus Forster zurückgegriffen, die sich auf Bildmaterial bezieht, jedoch auch auf audiovisuelle Inhalte übertragen werden kann: Die Manipulation audiovisuellen Materials bezeichnet somit eine „mit einer Täuschungsabsicht verbundene intentionale Veränderung von Informationen durch Auswahl, Zusätze oder Auslassungen“ (Forster 2003, 66). Im Kontrast dazu wird audiovisuelles Material, das keiner intentionalen Manipulation unterliegt, im weiteren Verlauf der Arbeit als authentisch bezeichnet. Technisch manipuliertes Material, das sich zum Beispiel in einem parodierenden oder satirischen Hintergrund als solches zu erkennen gibt, wird durch seine fehlende Täuschungsabsicht für Redaktionen nicht zu einer Herausforderung werden. Es lässt sich jedoch genauso einer der folgenden Manipulationsformen zuordnen.

Grundsätzlich herrscht in der Literatur Einigkeit darüber, auf welche Art und Weise audiovisuelles Material manipuliert werden kann. Eine einheitliche Systematisierung der verschiedenen Manipulationsformen existiert jedoch nicht. Es finden sich Kategorisierungen mit einem inhaltlichen Fokus auf die Wirkungsweise bzw. Zielsetzung der Manipulationsform (vgl. Papadopoulou et al. 2019, 200f.) und Kategorisierungen mit einer technischen Differenzierung. Die folgende Kategorisierung in vier Manipulationsformen ist das Ergebnis eines Abgleichs der unterschiedlichen Kategorisierungen nach derjenigen, die mit einem Fokus auf die Verifizierungsmöglichkeiten am sinnvollsten erscheint (vgl. Spangenberg u. Teyssou 2019, 13; Reuters). Natürlich können die unterschiedlichen Formen auch miteinander kombiniert werden.

|                                   |  |
|-----------------------------------|--|
| <b>Dekontextualisierte Medien</b> | Authentisches Material, das (un)bearbeitet in einen neuen Kontext eingebettet wird   |
| <b>Bearbeitete Medien</b>         | Material, bei dem Bild- bzw. Toninhalte entfernt, versteckt, kopiert oder hinzugefügt worden sind                          |
| <b>Inszenierte Medien</b>         | Zu diesem Zweck eigens produziertes Material   |
| <b>Synthetische Medien</b>        | Mithilfe von KI erzeugtes Material, das entweder vollständig computergeneriert ist oder auf authentischem Material basiert |

Tabelle 1: Überblick über Manipulationsformen (Grafik: Martha Peters)

Die ersten drei Manipulationsformen zählen zu den sogenannten *Shallowfakes*, mitunter auch als *Cheapfakes* bezeichnet.

### 2.3.1 *Shallowfakes / Cheapfakes*

Unter den Begriffen *Shallowfakes* oder *Cheapfakes* werden alle traditionellen Manipulationsformen zusammengefasst – also alle audiovisuellen Manipulationen, deren Erstellung ohne KI-Prozesse auskommt (vgl. Verdoliva 2020, 911; Johnson 26.03.19; Cozzolino et al. 2022, 281).

Am stärksten verbreitet sei **dekontextualisiertes Material** (vgl. Pörzgen 19.08.22). Möglicherweise auch deshalb, weil diese Manipulationsform weder großen Aufwand noch technische Hilfsmittel benötigt (vgl. Mezaris et al. 2019, vi; Apostolidis et al. 2019, 54). Unter dekontextualisierten Medien versteht die Nachrichtenagentur Reuters sowohl Material, das mit einem anderen Ereignis verbunden wird, als es tatsächlich darstellt, als auch Material, welches dadurch irreführend wirkt, dass es einen sehr selektiven Teil eines Ereignisses zeigt und damit aus dessen Kontext gerissen wird. Einen Spezialfall, auf den die BR-Redakteurin Janina Lückoff hinweist, stellt authentisches Material dar, dessen Aufnahmebedingungen korrekt genannt werden und die daraus nahegelegte Folgerung trotzdem falsch ist, da eine plausible Erklärung für die dargestellte Situation existiert, die aber nicht mitgeliefert wird (vgl. PresseClub München 07.04.22, 44:40min). Hier kann nur eine klassische journalistische Recherche zur

dargestellten Thematik helfen, da das Material für sich mitsamt dem genannten Aufnahmekontext authentisch ist.

Besonders aufwendig hingegen ist die Herstellung **inszenierten Materials**. Mit der Absicht täuschendes Material zu erstellen, wird eine Situation geplant, (selektiv) aufgenommen und dann verbreitet (vgl. Reuters): Inszenierungen sind gewissermaßen „Manipulationen, die vor der Kamera stattfinden“ (Büllesbach 2008, 125).

Als dritte traditionelle Manipulationsform ist **die Bearbeitung vorhandenen Materials** mit klassischen Mitteln zu nennen. Vom Aufwand her bewegen sich diese Manipulationsversuche meist zwischen Dekontextualisierung und Inszenierung, weisen jedoch eine große Bandbreite auf. Zunächst einmal kann audiovisuelles Material zugeschnitten werden. Die Auswahl eines selektiven Ausschnitts, bei dem der situative Rahmen vor und nach dem verbreiteten Moment entfernt wird, kann noch als Dekontextualisierung gesehen werden. Dagegen sind Veränderungen, für die Einzelbilder entfernt werden, Passagen dupliziert bzw. ergänzt werden oder auch ein bildlicher Zuschnitt klar als Bildbearbeitung zu verstehen. Bildmaterial kann zudem anders ausgerichtet bzw. gedreht werden und genauso können Lichtverhältnisse, Farben oder Struktur partiell oder in ihrer Gesamtheit verändert werden. Einzelne Bildelemente können entfernt bzw. eigentlich durch eine Ausweitung des Hintergrunds ersetzt (*inpainting*), dupliziert (*copy-move*) oder aus anderem Bildmaterial ergänzt (*splicing*) werden (vgl. Reuters; Verdoliva 2020, 911). Ton- oder Videodokumente können außerdem in ihrer Geschwindigkeit verändert werden. So entstanden bspw. die verbreiteten Videos der Sprecherin des US-Repräsentantenhauses Nancy Pelosi, die sie betrunken erscheinen lassen, durch eine simple Reduzierung der Framerate (vgl. Verdoliva 2020, 911).

Allerdings können Bearbeitungen auch ohne manipulative Intention erfolgen. Wichtig ist also, ob die Art der Bearbeitung wichtige inhaltliche Aspekte verändert, um eine verfälschte Darstellung von Ereignissen oder eine alternative Realität dazustellen (vgl. Reuters).

### 2.3.2 Synthetische Medien / Deepfakes

"Over the last several years, the rise of machine learning-based technologies has dramatically lowered the skill necessary to create compelling manipulations."  
(Sencar et al. 2022, 3f.)

Der folgende Abschnitt beschäftigt sich mit der vierten möglichen Manipulationsform in Form synthetischer Medien, auch als *Deepfakes* bekannt. Der Begriff Deepfake bezeichnet eine mithilfe von künstlicher Intelligenz erzeugte (Bewegt)Bild-, Ton- oder inzwischen mitunter sogar Textdatei, die authentisch wirkt, ohne es zu sein. Oft werden auch die verwendeten Methoden und Techniken zur Erstellung des Deepfakes unter dem Begriff zusammengefasst (vgl. Bendel 2020). Zurückzuführen ist die Bezeichnung Deepfake auf das Pseudonym eines Reddit<sup>7</sup>-Nutzers, der Ende 2017 unter diesem Namen einige der ersten Deepfakes veröffentlichte. Die Bezeichnung Deepfakes begann sich zu etablieren – vermutlich auch deshalb, weil der Begriff als Kofferwort aus *Deep Learning* (der dahinterstehenden Technologie) und *Fake* (Fälschung) durchaus als treffende Beschreibung verstanden werden kann.

Inzwischen wird der Begriff auch für die Produkte von Text-to-Image- bzw. Text-to-Video-Generatoren (T2I, T2V) angewandt, deren Ziel keine veränderten, sondern plausible neuartige Inhalte sind. Mehrere Systeme ermöglichen bereits eine mehr oder weniger überzeugende Bild- und auch Videosynthese basierend auf Texteingabeaufforderungen (vgl. Hendrix u. Morozoff 2022, 13f.). Seit der Vorstellung von *DALL-E* durch *OpenAI* im Januar 2021 (vgl. Ramesh et al. 05.01.21), dem ersten T2I-System, das in der Öffentlichkeit Aufmerksamkeit durch die Fähigkeit erlangte, komplexe Textbeschreibungen in annähernd realistisch aussehende Bilder umzuwandeln, hat sich nicht nur die Qualität der Ergebnisse deutlich verbessert, sondern auch die Anzahl der Anbieter vervielfacht. Im März 2023 hat nun auch Adobe (vgl. 04.05.23) einen T2I-Generator vorgestellt, der langfristig in die Programme der Adobe Cloud integriert werden soll, die Möglichkeiten generativer KI-Systeme mit klassischen Bildbearbeitungsprogrammen wie Adobe Photoshop verbindet und damit die Grenzen zwischen authentischen und synthetischen Inhalten weiter verschwimmen lässt. Seit letztem

---

<sup>7</sup> Reddit ist ein Social-News-Aggregator, eine Website, auf der registrierte Benutzer Inhalte einstellen und kommentieren können (siehe [www.reddit.com](http://www.reddit.com)).

Jahr gibt es zudem bereits erste Videogeneratoren, die aus Textbeschreibungen kurze Video-Clips erzeugen. Bisher sind die Ergebnisse der Videogeneratoren allerdings noch klar von authentischem Material unterscheidbar. Unklar ist, wie lange noch.

Das Besondere an synthetischen Medien ist der hohe Automatisierungsgrad, mit dem die Manipulationen erstellt werden können (vgl. Kietzmann et al. 2020, 138). Denn der Erfolg von Manipulationen hängt seit jeher vor allem von zwei Faktoren ab: Es stehen die verfügbaren Ressourcen und der mögliche Aufwand für die Manipulationserstellung der Zeit gegenüber, die für Verarbeitung und Überprüfung der Manipulation aufgewendet werden muss. Der technologische Fortschritt der letzten 30 Jahre hat diese Dynamik jedoch verändert. Nicht nur der Aufwand für die Verbreitung hat sich durch das Internet und Social Media stark verringert, sondern auch die (massenhafte) Erstellung manipulierter Inhalte hat sich durch KI-basierte Technologien stark vereinfacht (vgl. Hendrix u. Morozoff 2022, 8) – und man mag noch hinzufügen, auch zu einem ganz neuen Niveau an Manipulationen geführt.

Noch vor zehn Jahren war die Manipulation von Videomaterial, abgesehen von trivialen Operationen auf Bildebene, nur im Rahmen von Filmproduktionen möglich (vgl. Sencar et al. 2022, v). Die bisherigen Animationsprodukte in Filmen und Serien werden als *Computer Generated Imagery* (CGI) bezeichnet und sind ein klassisches Beispiel für aufwendige audiovisuelle Manipulationen. Sie bestehen aus 3D-Gerüsten, deren Generierung nicht nur mit einem enormen Aufwand verbunden ist (vgl. Verhoeven 2020, 233), sondern auch Expertise durch eine umfangreiche Ausbildung, teure Hardware und spezielle Software erfordert (vgl. Kietzmann et al. 2020, 136). Im Gegensatz dazu existiert für die Erstellung von Deepfakes mittlerweile kostenlose Software mit einer Benutzeroberfläche, die es Menschen erlaubt, qualitativ hochwertiges Material zu generieren, ohne tiefgehende technische Kenntnisse oder besonderes Equipment zu besitzen (vgl. Burkell u. Gosse 2019; Schade 2019, 95). Technologische Fortschritte verringern die notwendigen Fähigkeiten sowie die benötigte Zeit und Rechenleistung ständig weiter (vgl. Hendrix u. Morozoff 2022, 11).

In populärwissenschaftlichen Veröffentlichungen über Deepfakes werden häufig *künstliche Intelligenz*, *maschinelles Lernen* und *Deep Learning* – synonym und ohne weiter-

führende Erläuterung – als Begründung für die neuen Möglichkeiten audiovisueller Manipulation genannt. Das ist nicht falsch, aber unpräzise. Tatsächlich steht künstliche Intelligenz (KI) zunächst einmal für „ein Teilgebiet der Informatik, das sich mit der Automatisierung von intelligentem Verhalten befasst“ (Kirste u. Schürholz 2019, 21). Mit dieser Zielsetzung hat sich Ende des 20. Jahrhunderts ein zentrales Teilgebiet der KI-Forschung herausgebildet, dessen Ansatz darin besteht, Computerprogramme aus ihren eigenen Erfahrungen ‚lernen‘ zu lassen: Maschinelles Lernen (ML). Die Basis dieses Lernprozesses besteht in der Analyse gegebener Daten (je nach Kategorie des maschinellen Lernens mit oder ohne Vorgaben), aus denen Modelle entwickelt werden, die schließlich auf unbekannte Daten derselben Art angewendet werden können. Indem die Programme ihre Entscheidungsregeln über eine Rückkoppelung an das Erlernete anpassen, können sie sich im Gegensatz zu statischen Programmen verbessern oder auch Aufgaben lernen, die für Menschen zu komplex sind, als dass sie den Lösungsweg programmieren können. Der Begriff *Deep Learning* (DL) schließlich definiert einen Teilbereich des maschinellen Lernens, der auf künstlichen neuronalen Netzen beruht. Ihr vielschichtiger Aufbau, welcher sich an den Netzstrukturen menschlicher Nervenzellen orientiert, ermöglicht es, besonders komplexe Sachverhalte abzubilden (vgl. ebd.). Wie der Begriff *Deepfake* bereits nahelegt, entstammt die neue Form audiovisueller Manipulation diesem Teilbereich.

In der Theorie lassen sich Deepfakes nach Darstellungsform (Foto, Audio, Video) und Manipulationsweise (Lip-syncing, Face-swapping, Voice-swapping etc.) differenzieren, wie es unter anderem Kietzmann et al. (vgl. 2020, 142) tabellarisch tun. In der kommunikationswissenschaftlichen und auch gesellschaftlichen Auseinandersetzung mit Deepfakes stehen jedoch die möglichen Auswirkungen im Vordergrund, wenn man die verschiedenen Formen zu einem möglichst authentischen Eindruck kombiniert.

Anhand der bisher populärsten Deepfake-Beispiele, den Gesichtsmanipulationen (vgl. Schade 2019, 98), lässt sich die grundlegende Funktionsweise der Technologie skizzieren: Ziel ist, die Mimik eines Gesichts auf ein anderes Gesicht übertragen. Als Resultat kann entweder eine erste Person anhand ihres Gesichts in die Situation einer zweiten Person übertragen werden oder aber man geht noch einen Schritt weiter und über-



trägt das Gesicht der ersten Person, auf welches die Mimik der zweiten Person übertragen worden ist, wieder in das ursprüngliche Videomaterial – die Variante mit den potenziell größeren Auswirkungen: So passt das generierte Gesicht mit dem Rest des Videos zusammen und der Person können in einem realistischen Umfeld Worte in den Mund gelegt werden, die sie nie gesagt hat. Wie das funktioniert, erläutern die Autoren des Aufsatzes *Deepfakes: Trick or treat?* (vgl. Kietzmann et al. 2020) detailliert: Wie bei Anwendungen des maschinellen Lernens üblich, ist der eigentlichen Erstellung eines Deepfakes ein Lernprozess vorgelagert: Dafür wird zunächst eine möglichst große Menge Bildmaterial des Gesichts benötigt, das später zu sehen sein soll. Genauso wird möglichst viel Bildmaterial von einem zweiten Gesicht gebraucht, welches später die Handlung vorgibt, die durch das präsentierte Gesicht ausgeführt werden soll. Die Bilder müssen eine ausreichende Vielfalt an Ausdrücken, Winkeln und Perspektiven beinhalten, aber gleichzeitig eine möglichst ähnliche Beleuchtungssituation aufweisen (vgl. Burkell u. Gosse 2019).

Bisher weist die Technologie noch einige Schwachstellen auf, die eine Identifizierung der künstlichen Produkte als Manipulationen erleichtert. Laut einem Artikel des österreichischen Fachmagazins für Kommunikator:innen, ExtraDienst, werden Stirn, Haare und Ohren bei Facial Reenactments bisher meist nicht ersetzt: „Mit gutem Grund haben die Datenforensiker, die das Ibiza-Video auf seine Echtheit prüften, sehr genau auf die Ohren von Herrn Strache geschaut. Und dass bisher keine Deepfake-Experimente mit Brillenträgern veröffentlicht wurden, wird wohl auch seinen Grund haben“ (Krake 2020, 106). Auch die authentische Darstellung im Detailbereich (Haut, Gesichtsbehaarung) sei noch schwierig (vgl. Lyu 2022, 327).

Doch können die aktuellen Schwachstellen und notwendigen Voraussetzungen nur als Momentaufnahme gesehen werden. 2022 führte eine Studie der University of California und der Lancaster University zu dem Ergebnis, dass KI-generierte Gesichter inzwischen so perfekt darstellbar sind, dass von etwas mehr als 300 ungeschulten Proband:innen weniger als die Hälfte der 128 gezeigten Gesichter korrekt als authentisch oder synthetisch zugeordnet wurden (vgl. Nightingale u. Farid 2022).

Zukünftig könnten die Abbildungen ganzer Körper synthetisch erstellt werden. Dies sei nur eine Frage der Zeit, meinen manche (vgl. Burkell u. Gosse 2019). Der Journalist und Medientrainer Peter Welchering argumentiert dagegen, dass der technische Aufwand dafür zu groß sei im Vergleich zu den vorhandenen technischen Ressourcen: „Würde der Körper des Protagonisten mit einbezogen, könnten Deep Fakes nur noch in wochenlanger Arbeit in Höchstleistungsrechenzentren hergestellt werden. Die Produktionskosten wären exorbitant“ (Welchering 2020, 30). Wäre die KI-basierte Manipulation ganzer Körper erst einmal möglich, könnten dadurch nicht nur Aussagen, sondern ganze Handlungen imitiert werden: „Man kann in einem Video jetzt schon jeden alles sagen, bald wahrscheinlich auch jeden alles machen lassen“ (Krake 2020, 110). Damit würde sich auch die Limitierung auf bestehendes Videomaterial, in welches das generierte Video eingepasst werden kann, aufheben (vgl. Schade 2019, 100).

Die meisten bisher bekannten Deepfakes sind entweder als Unterhaltung, Satire oder aber mit der Intention generiert und veröffentlicht worden, auf die möglichen Konsequenzen der Technologie aufmerksam zu machen.<sup>8</sup> Eine Ausnahme stellen wirtschaftskriminelle Täuschungsaktionen mithilfe synthetischer Audiodateien dar. So überwies bereits 2019 ein britisches Energieunternehmen mehr als 200.000 US-Dollar nach einem Anruf, bei dem die Überweisung mithilfe der synthetisch generierten Stimme des CEOs angewiesen wurde (vgl. Hendrix u. Morozoff 2022, 19).

Doch auch bei der Verbreitung von Desinformation spielen Deepfakes als Untersuchungsgegenstand eine zunehmend relevante Rolle (vgl. Schade 2019, 95). Audiovisuelle Manipulationen haben das Potenzial, die öffentliche Kommunikation in einem Maß zu beeinflussen (vgl. Riess 2021, 343), dass der Technologie große Auswirkungen für die Gesellschaft, deren Bedrohung (vgl. Schade 2019, 103) und schließlich sogar eine Bedrohung der Demokratie (vgl. Kietzmann et al. 2020, 143) nachgesagt werden. Die Integrität von Informationen sei gefährdet (vgl. Sencar et al. 2022, v). Deepfakes könnten zu einem mächtigen Werkzeug werden, wenn sie die Erinnerungen

---

<sup>8</sup> Eines der bekanntesten Beispiele ist ein vom amerikanischen Medienunternehmen *buzzfeed* generiertes Video, in dem der Comedian Jordan Peele dem ehemaligen US-Präsidenten Barack Obama eine Rede in den Mund legt, die dieser so nie halten würde: <https://youtu.be/cQ54GDm1eL0>, zuletzt geprüft am 13.07.23

der Menschen an Ereignisse stören und verändern (vgl. Diakopoulos 2018), ihre Meinung zu Personen des öffentlichen Lebens verfälschen und so möglicherweise gar den Ausgang einer Wahl oder auch den Aktienmarkt beeinflussen (vgl. Kietzmann et al. 2020, 145). Mit KI-basierten Audio- und Videomanipulationen können genauso Wissenschaftler:innen Aussagen in den Mund gelegt werden, die sie nie getätigt haben, wie Politiker:innen rufschädigende Äußerungen. Beides kann nachhaltigen Einfluss auf die Meinungsbildung in der Gesellschaft haben.

"At its worst, deep fakes can damage brands and careers, manipulate and impact the political process, and go as far as destroy people's lives or even cause wars."  
(Spangenberg u. Teyssou 2019, 10)

Doch auch andersherum ist eine Bedrohung durch Deepfakes möglich: Lässt sich nicht mehr nachvollziehen, ob ein audiovisuelles Dokument echt ist oder nicht, können Beweise für Fehlritte von Politikern in Zweifel gezogen und echtes Material als falsch deklariert werden (vgl. Diakopoulos 2018): „Videoaufnahmen hatten bisher eine hohe Beweiskraft. Durch Deepfakes wird sich das ändern“ (Krake 2020, 104). Schon jetzt werden authentische Videoaufnahmen in Frage gestellt und als Deepfake betitelt, z. B. von Anhängern Donald Trumps im Falle eines Videos, in dem Trump die Stürmung des Kapitols verurteilt (vgl. Reuters 11.01.21). Deepfakes selbst würden in der Manipulationslandschaft bisher jedoch eine deutlich kleinere Rolle spielen, als die vielen Bedrohungsszenarien nahelegen (vgl. Hendrix u. Morozoff 2022, 14f.). Der manipulative Einsatz von Deepfakes bestehe bisher vor allem in der Generierung künstlicher Persönlichkeiten – bzw. ihrer Profildfotos – zur Ausgestaltung von Social-Media-Accounts, die Desinformation verbreiten (vgl. Hendrix u. Morozoff 2022, 15ff.; Grimm 16.03.22).

Eine qualitative Befragung von 18 Informatiker:innen und Journalist:innen legt nahe, dass Deepfakes als Risiko erkannt werden: „Die Risiken liegen für die Befragten zum einen in der Schwierigkeit begründet, Deep Fakes als Manipulationen zu entlarven, zum anderen im Glaubwürdigkeitsverlust, den Redaktionen und Medienhäuser bei einer Weiterverbreitung von Deep Fakes erleiden“ (Graßl et al. 2022, 15). Die Technologie wird jedoch gleichzeitig als Chance begriffen: „Sie versetzt Redaktionen in die Lage, Bildmaterial vorzuproduzieren oder im Nachgang zu glätten und so z. B. »Ähs« zu streichen, fehlende Worte einzusetzen oder eine bessere Synchronisierung herzustellen

len.“ Ob ein solcher Einsatz von Deepfakes in Hinblick auf Außenwirkung aber auch die Möglichkeit, authentisches Material von Fälschungen zu unterscheiden, hilfreich ist, muss hinterfragt werden. Einig zeigt sich die Forschungsliteratur in Bezug auf die hohe Geschwindigkeit, mit der sich die Technologie hinter synthetischen Medienprodukten weiterentwickelt – und dem Druck, der daraus für die Medienforensik entsteht:

„Die Betrachtung von komplett synthetisierten Deep Fakes hat [...] klar gemacht, dass der Technologie des Deep Learnings derzeit scheinbar keine Grenzen gesetzt sind und potentiell alles erdenkliche fotorealistic und von jedem Nutzer synthetisiert werden kann.“ (Schade 2019, 103)

Allein in den nächsten drei bis fünf Jahren werde sich die Manipulationslandschaft so rapide verändern, dass sich medienforensische Methoden anstrengen müssten, um mit den Manipulationsmöglichkeiten Schritt zu halten (vgl. Hendrix u. Morozoff 2022, 18). Um die medienforensischen Methoden geht es im folgenden Kapitel.

## 2.4 Der technische Rahmen für d. Verifikation audiovisuellen Materials

Nach einem Blick auf das zunehmende Manipulationspotenzial bei audiovisuellen Inhalten und den einhergehenden Herausforderungen im Journalismus ist kaum verwunderlich, dass „Bildforensik und anverwandte datengestützte Methoden ein wichtiger Bestandteil aktueller bildjournalistischer Arbeit“ seien, wie der Medienwissenschaftler Winfried Gerling (2022, 296) schreibt. Eine visuelle Überprüfung durch einen Menschen reiche heute nicht mehr aus, wenn das Material aus unbekanntem oder unseriösen Quellen stamme, urteilen auch IT-Forensiker:innen (vgl. Halvani et al. 2020, 122). Man ist zusätzlich auf klassische Methoden journalistischer Recherche, aber auch auf technische Verfahren der Multimedia-Forensik angewiesen, um audiovisuelles Material zu verifizieren. Das Forschungsfeld der **Multimedia-Forensik** ist ein Teilbereich der Informatik, der sich mit den technischen Möglichkeiten beschäftigt, Ursprung und Authentizität audiovisueller Medien zu ermitteln (vgl. Riess 2021, 343ff.). Eine Übersicht und Systematisierung multimedia-forensischer Methoden, aber auch der Möglichkeiten klassischer Recherche, die bei der Verifizierung audiovisueller Inhalte helfen, ist ausgehend von existierenden Publikationen mit den Interviewpartner:innen erarbeitet worden und findet sich in Kapitel 4.2.1.

Für die Beantwortung der Frage, inwieweit künftig noch absolute Sicherheit über die Authentizität audiovisuellen Materials möglich sein wird, ist es hilfreich, zwischen **Fact-Checking** und **Verifikation** als unterschiedlichen journalistische Vorgehensweisen zu unterscheiden, wie es der ehemalige Direktor des *International Fact-Checking Network (IFCN)*, Alexios Mantzarlis, tut: Als Fact-Checking bezeichnet er die nachträgliche Untersuchung einer Behauptung von öffentlicher Relevanz auf Logik, Kohärenz und ihren ursprünglichen Kontext, wenn diese bereits in den Medien aufgegriffen wurde. Die Verifikation sieht Mantzarlis hingegen als Handwerk, Quellen wie Tonaufnahmen, Fotos oder Videos vorab auf ihre Echtheit zu überprüfen, bevor sie journalistisch für Veröffentlichungen eingesetzt werden. Während es beim Fact-Checking nach dieser Differenzierung in erster Linie um eine inhaltliche Einordnung bereits bekannter Behauptungen geht, handelt es sich bei der Verifikation um die Prüfung von Material auf dessen Authentizität, meist mithilfe technischer Werkzeuge (vgl. Mantzarlis 2015; Stern

2020, 120ff.). Geht es darum, inwieweit authentisches Material in Zukunft noch von Fälschungen unterschieden werden kann, stellt sich zunächst die Frage nach den technischen Verifikationsmöglichkeiten.

#### 2.4.1 *Der Weg multimedia-forensischer Methoden in den Journalismus*

Eine technische und möglichst systematische Verifizierung digitaler Inhalte Dritter ist für Journalist:innen und Redaktionen relativ neu (vgl. Spangenberg u. Teysou 2019, 4). Zu Beginn des Arabischen Frühlings sei es Medienschaffenden noch so gut wie unmöglich erschienen, online verbreitete Inhalte ohne vertrauenswürdige Quellen vor Ort zu prüfen (vgl. Bischoff 2022b, 64). Und auch Matt Cooke, ein ehemaliger BBC-Journalist, der inzwischen für das Google News Lab arbeitet, meint: „Factchecking back in 2011 was generally limited to picking up a phone, calling someone and seeing who would answer“ (IKSŽ 2022, 44:30min). Das hat sich inzwischen geändert.

„In Deutschland forscht und arbeitet vor allem die Deutsche Welle (DW) an Tools zur Verifikation“, konstatiert die Redakteurin und Faktencheckerin Jenny Stern (2020, 123). Sie bezieht sich damit auf die Innovationsabteilung der DW, die sich als Schnittstelle zwischen Journalismus, IT und Medienwissenschaft begreift und bereits seit 2012 zahlreiche Projekte durchführt, die mehr oder weniger direkt mit den technischen Möglichkeiten zu tun haben, audiovisuelles Material zu verifizieren.

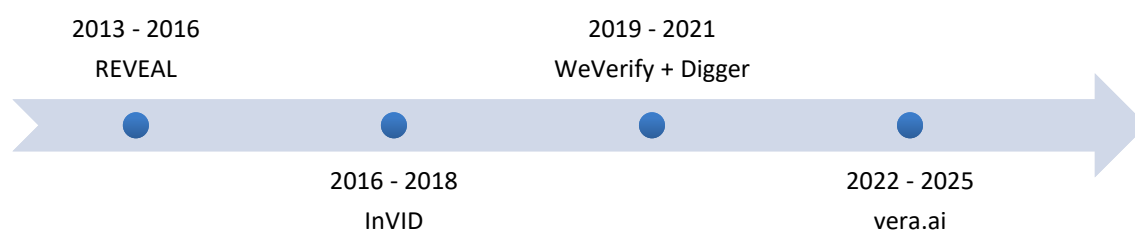


Abbildung 1: Projekte rund um die Verifizierung audiovisueller Inhalte unter Beteiligung der DW-Innovationsabteilung (Grafik: Martha Peters)

Von der Entwicklung einer Medienüberwachungsplattform über Projekte, die sich mit dem Potenzial der Blockchain-Technologie im Journalismus beschäftigen, hin zu Projekten, die sich spezifisch mit der Analyse von Desinformation und der Entwicklung (KI-gestützter) Verifizierungsinstrumente beschäftigen.

2019 betonten die Akteure des Projektes InVID noch, dass es einige Software gebe, die Faktenchecks oder die Verifizierung von Fotos unterstütze, jedoch gebe es kaum Software, die mit dem Ziel entwickelt worden ist, Videomaterial zu verifizieren. Stattdessen werde auf eine Vielzahl einzelner, unterschiedlicher Werkzeuge zurückgegriffen. So sind es vor allem Softwarelösungen der DW-Innovationsabteilung, die verschiedene Werkzeuge zur Verifizierung miteinander zu kombinieren versuchen. Ergebnisse vergangener Projekte sind die frei verfügbare Browsererweiterung *Fake news debunker by InVID & WeVerify* und die kollaborative Plattform *Truly Media* (vgl. DW Innovation; DW 28.09.22). Im Rahmen des aktuell laufenden Projektes *vera.ai (Verification Assisted by Artificial Intelligence)* sollen Plugin und Plattform durch KI-basierte Lösungen erweitert werden (vgl. Spangenberg).

Plattform und Plugin erleichtern die Abfrage, Analyse und Verifizierung vor allem nutzergenerierter Inhalte. Das Plugin wird von den Entwickler:innen beschrieben als „one of the major tools to debunk disinformation and verify videos and images“ (Mezaris et al. 2019, ix). Und auch im unter Journalist:innen verbreiteten *Verification Handbook* heißt es, InVID sei zusammen mit Bilder-Rückwärtssuchen das beste Hilfsmittel, um Fotos und Video zu untersuchen (vgl. Silverman u. Landesanstalt für Medien NRW 2020, 79). Der Ideengeber bezeichnet die Idee dahinter als Schweizer Taschenmesser der Verifikation. Die Browsererweiterung ist die Reaktion auf ein Projekt im Rahmen der französischen Präsidentschaftswahl 2017, an dem mehr als 30 Medienorganisationen gemeinsam an Faktenchecks und der Detektion von Desinformation arbeiteten und aufgrund dessen Möglichkeiten ausgemacht werden konnten, Verifikationsprozesse durch technische Werkzeuge effizienter zu gestalten. Inzwischen ist die Browsererweiterung durch das Folgeprojekt WeVerify weiter verbessert worden und die Menge der enthaltenen Werkzeuge sei inzwischen verdoppelt worden. Unter anderem wurde eine Assistenzfunktion ergänzt, welche die Auswahl der sinnvollen Werkzeuge erleichtert, zudem neue Elemente zur ganzheitlichen Analyse der Inhalte und Desinformationsströme Sozialer Netzwerke und eine dezentralisierte Datenbank mit bereits entlarvten Manipulationen. Ziel ist die Skalierung und Beschleunigung der Verifikationsprozesse (vgl. WeVerify; Marinova et al. 2020). Das Plugin ist kostenlos und für die Verwendung mit dem Chrome-Browser optimiert.

Die Plattform *Truly Media* wurde im Rahmen des Folgeprojektes von InVID, WeVerify, entwickelt, bei der die Organisation und Vernetzung im Zuge von Verifikationsprozessen im Vordergrund steht. Die Plattform erlaubt einen Datenaustausch in Echtzeit und ist mit selbst entwickelten, aber auch externen Tools verknüpft. Die Plattform ist außerhalb der DW inzwischen auch beim ZDF und bei Amnesty International im Einsatz (vgl. Stern 2020, 141).<sup>9</sup>

Es folgt ein Überblick über die grundsätzlichen, auch den Tools der DW zugrunde liegenden, Möglichkeiten und Technologien, die bei der Verifizierung audiovisueller Inhalte hilfreich sein können. Ein einheitliches Vorgehen bei der Verifizierung audiovisueller Inhalte ist bei dem aktuellen Stand der Technik nicht möglich. Stattdessen gibt es fünf Kernfragen, deren Bearbeitung u. a. von *First Draft*<sup>10</sup> und auch im *Verification Handbook*<sup>11</sup> vorgeschlagen wird und mithilfe derer man jeden Verifizierungsprozess (auch bei Audio- und Videodateien) strukturieren kann:

1. Sehen wir die Originalversion?
2. Wissen wir, wer es gemacht hat?
3. Wissen wir, wo es gemacht wurde?
4. Wissen wir, wann es gemacht wurde?
5. Und wissen wir, warum es gemacht wurde?

Bei der Beantwortung der Fragen helfen sowohl klassische Formen journalistischer Recherche als auch multimedia-forensische Verfahren. Alle nun folgenden – passiven – technischen Verifizierungsinstrumente haben eines gemeinsam: Ein tatsächlicher Beweis für die Authentizität audiovisuellen Materials ist nicht möglich. Schlussendlich lassen sich Inhalte nur falsifizieren oder aber Indizien für deren Authentizität sammeln (vgl. Welcherling 2020, 31; Primbs 2018, 118).

#### 2.4.2 Überprüfung des Publikationskontextes u. Prüfung früherer Verbreitung

„Möchte man herausfinden, ob eine Aufnahme echt ist, sollte man zuerst ihren Ursprung prüfen: Wo wurde sie erstmals veröffentlicht?“ (Bischoff 2022b, 66)

---

<sup>9</sup> Siehe <https://www.truly.media/>, zuletzt geprüft am 02.04.23

<sup>10</sup> Bei *First Draft* handelte es sich von 2015-2022 um eine unabhängige Non-Profit-Organisation, die sich auf die Bekämpfung von Falsch-/Desinformation im Internet konzentrierte. Gegründet wurde sie u. a. von der Google News Initiative und Bellingcat. Inzwischen wird die Mission als *Information Futures Lab* weitergeführt (vgl. First Draft; Urbani 22.09.22).

<sup>11</sup> Herausgegeben vom *European Journalism Centre*, geschrieben u. a. von Journalist:innen des BBC und als deutsche Übersetzung von der *Landesanstalt für Medien NRW* herausgegeben (vgl. 2020, 78).



Um die erste und wichtigste (vgl. Urbani 22.09.22) der fünf Kernfragen zu beantworten, kann man zunächst eine inverse Bildersuche (Image reverse search) durchführen. Von allen Methoden, die auf den folgenden Seiten noch beschrieben werden, ist dies eine Methode, die nahezu überall als erster Verifikationsschritt empfohlen wird. Lädt man bei geläufigen Suchmaschinen ein Bild hoch, werden KI-basiert verwandte Inhalte samt ähnlichen Bildern angezeigt. So können die gefundenen Bilder samt ihrem Publikationskontext häufig schon einen vermuteten Bildkontext widerlegen oder Anhaltspunkte für dessen Authentizität liefern. Besonders die Entkontextualisierung authentischen Materials kann so schnell enttarnt werden (vgl. Riess 2022; Gerling 2022, 305; Ratering 2022, 365; Sell u. Oswald 2021, 253; Mezaris et al. 2019, 6; Spangenberg 2014, 117).

Da jede Suchmaschine zu unterschiedlichen Ergebnissen führt, ist die Verwendung mehrerer Suchmaschinen sinnvoll<sup>12</sup>. Dafür gibt es inzwischen auch Tools, welche erlauben, mehrere Suchmaschinen gleichzeitig zu durchsuchen (vgl. Stern 2020, 131) bzw. anstelle des Bildes selbst die URL eines Bildes einzugeben.<sup>13</sup> Geht es um ein Video mit unterschiedlichen Szenen, muss dieses zunächst in einzelne Fragmente zerlegt werden. Dies ermöglicht zum einen eine Zusammenfassung des Videos, vor allem aber eine Verifizierung der Bestandteile des Videos auf Fragmentenebene (vgl. Mezaris et al. 2019, vi). Auch dafür wurden in den letzten Jahren Softwarelösungen entwickelt, so dass nicht mehr mit manuell erstellten Screenshots gearbeitet werden muss. Besonders die im Rahmen des Projektes InVID entwickelte Software erlaubt nach den Beteiligten eine sehr effiziente Zerlegung auch lokal gespeicherter Videos in die Anzahl von Einzelbildern, welche zur bildlichen Erfassung aller Szenen benötigt wird (vgl. Apostolidis et al. 2019, 55f.). Zudem sei ein automatisierter Abgleich mit bereits analysiertem Material implementiert worden, sowie eine Vergleichsmöglichkeit, die es erlaubt, zwei Videos parallel abzuspielen und anzuhalten (vgl. Fricke u. Thomsen 2019, 325).

---

<sup>12</sup> Zudem habe jede Suchmaschine andere Vorteile: Google zeige oft direkt das Datum einer Seite an, Bing könne gut Bildausschnitte vergrößern, Yandex sei besonders gut für Osteuropa und verfüge zudem über die beste Gesichtserkennung, während man bei TinEye direkt nach der ältesten und der am meisten bearbeiteten Version eines Fotos suchen könne (vgl. Sell u. Oswald 2021, 254).

<sup>13</sup> Z. B. RevEye, Karma Decay für Reddit Inhalte etc.

Bedroht werde die inverse Bildersuche durch die schnelle Entwicklung von T2I-Technologien, die es schon jetzt erlauben, eine unendliche Anzahl von Inhaltsvariationen zu synthetisieren. Noch sei sie jedoch eine der ganz wenigen zuverlässigen Techniken für Faktenchecker:innen (vgl. Hendrix u. Morozoff 2022, 11).

Wird das Material nicht bereits durch die inverse Bildersuche falsifiziert, kann eine Überprüfung des (Publikations-)kontextes und des Urhebers helfen. Auch ein Abgleich mit anderen Quellen, zum Beispiel weiteren Bildern der gleichen Situation oder Augenzeugen, kann unterstützen. Außerdem sollte die Analyse der Metadaten nicht vernachlässigt werden, auch wenn sie gefälscht werden können (vgl. Khan et al. 2023, 6757) und in aller Regel beim Upload von Inhalten auf Social Media entfernt werden (vgl. Bischoff 2022b, 67; Verdoliva 2020, 911). Die Frage, warum eine Aufnahme gemacht wurde, kann jedoch häufig nur durch direkten Kontakt mit der eigentlichen Quelle beantwortet werden. Dafür ist es notwendig herauszufinden, ob die Person, die den Inhalt verbreitet, auch diejenige ist, die ihn erstellt hat. Von der eigentlichen Quelle lassen sich auch die Umstände der Aufnahme (Ort, Zeit, Anlass) und das Originalmaterial erfragen, womit sich die Aussagen der Person durch weitere Plausibilitäts-Checks untermauern lassen (vgl. Khan et al. 2023, 6751f.; Sell u. Oswald 2021, 254; Papadopoulou et al. 2019, 191; Spangenberg 2014, 116; Diplaris et al. 2012, 1244).

Auch zur Überprüfung des Publikationskontextes können Softwarelösungen hilfreich sein. So haben Wissenschaftler:innen des Centre for Research and Technology Hellas im Rahmen des InVID-Projekts 2016 ein Werkzeug entwickelt, das den Online-Kontext um einen Videobeitrag auf YouTube, Facebook oder Twitter sammelt, filtert und zusammenfasst, um der verifizierenden Person mit einer übersichtlichen Darstellung die Prüfung und Analyse der verfügbaren Daten zu erleichtern. Das entsprechende Werkzeug wird als *Context Aggregation And Analysis (CAA) Service* vorgestellt<sup>14</sup>. Es sammelt Informationen über das Video und den uploadenden Account (Videotitel, Video- und Kanalbeschreibung, Erstellungszeit, ggf. möglicher Erstellungsort und Kommentare), aber auch Tweets, die das Zielvideo teilen, und, falls von der verifizierenden Person angegeben, Wetterinformationen zum Zeitpunkt und Ort, an dem das Ereignis angebe-

---

<sup>14</sup> <https://caa.iti.gr/>, zuletzt geprüft am 23.03.23

lich stattfand. Daraus erstellt es einen Bericht mit den Metadaten, einer Twitter-Zeitleiste sowie den Wetterdaten. Kommentare werden anhand von definierten aber auch selbst wählbaren Schlüsselwörtern direkt nach nützlichen Informationen für die Entscheidung über die Authentizität des Videos gefiltert. Je mehr Kommentare Schlüsselworte enthalten, welche die Echtheit des Materials in Frage stellen, umso wahrscheinlicher ist der Inhalt unglaubwürdig. Darüber hinaus gibt das Programm Links aus, die eine Bilder-Rückwärtssuche einfach machen. Bei Twitter-Videos wird zudem die Einschätzung des Tools *Tweet Verification Assistant* (fake/real) in den Bericht aufgenommen. Später wurde das Werkzeug noch um einen Abgleich mit der InVID-Datenbank an bereits verifiziertem und falsifiziertem Material erweitert, wonach bei einem Treffer bereits vermerkte Informationen in den Bericht aufgenommen werden (vgl. Papadopoulou et al. 2019, 210ff.). Das Resümee der Tool-Programmierer:innen:

"To sum up, the CAA tool does not provide a final decision and does not label the video as fake or real. It creates a verification report that the user should take into account, evaluate its different pieces of information and make the final decision [...] The fact that popular Internet platforms such as YouTube and Facebook are in the position to control who has programmatic access to data that is otherwise publicly available makes it very challenging to build automated solutions and tools that could help mitigate the problem of disinformation."

(Papadopoulou et al. 2019, 216ff.)

Die Wissenschaftler:innen vom *Centre for Research and Technology Hellas* stellten bei dem Abgleich ihrer Sammlung an verifiziertem und falsifiziertem Material mit den erhobenen Publikationsdaten fest, dass manipulierte Inhalte in der Regel deutlich kürzer seien als authentisches Material. Zudem seien YouTube-Kanäle und Twitter-Accounts mit gefälschtem Material im Durchschnitt deutlich jünger als solche, die keine gefälschten Inhalte verbreiten und hätten weniger Accounts, die ihnen folgen. Auch sei die Aktivitätshäufigkeit eines Kanals für echte Videos erheblich größer als die von Kanälen mit gefälschten Inhalten. Einen Unterschied in der sprachlichen Qualität und Emotionalität der Video-zugehörigen Textbeiträge konnten sie hingegen nicht ausmachen. Festhalten konnten sie jedoch, dass gefälschte Videos in ähnlicher Form mit einer viel höheren Rate über einen viel längeren Zeitraum gepostet werden, als das bei authentischem Material der Fall ist. Auch im Kommentarverhalten sind laut den For-

schers:innen Unterschiede zwischen gefälschten und authentischen Inhalten zu sehen (vgl. Papadopoulou et al. 2019, 205ff., 213).

Da audiovisuelle Inhalte, wie bereits angesprochen, selten ohne Kontext verbreitet werden, kann auch der Einsatz textkonzentrierter Verfahren zur (automatisierten) Erkennung von Desinformation sinnvoll sein. Einen Überblick über die Forschungslandschaft dort findet sich beispielsweise bei Katrin Hartwig und Christian Reuter (vgl. 2021) von der TU Dortmund.

### *2.4.3 Inhaltliche Überprüfung*

Obwohl die Analyse der Metadaten und auch die inverse Bildersuche als Methoden der Multimedia-Forensik angesehen werden können, wie die Experteninterviews im empirischen Teil dieser Arbeit ergeben (siehe Kapitel 4.2.1), handelt es sich bei den bisher vorgestellten Methoden größtenteils um Möglichkeiten klassischer Recherche, die wenig spezifische Software und technologisches Know-how voraussetzen. Im Folgenden geht es hingegen um Verfahren, die beides stärker voraussetzen und als „klassische Verfahren der Multimedia-Forensik“ (Maier u. Riess 2023, 215) bezeichnet werden. Dabei werden die Verfahren grob in zwei Kategorien unterteilt: in Verfahren, die auf dem in aller Regel sichtbaren Inhalt des Materials beruhen und solchen, die das Material auf Pixelebene analysieren (vgl. ebd.). Zunächst geht es um Möglichkeiten der inhaltlichen Überprüfung.

Ein Aspekt, der zwischen diesem und dem vorherigen Kapitel steht, weil er sowohl mit der Überprüfung des Publikationskontextes zu tun hat, als auch auf der Auswertung des Materials selbst beruht, ist die Identifizierung vorhandener Quellenangaben wie Logos. So weisen die Wissenschaftler Foteini Markatopoulou et al. vom *Centre for Research and Technology Hellas* darauf hin, dass es bei über 27.000 TV-Sendern weltweit oder auch unzähligen, wechselnden paramilitärischen Gruppen eine große Herausforderung sei, solche Quellenangaben in visuellem Material zu- und einzuordnen. Eine automatisierte Identifizierung des Logos und Bereitstellung zugehöriger Informationen könne die Bildanalyse deutlich beschleunigen (vgl. Markatopoulou et al. 2019, 129). Ein solches „logo detection system“ ist in der Folge auch in die InVID-Plattform integriert worden, allerdings enthält die zugehörige Datenbank bisher bloß die Nachweise

169 unterschiedlicher Quellen, kann aber von den Nutzer:innen selbstständig erweitert werden (vgl. Markatopoulou et al. 2019, 152–154; Fricke u. Thomsen 2019, 332).

Aber auch über Quellenangaben und Logos hinaus kann eine genauere Betrachtung des Materialinhalts bei der Verifizierung helfen, um Indizien für den Ort und die Zeit der fraglichen Aufnahme zu sammeln. Informationen für eine **Geolokalisierung** lassen sich beispielsweise über den Abgleich der abgebildeten Elemente und Umgebung mit anderen Aufnahmen der angeblichen Örtlichkeit, Straßenkarten, Satellitenbildern oder weiteren Datenbanken (bspw. von Straßenschildern, Flügen oder Reedereien) finden. Auch ein Vergleich des gesprochenen Dialekts kann helfen (vgl. Spangenberg 2014, 116). Nach einer erfolgten Geolokalisierung können (jahreszeitbedingte) Veränderungen, Witterungsbedingungen und Sonnenstand im Abgleich mit weiteren Daten eine **Chronolokalisierung** ermöglichen (vgl. Bischoff 2022b, 67ff.; Riess 2022, 208f.).

Unter dem Oberbegriff der **physikbasierten Verfahren** werden Methoden zusammengefasst, welche Bild- oder Videoinhalte auf physikalische Unstimmigkeiten hin analysieren. Vor allem die Wechselwirkung von Objekten und Licht steht dabei im Zentrum, um Hinweise auf die Konsistenz geometrischer und photometrischer Bildeigenschaften zu geben. Es wird die perspektivische Projektion einzelner Bildelemente auf seine Plausibilität hin geprüft, genauso wie Helligkeits- und Farbvariationen, auch unter Berücksichtigung möglicher Reflexionen innerhalb des Bildes (vgl. Riess 2022, 207ff.). Besonders einfach ist die perspektivische Überprüfung von Bildmaterial bei Szenen, zu denen weiteres Bildmaterial vorliegt oder die Text einer üblichen Schriftart enthalten. Dann kann das weitere Bildmaterial oder die Schriftart als Referenz verwendet werden. Eine Plausibilitätsprüfung anhand der Lichtsituation ist besonders einfach, wenn es nur eine einzige Lichtquelle im Bild gibt. Dann bietet sich die Überprüfung der Schatten an. Allerdings ist es bei einer solchen Lichtsituation auch vergleichsweise einfach, eine authentisch wirkende Manipulation zu erzeugen. Auch Lichtfarben können auf Konsistenz überprüft werden. Denn forensische Analysemethoden gehen davon aus, dass der Weißabgleich der Kamera eine globale Bildtransformation ist, die keine lokalen Farbinconsistenzen einführt. Auch Kamerabewegungen können laut Riess berücksichtigt werden, solange statische Objekte im Bild als Referenz dienen können. Unter dem

Ausschluss äußerer und unsichtbarer Einflüsse könne sogar die Konsistenz der Bewegung eines Objekts überprüft werden. Finden sich lokal begrenzte Unterschiede in der Perspektive und/oder der Lichtsituation, weist das darauf hin, dass das Bildmaterial zusammengesetzt oder zumindest bearbeitet wurde: „Geübte Beobachter\*innen können ohne Anwendung digitaler Methoden solche Inkonsistenzen im Bild erkennen, die dann algorithmisch überprüft werden sollten“ (Gerling 2022, 310; vgl. Riess 2022, 207ff., 219ff., 225ff.). In der journalistischen Praxis sei die Herangehensweise aktuell der Goldstandard, um Multimedia-Inhalte ohne viel Kontext zu verifizieren, brauche bei den derzeitigen technischen Möglichkeiten aber viel Zeit (vgl. Riess 2022, 208f.)

Die Anwendbarkeit physikbasierter Methoden ist jedoch auf Szenen beschränkt, bei denen eine Analyse anhand der vorgestellten Aspekte möglich ist. Zudem würden die meisten Methoden den manuellen Eingriff einer Analytiker:in erfordern, um die Anwendbarkeit der Methode für die Bildszene einzuschätzen, sogenanntes Weltwissen bereitzustellen und zu analysierende Bereiche auszuwählen. Einige der beschriebenen Methoden setzen auch die Kenntnis der Brennweite bei der Aufnahme voraus. Gleichzeitig seien physikbasierte Methoden meist unabhängig von Bild- oder Videoqualität und robust gegenüber vorangegangenen Bearbeitungen, was sie besonders attraktiv für die Analyse von Inhalten geringer Qualität oder sogar analoger Inhalte macht. Außerdem sind physikbasierte Methoden von Natur aus erklärbar durch die Verifizierung ihrer zugrunde liegenden Modelle. Aus diesen Gründen arbeiten Journalist:innen viel mit physikbasierten Methoden (vgl. Riess 2022, 210, 230f., 2021, 348f.).

Die Leistungsfähigkeit physikbasierter Methoden werde für die Zukunft vor allem durch zwei technologische Entwicklungen herausgefordert: Das eine sei die zunehmend weitreichendere Bildverarbeitung in Smartphones, welche die physikalisch begrenzte Kameraoptik softwareseitig auszugleichen versucht. Wenn das Bild selbst jedoch nicht mehr allein das Ergebnis eines physikalischen Aufnahmeprozesses ist, werde damit auch die Gültigkeit physikalisch basierter Modelle für die forensische Analyse grundsätzlich in Frage gestellt. Die zweite Herausforderung seien computergenerierte, physikalisch plausible Szenen. Der Einfluss beider Entwicklungen auf die Leistungsfähigkeit physikbasierter Multimedia-Forensik sei noch offen (vgl. Riess 2022, 231).

Eng verwandt mit den physikbasierten Methoden ist nach Riess (vgl. ebd., 210) die Inhaltsanalyse auf **Verhaltenshinweise der Personen** und **physiologische Merkmale**.

#### 2.4.4 Methoden auf Pixelebene

Bei pixelbasierten Verfahren wird Bildmaterial dagegen auf Unstimmigkeiten analysiert, die mit dem bloßen Auge in der Regel unsichtbar sind. Solche Artefakte entstehen üblicherweise sowohl bei der Bilderstellung, als auch bei anschließenden Bearbeitungs- oder Kompressionsverfahren (vgl. Maier u. Riess 2023, 215). Jedes Bild (oder Videoframe) kann dabei zunächst als 2D-Anordnung von RGB-Werten angesehen werden, wobei die Farben an sich für den Verifikationsprozess häufig gar nicht so entscheidend sind. Hinweise auf eventuelle Manipulationen können insbesondere **Auffälligkeiten im Bildrauschen und der Schärfe des Bildes** geben. Als Bildrauschen wird die zufällige Variation von Helligkeits- oder Farbinformationen bezeichnet, die im Allgemeinen eine Kombination aus den physikalischen Eigenschaften des Aufnahmegeräts und einer verlustbehafteten Bildkomprimierung ist. Der Begriff Schärfe ist intuitiver zu verstehen, setzt sich jedoch ebenfalls aus mehreren Komponenten in Form von Fokus, Sichtbarkeit und Bildqualität zusammen (vgl. Mercier et al. 2019, 164).

Mitunter kann es zur Verifizierung eines Fotos hilfreich sein zu wissen, **mit welchem Gerät**, also mit welcher Kamera oder welchem Smartphone, es aufgenommen wurde. Gleich 4 von 16 Kapiteln des Sammelwerks *Multimedia Forensics* (Sencar et al. 2022) beschäftigen sich mit der Identifizierung der Kameraquelle bei fotografischen Erzeugnissen. Dabei muss unterschieden werden, ob der Ursprung eines Fotos auf ein Kameramodell oder tatsächlich auf eine einzelne Kamera zurückgeführt werden soll. Gemeinsam ist den vier Methoden das zugrundeliegende Prinzip: „So erzeugt jeder Sensortyp ein spezifisches Rauschen, das als Muster erkannt werden kann. Hinzu kommen algorithmische Korrekturen wie z. B. Entzerrungen, die hersteller- oder kameratypisch sind“ (Gerling 2022, 309). Im Einzelnen:

Bilder, die mit dem gleichen **Kameramodell** aufgenommen wurden, weisen eine gemeinsame Sammlung an Artefakten auf, die es theoretisch möglich macht, sie anhand einer Pixelanalyse von Erzeugnissen anderer Modelle zu unterscheiden. Gründe dafür liegen in der unterschiedlichen physikalischen Beschaffenheit von Sensoren und Objek-

tiven genauso wie in der softwareseitigen Art und Weise der Bildverarbeitung innerhalb des Aufnahmeapparates (vgl. Mandelli et al. 2022, 133f.). Praktisch sei der Erfolg dieser Methode bislang jedoch sehr abhängig vom Grad der Nachbearbeitung, die das Foto durchlaufen hat. So würden Komprimierungsartefakte und andere Bearbeitungsformen, wie das Zuschneiden eines Fotos oder die Veränderung seiner Größe, den Prozess stark beeinflussen. Aber auch computergestützte Bildgebungsverfahren innerhalb des Aufnahmeprozesses – wie das Hinzufügen synthetischer Unschärfe im Porträtmodus vieler Smartphones – komplizieren laut der Informatikerin Sara Mandelli (vgl. 2022, 167f.) die Modellidentifizierung.

Mit der Unterscheidung **synthetisch erzeugter Unschärfe** von optischer Unschärfe beschäftigt sich der Informatiker Scott McCloskey (vgl. 2022) in seinem Beitrag zum Sammelwerk *Multimedia Forensics*. Durch die limitierten optischen Möglichkeiten von Smartphones sei synthetisch erzeugte Unschärfe dort inzwischen Standard. Es sei möglich, diese von optisch erzeugter Unschärfe zu unterscheiden. Allerdings könnten zukünftige computergestützte Bildgebungsverfahren die Wirksamkeit klassischer forensischer Techniken weiter in Frage stellen (vgl. ebd., 60).

Als erfolgreich, sehr effektiv und bislang unübertroffen wird eine Methode beschrieben, die es erlaubt, ein Foto seinem spezifischen Aufnahmegerät zuzuordnen. Der Informatiker Christian Riess (2021, 347) bezeichnet die Methode sogar als das „derzeit vielleicht stärkste multimedia-forensische Verfahren“. Dabei geht es nicht um die Feststellung eines Kamera- oder Smartphone-Modells, sondern um den Beleg, dass ein Foto aus einem bestimmten Aufnahmegerät stammt. Möglich geworden sei dies durch eine Entdeckung im Jahr 2005, die als **Photo Response Non-Uniformity (PRNU)** bezeichnet wird und auch als Fingerabdruck eines Bildsensors beschrieben werden kann. Man macht sich zunutze, dass jeder Bildsensor durch unvermeidliche Schwankungen im Herstellungsprozess physikalisch einzigartig ist und allen von ihm aufgenommenen Bildern ein gewisses Maß an Rauschen hinzufügt – leichte Schwankungen in der Lichtempfindlichkeit einzelner Pixel, selbst wenn die Sensorebene absolut homogen beleuchtet wurde. Der Sensor-Fingerabdruck sei zudem zeitlich stabil. Um ihn zu erkennen, brauche es IT-forensische Methoden der Signalverarbeitung. Für das menschliche



Auge seien die Empfindlichkeitsunterschiede nicht sichtbar. Durch einen Abgleich des im Foto festgestellten Rauschverhaltens mit dem Rauschverhalten einzelner Sensoren kann so der Sensor samt Kamera ausgemacht werden, auf dem das Foto beruht. Oder aber man nutzt aus, dass der Sensor-Fingerabdruck bei Bildänderungen beschädigt oder sogar entfernt wird, und schließt daraus auf eine Manipulation (vgl. Kirchner 2022, 65ff.; Cozzolino et al. 2022, 283f.; Riess 2021, 347).

Die Technologie gilt als Eckpfeiler der Multimedia-Forensik, wird als besonders robust gegenüber gängigen Formen der Nachbearbeitung, einschließlich verlustbehafteter Komprimierung und Filterung, beschrieben und habe es bisher geschafft, sich allen technischen Herausforderungen anzupassen. Auch wenn eine starke Komprimierung, geringe Auflösung sowie digitale Verfahren wie die Korrektur von Objektivverzerrungen, Bildstabilisierung oder eine HDR-Bildgebung durchaus die Kameraidentifizierung behindern würden. Voraussetzung für die Methode ist, dass sich die verdächtige Kamera im Besitz einer Forensiker:in befindet oder eine Bildserie an Referenzmaterial existiert (vgl. Kirchner 2022, 65ff., 81f.; Taspinar u. Memon 2022, 117).

Eine ML-basierte Variante dieses Ansatzes wird als *NoisePrint* beschrieben, zielt aber hauptsächlich auf Unstimmigkeiten innerhalb eines Bildes ab, als auf die Zuordnung zu einer speziellen Kamera: „Hierbei wird aus einer Menge von Beispielsbildern ein Filter gelernt, der den Bildinhalt unterdrückt, und gleichzeitig [...] *Bildrauschen* [...] stark hervorhebt“ (Riess 2021, 348).

Hat man hingegen keine Informationen, von welcher Kamera ein Foto stammen könnte oder möchte man herausfinden, welche anderen Fotos von derselben Kamera stammen, ist man auf Datenbanken angewiesen, die einen Abgleich fotografischer Erzeugnisse auf deren Sensoreigenschaften erlauben<sup>15</sup> (vgl. Taspinar u. Memon 2022).

Die Ausweitung der Methodik auf den Videobereich wird hingegen als offenes Problem beschrieben. Dies liege vor allem daran, dass zusätzliche Schritte bei der Videoerstellung das sensorspezifische Muster viel stärker beeinträchtigten, als das bei Fotos der

---

<sup>15</sup> Dies ist zum Beispiel für Ermittlungen zu Kinderpornografie relevant, jedoch weniger für den journalistischen Verifizierungsprozess. Daher wird es an dieser Stelle nicht weiter beachtet, der aktuelle Forschungsstand kann jedoch ebenfalls im Sammelwerk *Multimedia Forensics* nachgelesen werden.

Fall sei. Insbesondere würden zusätzliche Informationsverluste (Komprimierung, Codierung, Bildstabilisierung) die Kameraidentifizierung im Videobereich erschweren. Schlussendlich lässt sich als aktueller Forschungsstand festhalten, dass für eine erfolgreiche Kameraidentifizierung auf Basis von Videomaterial die in Frage kommende Kamera samt PRNU-Referenzmuster zur Verfügung stehen sollte. Gleichzeitig zeigt ein Studienergebnis des Informatikers Husrev Taha Sencar (vgl. 2022, 89f., 112f.), dass der Identifizierungsprozess bei Videos mit einer Bitrate unter 900 kBit/s beziehungsweise einer Auflösung unter 500 x 500 Pixeln nicht zuverlässig möglich ist. 900 kBit/s sind weniger als ein Achtel der empfohlenen Bitrate für YouTube-Uploads in FullHD-Qualität.

Wie bereits angesprochen, können pixelbasierte Verfahren nicht nur durch die Ermittlung des Bildursprungs bei der Verifizierung von Inhalten helfen, sondern es können auch **Inkonsistenzen innerhalb des Bildmaterials** ausgemacht werden, die auf eine nachträgliche Bearbeitung hindeuten. Dabei macht man sich zunutze, dass selbst gut durchgeführte Bearbeitungen immer Veränderungen auf Pixelebene, im Rauschverhalten oder in den Kompressionsmustern, hinterlassen würden (vgl. Cozzolino et al. 2022, 281f.; vgl. Mercier et al. 2019, 163). Grégoire Mercier et al. u. a. vom IT-Unternehmen eXo maKina unterscheiden die pixelbasierten Methoden nach der Manipulationsform, die sie erkennen können. So findet eine Dreiteilung statt:

"Three main groups of modifications are copy–moving, splicing or in-painting, and whole-image operations. In the first case, a part of the image is replicated and placed elsewhere in it — for example, the background is copied to remove an object or person, or a crowd is duplicated to appear larger. Copy–move detection algorithms attempt to capture the forgery by looking for self-similarities within the image [...]. In the case of splicing, a part of one image is placed within another. Splicing detection and localization algorithms are based on the premise that, on some possibly invisible level, the spliced area will differ from the rest of the image due to their different capturing and compression histories [...] Finally, whole-image operations such as rescaling, recompression, or filtering cannot be localized and thus are generally tackled with tampering detection algorithms"  
(Mercier et al. 2019, 166)

Bei der **Analyse von Kompressionsspuren (Error Level Analysis)** geht man davon aus, dass die Informationsreduktion im Rahmen des Kompressionsprozesses bspw. in Kameras anders erfolge als in Bildbearbeitungssoftware und sich zudem von Gerät zu

Gerät und von Software zu Software unterscheide. Dadurch könne erkannt werden, „ob ein Bild aus einer Kamera kommt, oder in einem Bearbeitungsprogramm verändert wurde“ (Gerling 2022, 309), aber auch ob einzelne Teile eines Bilder häufiger komprimiert worden sind als andere, was auf lokale Eingriffe hinweist (vgl. Riess 2021, 347f.). Verhindern könne man die Kompressionsspuren nur durch pixelweises Kopieren, was aber sehr aufwendig sei und deshalb von den meisten manipulierenden Akteur:innen gescheut werde (vgl. Welchering 2020, 24f.).

Obwohl die Manipulationsformen bei Fotos und **Videos** größtenteils gleich sind, weisen Mercier et al. mit Verweis auf eigene Experimente darauf hin, dass die für Fotos entwickelten Analysealgorithmen bei Videos nicht funktionieren, in erster Linie aufgrund der andersartigen Codierung des Videomaterials. Um die Datenmenge möglichst gering zu halten, werden bei der Codierung von Videomaterial die zeitlichen Wechselbeziehungen zwischen Einzelbildern ausgenutzt und damit die Spuren zerstört, die bei Fotos der Manipulationsdetektion dienen – erst recht bei zusätzlichen Codierungsschritten durch Social-Media-Plattformen. Mercier et al. sprechen sich daher für die Entwicklung zielgerichteter, videobasierter Algorithmen aus, die auf das jeweils übliche Codierungsformat spezialisiert sind. Die technischen Grundlagen, durch die entsprechende Manipulationsformen zu finden sind, bleiben jedoch sehr ähnlich. Genaue technische Erläuterungen der Filter, die es möglich machen die Art von Auffälligkeiten in Videos zu entdecken, welche auf den letzten Seiten vorgestellt worden sind, finden sich beispielsweise bei Mercier et al. (vgl. 2019, 171ff.). Sie stellen dort die Filter vor, welche sie im Rahmen des InVID-Projektes entwickelt haben. Für diese Arbeit ist jedoch in erster Linie die Möglichkeit relevant, mit solchen Filtern lokal begrenzte Veränderungen in Videos zu erkennen. Bisher muss die Auswertung der Filterdarstellung noch manuell erfolgen. Ein im Rahmen des InVID-Projekts erfolgter Versuch die Auswertung zu automatisieren führte nicht zu zufriedenstellenden Ergebnissen: „we cannot claim that we are ready for real-world application“ (Mercier et al. 2019, 185).

Logischerweise geht es zusätzlich zu der Suche nach Auffälligkeiten innerhalb eines (Stand-)Bildes hier auch um Auffälligkeiten innerhalb des zeitlichen Verlaufs (vgl. Mercier et al. 2019, 167f.). So können beispielsweise „copy-moving“-Manipulationen auch

zeitlich verschoben eingesetzt werden. Auch können Videos dadurch manipuliert werden, dass einzelne Frames entfernt oder hinzugefügt werden. Diese Manipulationsform ist durch forensische Analysen eines einzelnen Standbildes logischerweise nicht erkennbar, daher existieren hierfür andere Ansätze (vgl. Long et al. 2022, 333, 359). Einer davon ist die Betrachtung mehrerer aufeinander folgender Frames, eine sogenannte Einzelbildanalyse: „Eine Kopie des zu untersuchenden Videos wird in eines der marktüblichen Videoschnittprogramme geladen. An der kritischen Stelle wird die Datei dann eben nicht normal abgespielt, sondern mit dem Knopf ‚nächster Frame‘. Bei dieser Sichtungsmethode fallen dann ‚gestreckte Videos‘, also einkopierte Frames oder Bilder sofort auf“ (Welchering 2020, 26f.). Die Manipulationsform kann jedoch sowieso nur zu einem unauffälligen Ergebnis führen, wenn das Ausgangsmaterial bereits Schnitte enthält und/oder es sich um eine statische Aufnahmesituation mit gleichbleibendem Hintergrund handelt und ist damit für die meisten nutzergenerierten Inhalte kaum von Relevanz (vgl. Mercier et al. 2019, 167, 169).

Während die bisher beschriebenen Methoden auf der Analyse des audiovisuellen Dateiinhalts beruhen, hätten jüngste Forschungen gezeigt, dass auch in der **Dateistruktur** Hinweise auf den Hintergrund einer Datei zu finden sein können. Gerade wenn die Inhaltsanalyse aufgrund von stark komprimiertem oder niedrig aufgelöstem Material wirkungslos sei, ermögliche die Analyse des Dateiformats und der Metadaten, deren Kompatibilität, Vollständigkeit und Konsistenz basierend auf der erwarteten Medienhistorie zu bestimmen. Teilweise sei es möglich, die Marke oder das Modell des Aufnahmegeräts zu bestimmen und Rückschlüsse auf Soziale Netzwerke zu ziehen, über welche die Datei verbreitet wurde. Dabei sei das Verfahren unabhängig von der Dateigröße sehr schnell und von geringem Aufwand und eine Fälschung der hierbei zu gewinnenden Daten komplex: „We are currently not aware of any publicly available software that would allow users to consistently forge such information without advanced programming skills“ (Piva u. Iuliani 2022, 364). Allerdings gebe es aktuell keine Bearbeitungswerkzeuge oder Metadaten-Editoren, die in der Lage sind, diese Informationen zu extrahieren. Hinzu kommt, dass es an Trainings-Datensätzen mangle, spätere Verarbeitungsschritte dazu neigen würden die Spuren früherer Schritte zu entfernen und besonders Referenzdaten, die eine Zuordnung zu einem Sozialen Netzwerk erlau-

ben, aufgrund ihrer ständigen Veränderungen schnell veralten würden. Außerdem ermögliche die Methode nur eine Fälschungserkennung, aber nicht die Lokalisierung der Fälschung (vgl. Piva u. Iuliani 2022, 363f., 384f.).

Ein weiterer Ansatz, der vorrangig für Audiosignale entwickelt wurde, besteht in der **Analyse** sogenannter **ENF-Signale** (*Electrical Network Frequency*). Dabei nutzt man aus, dass sich die Frequenz des Stromnetzes zwischen dem größten Teil Amerikas (60Hz) und dem überwiegenden Rest der Welt (50Hz) unterscheidet, aber auch, dass die tatsächliche Netzfrequenz vor Ort leichten Schwankungen um ihren Nennwert herum unterworfen ist. Die Kette an tatsächlichen Momentanwerten ist als ENF-Signal definiert. Da die Schwankungen zufällig, zeitlich einzigartig und typischerweise an allen Standorten desselben Stromnetzes sehr ähnlich seien, stelle das ENF-Signal – sofern es in einer Aufnahme enthalten ist – eine Umgebungssignatur zum Zeitpunkt der Aufnahme dar. Die häufigste Ursache für ENF-Signale bei Audioaufnahmen sei ein akustisches Netzbrummen, das von netzbetriebenen Geräten am Ort der Aufnahme erzeugt wird. Wobei der Aufnahmeort und die Brummquelle sich Studien zufolge durchaus in unterschiedlichen Räumen und einer Distanz von zehn Metern befinden dürfen (vgl. Hajj-Ahmad et al. 2022, 235f., 245).

Ist sowohl das ENF-Signal einer Aufnahme vorhanden, als auch ein Referenzsignal vom Ort der Aufnahme zum vermeintlichen Aufnahmezeitpunkt, können beide Signale miteinander verglichen werden. Es gibt zusätzlich Ansätze, den Aufnahmeort ohne Referenzsignal auf ein Stromnetz einzugrenzen, indem beispielsweise von der Größe der Schwankungen auf die Größe des Stromnetzes geschlossen wird: Je größer ein Stromnetz, desto geringer seien die Schwankungen. Lokale Lasteigenschaften einer Stadt oder Netzstörungen würden sogar die Lokalisierung des Aufnahmeortes innerhalb eines Stromnetzes ermöglichen. Ein Abgleich zweier ENF-Signale könne auch Rückschlüsse auf den Abstand zwischen beiden Aufzeichnungsorten erlauben. Ebenso sollen Manipulationen in Aufnahmen durch Auffälligkeiten auf der Ebene ihres ENF-Signals erkannt werden und sogar Aufschluss darüber geben, ob an der auffälligen Stelle etwas entfernt oder hinzugefügt worden ist. Weitere Forschungsansätze beschäftigen sich

damit, wie ENF-Signale aus digital aufgenommenem Bildmaterial extrahiert werden können (vgl. Hajj-Ahmad et al. 2022, 248, 258ff.).

#### 2.4.5 Identifizierung Synthetischer Medien / Deepfakes

Was die Schwierigkeit angeht, synthetische Medien als solche zu erkennen, und auch hinsichtlich der dazu geeigneten Methoden, unterscheiden sich die Einschätzungen in der Literatur deutlich. Klar ist: Auf synthetische Medien sind die zuletzt beschriebenen technischen Ansätze nur bedingt anwendbar, weil synthetische Inhalte häufig in einem Guss entstehen und damit herkömmliche Inkonsistenzen innerhalb des Materials nicht existieren. Stattdessen gehe es meistens darum, ein synthetisches Produkt in seiner Gänze als solches zu erkennen (vgl. Mercier et al. 2019, 170):

"Despite the continuous research efforts and the numerous forensic tools developed in the past, the advent of deep learning is changing the rules of the game and asking multimedia forensics for new and timely solutions."

(Verdoliva 2020, 911)

Eine Ausnahme stellen synthetische Manipulationen dar, die sich nur auf Teile eines Bildes oder Videos, zum Beispiel das Gesicht, beziehen (vgl. Mercier et al. 2019, 170). Auch physikbasierte Methoden werden zur Detektion synthetischer Medien vorgeschlagen. Gerade bei gelungenen Fälschungen helfe es derzeit noch, die Ausrichtung von Schatten zu überprüfen oder auch die Stimmigkeit der Audiospur gegenüber der dargestellten Szenerie, meint der OSINT-Journalist Forrest Rogers (vgl. Bischoff 2022a, 73). Genauso lohne ein Blick, „ob das Gesicht zum Körper passt und die Körperhaltung zur geäußerten Mimik“ (Welchering 2020, 30). Unstimmigkeiten durch abrupt endende Tattoos, ein Wechsel in der Hautfarbe oder bei der Muskelkontur, ein unscharfer Mundinnenraum und fehlendes oder künstlich wirkendes Blinzeln seien weitere Anzeichen (vgl. ebd.). All diese Merkmale können allerdings mit genügend zielgerichtetem Training der neuronalen Netzwerke perfektioniert werden, womit sie als Unterscheidungsmerkmale an Wert verlieren.

Ein weiterer Ansatz zur Detektion synthetischer Medien beschäftigt sich mit der Suche nach charakteristischen Merkmalen authentischer Medieninhalte, wie zum Beispiel dem spezifischen Rauschabdruck eines jeden Bildsensors. Fehlen solche Merkmale und/oder finden sich stattdessen Spuren eines synthetischen Herstellungsprozesses,

kann auf einen Deepfake geschlossen werden. Anders als für Menschen sei das Erkennen von Spuren sogenannter *Generative Adversarial Networks (GAN)*<sup>16</sup>, die vielen synthetischen Gesichtern zugrunde liegen, für moderne Erkennungssysteme sehr gut möglich. Daher scheine gerade die automatisierte Erkennung synthetischer Medien vorerst relativ einfach zu sein, gerade im Vergleich zur Verifizierung von Shallowfakes, konstatieren mehrere Informatiker: „Yet, detecting computer-generated material seems to be relatively simple, for the time being“ (Cozzolino et al. 2022, 305; vgl. Neves et al. 2022, 177). Zudem führe der automatisierte Generierungsprozess bei synthetischen Medien dazu, dass KI-basierte Detektionssoftware eine nahezu unbegrenzte Menge an Trainingsdaten zur Verfügung habe (vgl. ebd.).

Die Suche nach Merkmalen, die für oder gegen die Authentizität des Materials sprechen, kann entweder unter Vorgabe definierter Anomalien auf Signalebene (*Signal Feature-Based Methods*) durchgeführt werden oder auch ohne die Vorgabe definierter Unterscheidungsmerkmale, allein auf der Basis maschineller Lernprozesse, denen Datenbanken mit Material zugrunde liegen, das als *echt* oder *synthetisch* klassifiziert wurde. Die zweite Variante führt aktuell zu den besseren Ergebnissen (vgl. Lyu 2022, 318ff.). Trotzdem beträgt die durchschnittliche Genauigkeit dieser Systeme bei unbekanntem Videomaterial bisher nur 65 Prozent (vgl. ebd., 315f.). Damit stehen die bisherigen Testergebnisse automatisierter Erkennungssysteme im Widerspruch zu der vermeintlich einfachen Identifizierung synthetischer Inhalte. Bei einer erheblich reduzierten Bildauflösung sinkt die Erkennungsrate sogar um mehr als das Siebenfache. Ebenfalls sinkt die Genauigkeit erheblich, wenn ein Autoencoder zum Entfernen der GAN-Spuren (GANprintR) bei der Erstellung der Gesichtsmanipulationen eingesetzt wird (vgl. Neves et al. 2022, 198ff.). Hinzu kommt, dass die meisten bisherigen Lösungen als „black box“ zu sehen sind, denen es an der Erklärbarkeit der Ergebnisse mangelt (vgl. Lyu 2022, 324f.).

Nach dieser Darlegung ist es nicht verwunderlich, dass die Informatikerin Siwei Lyu, die bereits an den ersten Algorithmen zur Deepfake-Erkennung beteiligt war, eine Reihe

---

<sup>16</sup> GAN beruhen darauf, dass ein zweites künstliches neuronales Netz zeitgleich versucht und trainiert, die Ergebnisse des ersten Netzes von den Ausgangsbildern zu unterscheiden, woraufhin das erste Netz seine Technik anpasst (vgl. Neves et al. 2022, 178; Stamm u. Zhao 2022, 468).

kritischer Probleme sieht, die für eine erfolgreiche Detektion noch gelöst werden müssen. Für eine breite praktische Anwendung müssen nach Lyu (vgl. 2022, 313, 328) Gesamteffizienz, Erkennungsgenauigkeit und vor allem die Falsch-Positiv-Rate der Erkennungssysteme verbessert werden. Außerdem müssen die Systeme robuster gegenüber Veränderungen des zu analysierenden Materials werden. Dies mögen Gründe dafür sein, dass der Journalist und Medientrainer Peter Welchering (vgl. 2020, 27ff.) automatisierte Detektionssysteme bisher nicht alltagstauglich genug für den Einsatz in Redaktionen hält. Deshalb sei die Einzelbildanalyse bisher noch die Methode der Wahl, um synthetische Medien als solche zu erkennen.

Es bleibt festzuhalten, dass derzeit viele Forschungsansätze zur Detektion synthetischer Medien verfolgt werden, eine langfristige Lösung aufgrund der anpassungsfähigen und sich stetig verbessernden Software zur Erstellung von Deepfakes jedoch schwierig ist. Die Folge ist ein „potentiell unendlich andauerndes Wettrüsten von Fälschungs- und Erkennungssoftware“ (Schade 2019, 101; vgl. Lyu 2022, 328), auch von einem Katz-und-Maus-Spiel ist die Rede (vgl. Kietzmann et al. 2020, 144):

"In synthesis, deep learning is part of the problem but also part of the solution"  
(Sencar et al. 2022, vi)

#### *2.4.6 Automatisierte Manipulationserkennung*

Funktionierende automatisierte Detektionssysteme könnten nicht nur die Erkennung synthetischer Medien stark vereinfachen, sondern allgemein die Verifizierung audiovisueller Inhalte. Sie seien sogar notwendig, sagen manche Akteure:

"We will require automated tools to augment our human abilities to make sense of what we read, hear, and see." (Hendrix u. Morozoff 2022, 8)

Die Vision eines vollautomatisierten Verifikationsprozesses für Videos, der die Glaubwürdigkeit des Materials mit einem oder mehreren numerischen Werten angibt, wird auch im Rahmen des InVID-Projektes angesprochen. Bisher sei eine technische Lösung allerdings nicht in Reichweite und zudem unklar, ob eine solche überhaupt erreicht werden kann (vgl. Fricke u. Thomsen 2019, 307, 310). Im Vergleich zur Erkennung synthetischer Medien sei die Identifizierung von Shallowfakes herausfordernder. KI-basierte Ansätze zur Erkennung herkömmlicher Manipulationen ständen unter ande-



rem vor dem Problem, erst einmal an einen repräsentativen Datensatz dieser Manipulationen zu kommen, der nicht einfach generiert werden kann (vgl. Cozzolino et al. 2022, 305). Außerdem liegt eine Herausforderung in der Unterscheidung zwischen bösartigen Bearbeitungsprozessen und solchen, die das Material bloß aufwerten oder als Teil eines Sendeformats (Branding) markieren sollen (vgl. Mercier et al. 2019, 163):

"A good forensic algorithm should be able to single out only malicious attacks, based on the type of manipulation, often concerning only a small part of the image, and especially based on the context, including all other related media and textual information." (Cozzolino et al. 2022, 304).

Eine weitere Herausforderung für den Einsatz DL-basierter Methoden ist derzeit die Bildgröße: Um Komplexität und Speicher zu begrenzen, muss das zu verifizierende Material entweder verkleinert oder abschnittsweise analysiert werden. Während ersteres mögliche Beweisspuren zerstört, verhindert zweiteres eine Analyse auf Bildebene (vgl. Cozzolino et al. 2022, 303). Es muss also zwischen Softwarelösungen, die auf Pixelebene arbeiten, und solchen, die auf Bildebene arbeiten, unterschieden werden. Während Softwarelösungen auf Bildebene ihre Analyse in der Regel zu einem Wahrscheinlichkeitswert für die Authentizität des Materials zusammenfassen, zeigen Softwarelösungen auf Pixelebene auch die Stellen vermuteter Manipulationen (vgl. Cozzolino et al. 2022, 283, 290; Mercier et al. 2019, 165f.).

Trotz der Herausforderungen sind IT-Expert:innen auf dem Gebiet durchaus optimistisch, dass automatisierte Detektionsmethoden einen wichtigen Beitrag für die Verifizierung audiovisuellen Materials leisten können (vgl. Cozzolino et al. 2022, 306; Papadopoulou et al. 2019, 195). Im Kontext der Analyse von JPEG-Kompressionsmustern weist Cozzolino (vgl. 2022, 288f.) darauf hin, dass DL-basierte Verfahren auch oder gerade in Fällen, wo herkömmliche Methoden weitgehend versagen, gute Ergebnisse liefern würden. Für die Detektion identischer Stellen innerhalb eines Bildes (copy-moves) mit geringer Auflösung überträfen DL-basierte Methoden bereits die herkömmlichen.

Gerade die hierbei zugrunde liegenden *Convolutional Neural Networks* (CNN) werden jedoch gleichzeitig für besonders vulnerabel gegenüber bewussten Anpassungen der Manipulationsmechanismen gehalten. So sei es einfach, mit partiellem Wissen über

ein Analyseprogramm Material zu erstellen, das manipuliert ist, aber vom Analyseprogramm nicht mehr als solches klassifiziert wird: „Such a problem is currently the subject of an intense research activity, yet no satisfactory solution has been found yet“ (Barni et al. 2022, 436). Eine stärkere Robustheit DL-basierter Verfahren gegenüber bewussten Täuschungsversuchen wird daher als eine der dringendsten Anforderungen auf dem Weg zu funktionalen automatisierten Detektionssystemen beschrieben (vgl. ebd., 461f.; Cozzolino et al. 2022, 306).

Zudem kommt der Informatiker Oren Halvani (vgl. 2020, 142f.) zu dem Schluss, dass Maschinelles Lernen im Kontext von Desinformation erst dann wirklich erfolgreich sein könne, wenn große Mengen an Medien, die bereits als Desinformation erkannt wurden, in Datenbanken entsprechend abgelegt sind und als Trainingsmaterial für zukünftige Detektionsverfahren zur Verfügung stehen. Die Nutzung erster Ergebnisse durch Journalist:innen könne aber bereits in naher Zukunft erfolgen. Ein Automatisierungsgrad, der auch vollkommenen Laien die Nutzung ermögliche, sei später ebenfalls denkbar. Auch funktionale automatisierte Detektionssysteme können jedoch kein „universeller ‚Wahrheitsfinder‘“ sein, gibt Christian Riess (2021, 343) zu bedenken. Zwar könnten maschinelle Lernverfahren in einem vorgegebenen Rahmen „bemerkenswertes leisten und sowohl in Geschwindigkeit als auch Genauigkeit Menschen in vielen Fällen übertreffen“ (ebd., 355), aber weder seien sie im Stande „abstrakte Konzepte wie ‚Wahrheit‘ zuverlässig zu erfassen, noch die Grenzen der eigenen Anwendbarkeit zu erkennen“ (ebd., 343). So sind automatisierte Detektionssysteme ein Werkzeug, das Journalist:innen in Verifizierungsprozessen unterstützen kann. Dafür müssen die Ergebnisse des Systems allerdings für diese nachvollziehbar sein (vgl. ebd.): „Improving explainability, though largely unexplored, is therefore a major topic for current and future work in this field“ (Cozzolino et al. 2022, 306).

Als Indikator für die momentane Überlegenheit von Menschen über Maschinen bei der Einordnung sensibler Inhalte führen die Journalismusforscher:innen Wiebke Loosen und Paul Loosbach auf, dass sogar ein digitaler Großkonzern wie Facebook die Verifikation seiner Inhalte bisher an Dienstleister wie Arvato oder die Deutsche Presse-Agentur auslagere (vgl. Loosen u. Solbach 2020, 186).

#### 2.4.7 Aktive Forensik

Unter dem Begriff der aktiven Multimedia-Forensik (engl. *active forensics*) werden hingegen Methoden zusammengefasst, welche die Integrität audiovisuellen Materials präventiv überwachen. Dafür werden bestenfalls bereits zum Zeitpunkt der Aufnahme Informationen in die Datei integriert, anhand derer später sichergestellt werden kann, dass das Material nicht verändert worden ist und eventuell sogar, von wem es stammt. Im Gegensatz dazu werden alle bisher genannten Bemühungen, Inhalte ohne solche vorbeugenden Maßnahmen allein aus dem Material selbst heraus zu verifizieren, als passive Forensik zusammengefasst. Da aktive Ansätze bisher kaum verbreitet sind, kommen für die Verifizierung audiovisueller Inhalte bisher meist ausschließlich passive Methoden in Frage (vgl. Mercier et al. 2019, 165; Papadopoulou et al. 2019, 194). Jedoch bergen zuverlässige Verfahren aktiver Forensik das Potenzial, audiovisuelles Material tatsächlich zu verifizieren, statt bloß Indizien für dessen Authentizität zu sammeln. Eine Umsetzung kann beispielsweise auf der Basis von Wasserzeichen oder auch von Signaturen erfolgen. Gerade für die Verifizierung synthetischer Inhalte werden aktive Ansätze teilweise für vielversprechender gehalten, als die bereits vorgestellten passiven Ansätze (vgl. Schade 2019, 101f.). Auch vom BSI (vgl. ) werden kryptografische Verfahren, zum Beispiel in Form einer digitalen Signatur beim Aufnahmeprozess, als mögliche Gegenmaßnahme zu audiovisuellen Manipulationen vorgestellt. Ebenso sieht Elliot Higgins (vgl. 2021, 271f.) die Integration von Verifikationsdaten in audiovisuelle Aufnahmen als Möglichkeit, auf synthetische Medien zu reagieren.

Ein Ansatz, der auch im Journalismus nicht unbeachtet bleibt, wie bspw. das *News Provenance Project* aus einer Kooperation der The New York Times mit IBM zeigt. Das Projekt, welches aber anscheinend seit 2020 nicht weiterentwickelt wurde, setzt unter anderem auf einen Echtheitsnachweis bei Fotos mithilfe der Blockchain-Technologie:

"We chose blockchain because its data structure can help maintain a transparent and immutable record of a photo's origins: when, where and by whom it was taken, who published it and how it has been used across a network of news organizations." (NPP 2020)

Das Projekt stellt weder die einzigen noch die umfangreichsten Bemühungen in diesem Bereich dar. Eine inzwischen sehr umfangreiche Initiative ist die von Adobe 2019 ins

Leben gerufene *Content Authenticity Initiative* (CAI). Sie hat gemeinsam mit den Beteiligten von *Project Origin 2021* die *Coalition for Content Provenance and Authenticity* (C2PA) gegründet, welche im selben Jahr die erste Version eines technischen Standards für den Nachweis der Herkunft und Authentizität digitaler Inhalte vorgestellt hat. Die technologische Grundlage für den C2PA-Standard bilden laut den beteiligten Akteur:innen kryptographisch sichere Signaturen, aufgrund derer die Quell- und Bearbeitungsdaten des erstellten Materials manipulationssicher in die jeweilige Datei eingebettet und später eingesehen werden können. Der Standard setzt keine Verwendung der Blockchain-Technologie voraus, ist aber mit ihr kompatibel. Teil der beiden Zusammenschlüsse zur Gestaltung, Förderung und Verbreitung des C2PA-Standards sind außer *Adobe* sowohl IT-Firmen wie *arm*, *Intel*, *Microsoft* und *nvidia* als auch Kamerahersteller wie *Canon*, *Leica*, *Nikon*, *Panasonic* und *Sony*, Anbieter von Bilddatenbanken wie *Getty Images*, aber auch journalistische Institutionen wie *AFP*, *AP*, *axel springer*, *BBC*, *dpa*, *REUTERS*, *stern*, *The Wall Street Journal*, *Washington Post* und *The New York Times* (vgl. Hendrix u. Morozoff 2022, 28f.; CAI; C2PA; C2PA). Dass es sich nicht allein um theoretische Überlegungen handelt, sieht man daran, dass das Unternehmen *Truepic*, das auch Gründungsmitglied von C2PA ist, bereits eine Sammlung von Programmierhilfen anbietet, die es erlaubt den C2PA-Standard in bestehende iOS-, Android- oder Webanwendungen zu integrieren (vgl. Truepic 2023). Bisher scheint sich eine solche Lösung jedoch nicht durchzusetzen.

## 2.5 Die Verifizierungspraxis in Redaktionen

Mit dem veränderten Einsatz und den veränderten Manipulationsmöglichkeiten audiovisuellen Materials sind Redaktionen weltweit konfrontiert – und sie reagieren. Das folgende Kapitel soll einen Überblick über den derzeitigen Umgang deutscher Redaktionen mit den beschriebenen Herausforderungen geben. Dass die Verifizierung von Inhalten eine zentrale Aufgabe im Journalismus ist und auch früher schon war, darüber herrscht Einigkeit. Und trotzdem sind dezidierte Einheiten, kleinere Teams oder ganze Redaktionen, die sich 'nur' mit der Prüfung von Inhalten beschäftigen, vor allem in Deutschland eine recht neue Erscheinung.

Das *Reporter's Lab* am *DeWitt Wallace Center for Media & Democracy* innerhalb der *Duke University* versucht seit 2014 über eine Datenbank und damit verbundene Karte Überblick über diese Entwicklung zu geben. Um in die Datenbank des *Reporter's Lab* aufgenommen zu werden, sind die Kriterien denen des *International Fact-Checking Networks* (IFCN) sehr ähnlich: Es handelt sich ausschließlich um Organisationen, deren primäre Aufgabe darin besteht, Informationen bereitzustellen und zudem regelmäßig Beiträge publizieren, in denen verbreitete Inhalte überprüft werden. Kontrolliert wird auch, ob die Organisationen ausgeglichen und transparent arbeiten. Anders als beim IFCN, wo nur die Dachorganisationen gezählt werden, gilt für das *Reporter's Lab* jedoch jede regionale Einheit als eigene Organisation (vgl. Adair u. Stencel 2020). So kommt es, dass sich die Anzahl eingetragener Organisationen zwischen beiden Institutionen unterscheidet, der starke Zuwachs in den letzten Jahren ist jedoch bei beiden zu beobachten.

Orientiert man sich an der jährlichen Erhebung des *Duke Reporter's Lab*, gibt es 2023 mittlerweile 417 aktive Faktencheck-Einheiten in insgesamt 108 Ländern. Das entspricht mehr als einer Verdopplung seit 2016 (188). Inzwischen nimmt die Anzahl an Einheiten nicht mehr bedeutend zu. Das ist nicht zwangsläufig ein schlechtes Zeichen, sondern deutet vielmehr darauf hin, dass sich inzwischen eine feste Struktur an Faktencheck-Einheiten zu etablieren beginnt. Fast die Hälfte der erfassten Einheiten ist inzwischen seit min. fünf Jahren aktiv (vgl. Stencel et al. 21.06.23). Bekannte internationale Medienhäuser wie *The New York Times*, *BBC*, *The Wall Street Journal* oder auch

die Nachrichtenagenturen *Reuters* und *The Associated Press* haben inzwischen eigene Abteilungen zur Auswertung und Verifizierung audiovisueller Inhalte (vgl. Khan et al. 2023, 6749; Higgins 2021, 279f.). In Deutschland sind es vor allem *ARD* und *ZDF*, die Faktencheck-Einheiten betreiben, aber auch Agenturen wie *dpa* und *AFP*. Auch die *Deutsche Welle*, das Magazin *Stern* und das Recherchenetzwerk *Correctiv* werden vom *Reporter's Lab* als aktive Faktencheck-Einheiten aufgeführt.

Zu beachten ist, wie bereits beschrieben, dass in der Statistik nur Strukturen erfasst werden, die ihre Rechercheergebnisse auch als eigenständigen Beitrag publizieren. Einheiten, die ausschließlich für eine interne Überprüfung von Inhalten zuständig sind, müssen bei einer Gesamtbetrachtung der journalistischen Verifikationslandschaft zusätzlich bedacht werden. Nach einer Befragung von etwas mehr als 4100 Journalist:innen und Redaktionsleiter:innen in 149 Ländern durch das *International Center for Journalists (ICFJ)* aus dem Jahr 2019 gibt es in einem Drittel aller befragten Nachrichtenorganisationen dezidierte Faktenchecker:innen und mehr als die Hälfte der befragten Journalist:innen gibt an, regelmäßig auf digitale Hilfsmittel zur Überprüfung von Fakten zurückzugreifen. Nur vier Prozent aller Redaktionsmitarbeiter:innen seien jedoch tatsächlich technische Fachleute wie z. B. Produktentwickler, Tendenz leicht steigend (vgl. ICFJ 2019, IVff.).

Wie bereits durch die Begriffe 'Faktencheck-Einheiten' und 'Faktenchecker:innen' deutlich wird, geht es in den Redaktionen überwiegend gar nicht darum, das Bild- oder Tonmaterial zu verifizieren, sondern vielmehr darum, die darin eingebettete Aussage zu belegen oder zu entkräften. Es handelt sich also vielmehr um Fact-Checking als um Verifikation.<sup>17</sup> Entsprechend wird nie das entsprechende Material allein für eine Beurteilung herangezogen (vgl. Stern 2020, 130ff.). Es wird deutlich, dass traditionelle Recherchemethoden wie die direkte Kontaktaufnahme mit den Quellen trotz der technologischen Entwicklungen weiterhin einen sehr großen Teil in redaktionellen Verifikati-

---

<sup>17</sup> Das wird sowohl durch die Beschreibungen der Verifikationsprozesse deutlich, als auch dezidiert in Podiumsdiskussionen und Vorträgen betont. Siehe bspw. *PresseClub München (2022): BJV-Podiumsdiskussion: Fotos und Videos: wahr oder falsch? – Bilder im Röntgenblick. Podiumsdiskussion mit Dennis Amour, Janina Lückoff, Dr.-Ing. Christian Riess, Lea Weinmann. Online unter [youtu.be/XfeoOCcObq](https://youtu.be/XfeoOCcObq), zuletzt geprüft am 07.06.23*

onsprozessen ausmachen (vgl. Riess 2022, 208; Sell u. Oswald 2021, 249). Die bereits auf Seite 25 vorgeschlagene Bearbeitung der fünf Kernfragen zu Originalität, Quelle, Ort, Datum und Motivation hinter dem Erzeugnis und seiner Verbreitung scheint auch in der Praxis üblich zu sein (vgl. Khan et al. 2023, 6751; Sell u. Oswald 2021, 253). Dafür wird laut den Berichten der Faktenchecker:innen<sup>18</sup> außer klassischer Recherche (Lesen zugehöriger Kommentare, Übersetzen fremdsprachiger Inhalte, Kontaktaufnahme zu potenziellen Quellen etc.) vor allem mit frei verfügbaren Werkzeugen gearbeitet, mithilfe derer man sich die Metadaten anschauen und den Materialinhalt auf Plausibilität prüfen kann.

Alle Informationen und Daten, die aus öffentlichen bzw. frei verfügbaren Quellen stammen, werden auch als Open Source Intelligence (OSINT) bezeichnet. Es handelt sich dabei also um digitale Werkzeuge (Software, Plattformen), die Recherchen und Ermittlungen erleichtern (vgl. Spangenberg u. Teyssou 2019, 6f.). Sie begründen eine Recherchebewegung, die inzwischen auch von Redaktionen in Anspruch genommen wird. OSINT-Recherchen zeichnen sich in der Regel nicht nur durch die Recherche mithilfe frei verfügbarer Quellen aus, sondern auch durch die transparente Darstellung des Rechercheprozesses und die Einbeziehung Freiwilliger: „Ein entscheidender Faktor von OSINT ist die Schwarmintelligenz“ (Bischoff 2022b, 66, vgl. 64f.). Die Zuhilfenahme des Publikums durch professionelle Journalist:innen wird auch als kollaborativer Journalismus bezeichnet (engl. collaborative journalism, participatory journalism, networked journalism) und klar von der Informationsbereitstellung und dem Informationskonsum durch Laien abgegrenzt. Letzteres wird mitunter als Graswurzel-Journalismus (engl. grassroots journalism, citizen journalism, public journalism, guerrilla journalism) benannt (vgl. Spangenberg u. Heise 2014, 765f.). Um das Vorgehen bei OSINT-Recherchen zu strukturieren und Rechtsverletzungen im Rahmen der Recherchen zu verhindern, hat die University of California in Berkeley zusammen mit den Vereinten Nationen 2020 im Rahmen des Berkeley-Protokolls einen Leitfaden für OSINT-Recherchen herausgegeben. Darin wird der Fokus auf eine lückenlose Doku-

---

<sup>18</sup> Große Teile der folgenden Informationen zu den Verifikationsprozessen in Redaktionen stammen aus journalistischen Berichten und nicht aus wissenschaftlicher Fachliteratur.

mentation, Offenheit gegenüber Wissenslücken und eine ordnungsgemäße Sicherung der Daten gelegt (vgl. Bischoff 2022b, 69f.).

"The availability of so-called 'Open Source Intelligence' (OSINT) tools have particularly changed the verification sphere and respective possibilities profoundly."  
(Spangenberg u. Teyssou 2019, 6)

Besonders häufig wird in Redaktionen auf inverse Bildersuchen und die Geo- und Chronolokalisierung insbesondere über den Abgleich mit existierendem Bild- und Kartenmaterial oder Wetterdaten verwiesen (vgl. Baig 21.03.23; Maas u. Ratering 2022, 502; Ratering 2022, 360; Heine 2022; Obermann 03.03.22; Grimm 16.03.22; Pörzgen 19.08.22; Sell u. Oswald 2021, 252ff.; Stern 2020; PresseClub München 07.04.22).<sup>19</sup> Für die Organisation der Verifikations- und Faktencheck-Prozesse wird zudem auf weitere Software zurückgegriffen (vgl. Pörzgen 19.08.22), u. a. die von der DW mitentwickelte Plattform *Truly Media* (vgl. Truly Media 09.06.2019). Von Bernd Oswald, Medientrainer und freier Journalist u. a. für BR24, heißt es, der schwierigste Teil im Verifikationsprozess sei die Überprüfung des Aufnahmezeitpunktes (vgl. Sell u. Oswald 2021, 245, 257). Zudem erreiche man bei synthetischen Medien technische Grenzen der Verifizierbarkeit (vgl. Stern 2020, 135f.). Auch von dpa-Seite heißt es, dass man nicht jetzt, aber für die Zukunft große Angst habe, auf einen Deepfake hereinzufallen (vgl. Pörzgen 19.08.22). Insgesamt sei außerdem der Konflikt zwischen Schnelligkeit und Gewissenhaftigkeit eine Herausforderung (vgl. Pörzgen 19.08.22; Weinmann 02.03.22; Fricke u. Thomsen 2019, 302).

In 80-90 % der Fälle kämen sie mit dem beschriebenen Methodenmix derzeit an die gesuchten Informationen, heißt es vom Leiter des RTL News-Verifizierungsteams, Andreas Greuel. Die Verifizierungseinheit existiert seit 2016 und besteht aus drei Mitarbeitenden, die aufgrund des Krieges in der Ukraine zeitweise durch fünf weitere Mitarbeiter:innen aus anderen Abteilungen verstärkt werden (vgl. Heine 2022). Bei der

---

<sup>19</sup> Aus den Quellen ergibt sich eine solche Verifizierungspraxis mindestens für die entsprechenden Einheiten von ARD inkl. BR24, ZDF, DW, RTL/ntv, dpa, RND und SZ. Besonders bei BR24 und SZ wird darüber hinaus sehr deutlich, dass sie zusätzlich je nach Fall weitere, teils nicht öffentliche, Hilfsmittel wie spezielle Satellitenbilder oder Karten zur Hilfe nehmen. Das ZDF betont, auch KI-basierte Tools wie Microsoft Azure, Google Lens und Mapillary zu nutzen (vgl. Ratering 2022). Ein Überblick über die öffentlich zugänglichen Hilfsmittel, die BR24 nutzt, findet sich unter <https://start.me/p/xbk65R/verifikations-toolkit-br24-faktenfuchs>, zuletzt geprüft am 19.04.23



Gemeinschaftsredaktion ARD-aktuell gebe es hingegen bereits seit 2012 eine Verifikationseinheit, die Videos auf ihre Richtigkeit prüft. Aus wie vielen Personen sich das Team zusammensetzt, macht der Sender jedoch nicht öffentlich (vgl. Pörzgen 19.08.22). Hinzugekommen ist 2017 zusätzlich eine eigene Verifikationseinheit durch den BR, die aus acht oder neun Personen besteht (vgl. BR24 2023; Pörzgen 19.08.22). Im ZDF gibt es erst seit 2020 ein Faktencheck-Team, das sich aus Journalist:innen und Dokumentar:innen zusammensetzt (vgl. Ratering 2022, 360; Stern 2020, 143). Für die ZDF-Online-Nachrichtenredaktion ist von fünf Faktenchecker:innen die Rede (vgl. Pörzgen 19.08.22; Wiesel 28.09.21). Wie viele Personen darüber hinaus im ZDF mit der Verifizierung audiovisueller Inhalte betraut sind, bleibt unklar. Auch die DW-Programme werden seit 2020 durch zwölf Faktenchecker:innen unterstützt (vgl. Pörzgen 19.08.22).

In den Zeitungsredaktionen der Frankfurter Allgemeinen Zeitung (FAZ) und der ZEIT gibt es hingegen keine dezidierten Faktencheck-Einheiten. Stattdessen setzen die Bildredaktionen vorwiegend auf zuverlässige Quellen und würden Bildmaterial aus unklaren Quellen nur verwenden, wenn es sich gut mithilfe klassischer Recherche oder Metadaten verifizieren lässt (vgl. Pörzgen 19.08.22; Bartlitz 2022). Die Süddeutsche Zeitung (SZ) ist hingegen der Meinung, dass es heutzutage zu einer fundierten Berichterstattung gehöre, Inhalte aus sozialen Medien nicht auszuklammern und baut bspw. auf Recherchen von Bellingcat auf, nachdem sie den Rechercheweg selbst noch einmal vollzogen hat (vgl. Weinmann 02.03.22). Das investigative Rechercheprojekt zu den *Xinjiang Police Files* zeigt beispielhaft, dass für die Überprüfung von Inhalten schon heute nicht nur zwischen verschiedenen Medienhäusern kooperiert wird, sondern in Einzelfällen auch auf die Expertise von IT-Forensiker:innen zurückgegriffen wird. So ließ der SPIEGEL im vorliegenden Fall ausgewählte Bilder durch das Fraunhofer-Institut für Sichere Informationstechnologie (SIT) analysieren. Außerdem zeigt das Projekt, dass auch der SPIEGEL in Einzelfällen auf OSINT-Verfahren zurückgreift, obwohl gerade die Printausgabe normalerweise ähnlich wie FAZ und ZEIT hauptsächlich Bildmaterial aus zuverlässigen Quellen verwendet (vgl. Pörzgen 19.08.22; Epp u. Höfner 24.05.22). Der alltägliche Umgang des SPIEGEL mit audiovisuellem Material wird im empirischen Teil der Arbeit noch näher beleuchtet.

Besonders unter Druck stehen Agenturen, da Agenturinhalte ohne weitere Prüfung von Redaktionen übernommen werden (dürfen). Umso wichtiger ist es, dass Agenturen keine falschen Inhalte verbreiten, um einen „Gießkanneneffekt“ (Jan Ludwig - dpa, zit. n. Pörzgen 19.08.22) zu vermeiden. Wahrscheinlich ist damit zu erklären, dass bei der Nachrichtenagentur dpa insgesamt rund 30 Personen im Faktencheck-Team arbeiten (vgl. ebd.) und die Expertise langfristig auch an alle Redakteur:innen im Haus weitergegeben werden soll (vgl. Stern 2020, 140). Die AFP als eine der größten Nachrichtenagenturen der Welt gibt an, über ein Faktencheck-Netzwerk mit 140 Journalist:innen zu verfügen und damit weltweit führend zu sein (vgl. AFP 2023).

Unterschiede zwischen den Redaktionen gibt es nicht nur in der Existenz von Faktencheck-Einheiten und deren Personalstärke, sondern auch in der Menge der Inhalte, die geprüft werden und der Entscheidung, in welchem Fall audiovisuelles Material verwendet wird. Während beim SPIEGEL bspw. nur die Inhalte der Printausgabe durch die Dokumentarabteilung vollständig überprüft werden und audiovisuelle Inhalte aus der Onlineredaktion keinen zusätzlichen Verifikationsprozess durchlaufen (vgl. Pörzgen 19.08.22) und auch bei der ZEIT für das Printprodukt stärker auf die einzelnen Bilder geschaut werde als für die Online-Inhalte (vgl. Bartlitz 2022), würden für das Nachrichtenangebot des ZDF nicht nur die Inhalte für Sendungen und Website, sondern auch für die Social-Media-Formate überprüft (vgl. ZDF 15.02.23). Auch zeigen sich leichte Variationen darin, wie sehr man sich auf den Verifizierungsprozess bei einer einzelnen Quelle verlässt und wie transparent der vorhergehende Rechercheprozess dargestellt wird (vgl. Bartlitz 2022; Maas u. Ratering 2022, 502; Grimm 16.03.22; Sell u. Oswald 2021, 258; Stern 2020, 126f.).

Der Beschreibung der Verifikationspraxis auf den vergangenen Seiten steht das Urteil des Journalisten und Journalistik-Dozenten Peter Welchering gegenüber, der zumindest bezüglich der Verifikation von Fotos 2020 noch die Meinung vertritt:

„Fälschungen kommt man schon mit den in jeder Redaktion vorhandenen Bordmitteln auf die Spur. Allein: Oftmals fehlen Wille und Know-How.“  
(Welchering 2020, 21)

Problematisch wird in jedem Fall gesehen, dass in kleinen Medienhäusern wie Regionalzeitungen manchmal niemand mit Faktenchecks befasst sei (vgl. Pörzgen 19.08.22).

Es gebe bereits Agenturen, deren Service in der Bereitstellung verifizierter nutzergenerierter Inhalte besteht<sup>20</sup>, doch könnten sie nicht alle Bereiche abdecken (vgl. Fricke u. Thomsen 2019, 302; Sterz 29.10.16). Und vor allem gelten technische Herausforderungen und Möglichkeiten genauso für diese Agenturen wie für Journalist:innen. Eine Untersuchung der Medienethiker:innen Nina Köberer und Marc Sehr stuft das Outsourcing von Verifikationsprozessen per se nicht als problematisch ein, solange „Kriterien zur Verifizierung angesetzt werden wie auch in Redaktionen und darüber hinaus die Themensetzung nach wie vor bei Redaktionen liegt“ (Köberer u. Sehr 2017, 176).

Trotz der beschriebenen Veränderungen, der Bildung von Faktencheckeinheiten innerhalb von Redaktionen und in Form eigenständiger Agenturen und Recherchenetzwerke stellen die Informatiker:innen Khan et. al (vgl. 2023, 6748f.) fest, dass Medienschaffende sich nach wie vor schwer damit täten, visuelle nutzergenerierte Inhalte schnell und gründlich zu verifizieren, da die Überprüfung ein hohes Maß an Fachwissen erfordere. Auch wenn es forensische Software zur Verifizierung visueller Inhalte gebe, würden in Redaktionen vorwiegend einfache Werkzeuge genutzt. Um das zu ändern seien zwei Dinge wichtig: Zum einen müssten die Softwareentwickler:innen den Verifikationsablauf in Redaktionen vollständig verstehen, um bedarfsgerechte Hilfsmittel entwickeln zu können. Zum anderen sei es wichtig, die Akzeptanz, Einbettung und Schulung der durch die IT-Forensik vorgeschlagenen Hilfsmittel in den Redaktionen zu steigern. Schlussendlich kommen sie auf Grundlage ihrer Analyse der journalistischen Verifizierungspraktiken für audiovisuelle Inhalte zu dem Schluss dass es (halb-)automatisierte Verifizierungstools brauche, um Journalist:innen im Kampf gegen die zunehmende Menge an visuellen Fehlinformationen im Internet zu unterstützen (vgl. Khan et al. 2023, 6748ff.).

---

<sup>20</sup> Z. B. Storyful (<https://storyful.com/>), zuletzt geprüft am 24.03.23

### 3 Methodik der empirischen Erhebung

#### 3.1 Leitfragen

Aus Forschungs- und Sachstand ergibt sich eine Vielzahl unterschiedlicher Fragen, deren Bearbeitung für eine Beantwortung der eingangs genannten Forschungsfrage

**Inwiefern ist und bleibt es möglich, die Authentizität audiovisuellen Materials in Redaktionen sicherzustellen?**

notwendig ist und sich bisher nicht ausreichend in der Forschungsliteratur wiederfindet. Als besonders wichtig wurden die folgenden Fragestellungen empfunden. Sie werden nach Jochen Gläser und Grit Laudel in der vorliegenden Arbeit als **Leitfragen** zum „Bindeglied zwischen den theoretischen Vorüberlegungen und qualitativen Erhebungsmethoden“ und „charakterisieren das Wissen, das beschafft werden muss, um die Forschungsfrage zu beantworten“ (2010, 90f.):

1. Welche Manipulationsformen sind unter welchen Voraussetzungen zurzeit technisch möglich?
2. Mit welchen Manipulationsformen sind Gesellschaft und Journalismus aktuell tatsächlich konfrontiert?
3. Wie lassen sich die unterschiedlichen Möglichkeiten, audiovisuelles Material zu verifizieren, im Rahmen der Multimedia-Forensik verorten?
4. Inwiefern sind die theoretisch betrachteten Methoden der Multimedia-Forensik in der Praxis dafür geeignet, zwischen authentischem Material und Manipulationen zu unterscheiden?
5. Wie wird mit audiovisuellem Material in Redaktionen und Recherchenetzwerken umgegangen, welche technischen Verifizierungsinstrumente werden genutzt und welche Probleme gibt es?
6. Wie wird die aktuelle und zukünftige Detektierbarkeit audiovisuellen Materials schlussendlich eingeschätzt?
7. Wie wird das Risiko audiovisueller Manipulationen für den Journalismus und die Gesellschaft allgemein wahrgenommen?

#### 8. Was muss in welchen Bereichen getan werden?

Um sich den genannten Fragen zu nähern, wurden **qualitative Methoden der empirischen Sozialforschung** (vgl. Lamnek u. Krell 2016; Gläser u. Laudel 2010, 36f.) eingesetzt.<sup>21</sup> Qualitative Forschung ist in der Lage „eine sinnvolle Erkundung neuer und theoretisch noch wenig strukturierter Gegenstandsbereiche zu leisten“ (Lamnek u. Krell 2016, 95). Mit quantitativen Verfahren kann hingegen nur das erhoben werden „von dem wir vorab wissen oder vermuten, dass es alle Fälle gemeinsam haben“ (Gläser u. Laudel 2010, 37), was sich aufgrund des komplexen, unzureichend erschlossenen und stark dynamischen Forschungsfeldes nicht anbietet<sup>22</sup>. Die Zielsetzung des **induktiven** Vorgehens besteht darin, mithilfe **explorativer** Befragungen alle Informationen zu sammeln und zu analysieren, die für eine erste Beantwortung der Forschungsfrage nötig sind, und gleichzeitig weitere notwendige Forschungsbereiche und eventuell auch bereits mögliche Handlungsvorschläge zu extrahieren: „Die erhobenen Daten werden nicht zur Falsifikation von vorab formulierten Hypothesen verwendet, sondern zur Gewinnung solcher Hypothesen auf der Basis des Materials und auf dem Wege der Interpretation genutzt.“ (Lamnek u. Krell 2016, 479)

Um eine **intersubjektive Kontrollierbarkeit** des Forschungsvorgehens zu gewährleisten (vgl. Lamnek u. Krell 2016, 36f.; Gläser u. Laudel 2010, 32), wird im Folgenden das systematische Vorgehen bei der Vorbereitung, Durchführung und Auswertung der geführten Interviews beschrieben.

### 3.2 Qualitative Experteninterviews

**Qualitative Interviews** gelten als Kernelement empirischer Sozialforschung und als „wichtige Instrumente, um detailreiches Wissen [...] über gesellschaftliche Problemlagen zu erhalten“ (Misoch 2019, V) – vorausgesetzt, die angestrebten Informationen „können durch Akteure des Feldes verbalisiert werden und sind damit kommunizier-

---

<sup>21</sup> Dazu wurde außer der Methodenliteratur von Jochen Gläser und Grit Laudel (2010) sowie von Siegfried Lamnek und Claudia Krell (2016) für die Durchführung der Interviews auf eine Publikation von Sabine Misoch (2019) zurückgegriffen, während die Auswertung der Interviews auf der Methodik von Udo Kuckartz (2018) beruht.

<sup>22</sup> Zumindest mit den natürlichen Möglichkeiten und dem vorgegebenen Umfang dieser Arbeit – Ideen für weitergehende Forschung, auch quantitativ, finden sich am Ende dieser Arbeit

bar“ (Gläser u. Laudel 2010, 103). Dies ist hier der Fall. Für einen Großteil der Leitfragen käme eine andere Erhebungsmethode, bspw. eine Beobachtung (vgl. ebd.), auch gar nicht in Frage. Nach der Differenzierung von Siegfried Lamnek und Claudia Krell begründet die Intention der Erhebung die Durchführung **ermittelnder, informatorischer Interviews**, wobei sich Lamnek und Krell hierbei auf die Definition Jürgen von Koolwijks (vgl. 1974, zit. n. Lamnek u. Krell 2016, 315ff.) stützen.

Als Interviewpartner:innen wurden gezielt Personen ausgewählt, die sich „inhaltlich adäquat im Hinblick auf die Forschungsfrage erweisen und die reichhaltige Informationen zu dieser zu liefern versprechen“ (Misoch 2019, 200), wie es beim **Sampling** im Rahmen qualitativer Forschung gefordert wird. Eine herkömmliche Stichprobe und entsprechende Stichprobenziehung wäre bei dem vorliegenden Forschungsvorhaben weder machbar noch zielführend. Insgesamt „richtet sich doch bei ihnen das Interesse weniger auf die zahlenmäßige Verteilung bestimmter Merkmale als auf die Erkenntnis wesentlicher und typischer Zusammenhänge“ (Lamnek u. Krell 2016, 177). Im Gegensatz zu quantitativen Erhebungen wird also keine statistische Repräsentativität angestrebt, sondern eher eine exemplarische bzw. inhaltliche, welche durch die entsprechende Auswahl der Interviewpartner:innen erreicht werden kann (vgl. Misoch 2019, 202). Dafür wurde darauf geachtet, dass die Interviewpartner:innen möglichst weit gefächerte Perspektiven auf das Forschungsgebiet bieten.

Aus dem Zweck der Interviews folgt die Entscheidung zur Durchführung **vorwiegend rekonstruierender Experteninterviews** (vgl. Misoch 2019, 280f.; Gläser u. Laudel 2010, 13). Anders als andere Interviewformen ist das Experteninterview nicht inhaltlich oder methodologisch definiert, sondern bloß im Hinblick auf die Funktion der Interviewten: „Im Experteninterview selbst stehen die Interviewten nicht als individuelle Personen im Vordergrund, [...] sondern als Funktionsvertreter und somit als spezielle Wissensträger“ (Misoch 2019, 120; vgl. 119ff.; vgl. Lamnek u. Krell 2016, 687). Umso wichtiger ist es klar zu definieren, worum es sich bei den Wissensträger:innen handelt. In dieser Arbeit wird ein eng gefasster **Expertenbegriff** angewandt, so dass alle interviewten Personen „durch Sonderwissen und institutionalisierte Kompetenz, d. h. ihre Berufsrolle, gekennzeichnet sind“ (Misoch 2019, 126).

Gleichzeitig erfüllen die Interviews durch ihre Problemzentrierung und ihre Prozesshaftigkeit durchaus einzelne Kriterien des ‚**problemzentrierten Interviews**‘ nach der Definition Andreas Witzels (zit. n. Misoch 2019, 71ff.). Die Problemzentrierung bedarf wohl keiner weiteren Erläuterung. Es wurden zudem die Leitfäden der folgenden Interviews nach jedem Interview geprüft und ggf. ergänzt. Außerdem lag der Fokus der Erhebung überwiegend auf der Rekonstruktion sozialer Sachverhalte, aber nicht ausschließlich. Einzelne Leitfragen zielen auch auf die Ermittlung subjektiver Sichtweisen und Einschätzungen, um die Beantwortung der Forschungsfrage nicht nur für den Moment zu erreichen, sondern zusätzlich einen Ausblick in die Zukunft zu ermöglichen. Auch das sorgt für eine gewisse Nähe zur Methodik problemzentrierter Interviews. Es zeigt sich, dass eine scharfe Trennung zwischen Interviews mit rekonstruierender Funktion und solchen, die „Deutungen, Sichtweisen und Einstellungen von Menschen“ erfassen, hier die Möglichkeiten der Untersuchung künstlich beschneiden würde und daher nicht sinnvoll erscheint. Dieser Fall wird auch in der Forschungsliteratur beschrieben: In der Regel dominiere eine der beiden Funktionen, an der das Interview zu orientieren sei (vgl. Gläser u. Laudel 2010, 13f.). Daher werden die geführten Interviews trotz der genannten Abweichungen als Experteninterviews verstanden, welche als **Einzelbefragungen** geplant und durchgeführt wurden (vgl. ebd., 43).

### *3.2.1 Vorstellung der Interviewpartner:innen*

Insgesamt wurden für den empirischen Teil dieser Arbeit sieben Interviews geführt. Zusammen ergibt sich eine Kombination aus redaktionellen und technischen Expert:innen, die es erlaubt, Informationen zu allen formulierten Leitfragen zu sammeln. Soweit möglich, wurden Personen ausgewählt, die Wissen zu beiden Bereichen vereinen. Zum einen, um möglichst spezifisches Wissen zu erhalten, aber auch, damit ein möglichst großer Teil der angestrebten Erkenntnisse durch mehrere Interviewte abgedeckt werden kann. Auf diese Weise können die Ergebnisse stärker empirisch abgesichert werden, was laut Gläser und Laudel bereits als **Triangulation** verstanden werden kann (vgl. Gläser u. Laudel 2010, 104f., 117). So ist beispielsweise der IT-Forensiker Dr. Sascha Zmudzinski vom Fraunhofer SIT auch mit der Schulung von Journalist:innen betraut und Dr. Christian Riess war bereits als Gast des Bayerischen Journalisten-

Verbandes mit der redaktionellen Seite im Austausch<sup>23</sup>, während Johanna Wild als Mitarbeiterin des Recherchenetzwerks Bellingcat sich dort auch um die Entwicklung technischer Hilfsmitteln kümmert. Eine Differenzierung der Interviews in die Erhebung von **Kontext- und Betriebswissen** (vgl. Misoch 2019, 120f.) lässt sich kaum vornehmen, da die redaktionellen Interviewpartner:innen als Expert:innen auf ihrem Gebiet teilweise auch technisches Kontextwissen mitbringen, genauso wie die Informatiker:innen teils selbst Betriebswissen, beispielsweise in Form der Arbeitsweise des Fraunhofer SIT, beisteuern. Es folgt eine detailliertere Beschreibung der Expert:innen mit ihrem jeweils beschriebenen Kontext- und/oder Betriebswissen:

---

### **01 Dr. Dominique Dresen – Bundesamt für Sicherheit in der Informationstechnik**

---

Dr. Dominique Dresen ist Referent beim Bundesamt für Sicherheit in der Informationstechnik (BSI). Dort arbeitet er im Referat Bewertungsverfahren für eID-Technologien in der Digitalisierung und beschäftigt sich mit dem Thema Deepfakes, insbesondere der Gesichtsmanipulation.

Das BSI ist die Cyber-Sicherheitsbehörde des Bundes. Der Schutz der IT-Systeme des Bundes ist daher Aufgabenbereich des BSI. Darüber hinaus informiert und berät das BSI Behörden, Unternehmen und Bürger:innen in Fragen der IT- und Internetsicherheit. Es prüft, bewertet und zertifiziert IT-Systeme, Dienstleistungen und Personen. Genauso beaufsichtigt es Betreiber Kritischer Infrastrukturen und digitaler Dienste. Außerdem werden Standards, IT-Sicherheitsanwendungen und -produkte entwickelt. Und zu guter Letzt ist das BSI auch die zentrale Meldestelle für IT-Sicherheitsvorfälle und Cyber-Angriffe in Deutschland. Insgesamt wird das Ziel des BSI beschrieben als „präventive Förderung der Informations- und Cyber-Sicherheit, um den sicheren Einsatz von Informations- und Kommunikationstechnik in unserer Gesellschaft zu ermöglichen und voranzutreiben“ (BSI). Dafür beschäftigt sich das BSI mit den Risiken der Digitalisierung und wie man sie beherrschbar macht (vgl. ebd.).

---

---

<sup>23</sup> Siehe <https://www.bjv.de/news/podiumsdiskussion-fotografie-wahr-oder-falsch-bilder-im-roentgenblick>, zuletzt geprüft am 03.06.2023



---

## **02 Johanna Wild – Bellingcat**

---

Johanna Wild war bereits als Journalistin u. a. für die SZ, den BR, ARTE und weitere Medien im Ausland tätig, arbeitet inzwischen für das Recherchenetzwerk Bellingcat und beschäftigt sich dort mit datenintensiven Recherchen, aber auch mit der Entwicklung von Technologien und Tools für digitale Ermittlungen.

Bellingcat ist ein international ausgerichtetes, investigatives Recherchenetzwerk, das von seinem Gründer auch als „Nachrichtendienst für jedermann“ (Higgins 2021, 15; vgl. Stern 2020, 122) bezeichnet wird. Das Recherchenetzwerk gilt als Vorreiter im Bereich der OSINT-Recherchen und hält vermutlich die größte Sammlung an frei zugänglichen Verifizierungsinstrumenten bereit (vgl. Stern 2020, 122; PresseClub München 07.04.22, Weinmann). Das Netzwerk ist eine gemeinnützige Organisation. Insgesamt arbeitet bei Bellingcat, Stand 2021 (vgl. Higgins 2021, 8), eine Kernmannschaft von 18 Personen, die um viele Freiwillige weltweit ergänzt wird. Die ersten Festangestellten wurden vor mehreren Jahren mit finanzieller Unterstützung von Google eingestellt. Insgesamt werde zum Wohl der Unabhängigkeit aber darauf geachtet, dass das Geld aus möglichst vielen verschiedenen Quellen komme (vgl. Higgins 2021, 8, 148, 163f.; Bischoff 2022b, 65). Zum Durchbruch verhalf dem Netzwerk die Recherche zum Abschuss des Passagierflugzeugs MH17, deren Ergebnisse sie im November 2014 veröffentlichten (vgl. Higgins 2021, 130, 141).

„Wir sind keine richtigen Journalisten, wir sind auch keine hauptamtlichen Menschenrechtsaktivisten, auch keine Computerexperten, Archivare, keine wissenschaftlichen Forscher, justiziellen Ermittler, sondern eine neuartige Mischung aus allen diesen Bereichen.“ (Higgins 2021, 287)

---

## **03 Dr. Christian Riess – Friedrich-Alexander-Universität Erlangen-Nürnberg**

---

Dr. Christian Riess ist Gruppenleiter für Multimediasicherheit am Lehrstuhl für IT-Sicherheitsinfrastrukturen der Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg. Riess publiziert seit mehr als einem Jahrzehnt zahlreiche Forschungsergebnisse zu verschiedenen Ansätzen, die sich mit der Verifizierung audiovisuellen Materials beschäftigen. Viele Publikationen beschäftigen sich mit physikbasierten

---

Detektionsmethoden klassischer Manipulationen, aber auch synthetische Medien, der Einsatz maschinellen Lernens in der Multimedia-Forensik und Verifikationsprozesse in Redaktionen werden thematisiert.

---

#### **04 Dr. Jutta Jahnel – Karlsruher Institut für Technologie (KIT)**

---

Dr. Jutta Jahnel beschäftigt sich als wissenschaftliche Mitarbeiterin am Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) des Karlsruher Instituts für Technologie (KIT) mit der gesellschaftlichen Dimension lernender Systeme. In diesem Zusammenhang berät sie auch das EU-Parlament als Mitautorin einer internationalen Studie in Bezug auf die Regulierung von KI, fokussiert auf Deepfakes.

---

#### **05 Andrea Sauerbier – SPIEGEL**

---

Andrea Sauerbier arbeitet bei der Zeitschrift SPIEGEL als Dokumentarin. Bis 2022 war sie Teil der Abteilung Bilddokumentation und prüfte ausschließlich Fotos und Bildunterschriften für den SPIEGEL, SPIEGEL Geschichte und DEIN SPIEGEL sowie für weitere Print-Nebenprodukte nach Absprache. Seit 2022 gibt es die Abteilung Bilddokumentation beim SPIEGEL nicht mehr. Stattdessen findet die Prüfung des Bildmaterials und der Bildunterschriften für die Print-Produkte des SPIEGELS in der Dokumentationsabteilung zusammen mit der Verifizierung der Textbeiträge statt.

---

#### **06 Dr. Sascha Zmudzinski – Fraunhofer-Institut für Sichere Informationstechnologie**

---

Dr. Sascha Zmudzinski arbeitet als wissenschaftlicher Mitarbeiter des Forschungsbereichs Multimedia-Sicherheit und IT-Forensik am Fraunhofer-Institut für Sichere Informationstechnologie SIT.<sup>24</sup> In dieser Funktion ist er u. a. am laufenden Projekt Disco beteiligt, das einen Überblick über die typischen Vorgehensweisen und technischen Methoden bei der Verbreitung von Desinformation im Kontext aktueller Krisen geben soll. Ziel ist herauszufinden, welche forensischen Methoden sich am bes-

---

<sup>24</sup> Seine erste Publikation zu digitalen Wasserzeichen stammt bereits aus dem Jahr 2004, siehe *Steinebach, Martin et al. (2004): The digital watermarking container: secure and efficient embedding. In: International Multimedia Conference: Proceedings of the 2004 workshop on Multimedia and security; 20-21 Sept. 2004. 2004 ACM. pp. 199–205*

ten für die Detektion von Desinformation eignen. Dafür soll bspw. auch das Leistungsvermögen inverser Bildersuchen über unterschiedliche Suchmaschinen beurteilt, verglichen und der Einfluss zusätzlicher Bearbeitungsschritte auf die (Wieder-)Erkennungsraten untersucht werden (vgl. Fraunhofer SIT).

Zusätzlich arbeitet Zmudzinski als Referent zum Thema Multimedia-Forensik für die Fraunhofer Academy und schult IT-Forensiker:innen, Ermittler:innen und Journalist:innen. Auch bei konkreten Recherchen ist das Fraunhofer SIT Ansprechpartnerin für Redaktionen. So prüfte es bspw. für SPIEGEL und SZ die Echtheit des Ibiza-Videos, das zum Regierungswechsel in Österreich führte und analysierte auch Foto-Material aus den Xinjiang Police Files zur Uiguren-Politik in China.

---

### **07 Jochen Spangenberg – u. a. DW Innovation**

---

Jochen Spangenberg ist stellv. Leiter der Abteilung DW Research and Cooperations Projects (DW ReCo), welche zusammen mit dem DW Lab die Innovationseinheit der DW bildet. Bereits seit 2013 beschäftigt sich Spangenberg insbesondere mit der Bekämpfung von Desinformation und der Verifizierung nutzergenerierter Inhalte und koordiniert(e) in seiner Funktion die bereits in Kapitel 2.4.1 angesprochenen Forschungs- und Innovationsprojekte für die DW.

Außerdem lehrt Jochen Spangenberg an der Freien Universität Berlin im Fachbereich Medien- und Kommunikationswissenschaften, trainiert im Rahmen der DW Akademie ausländische Journalist:innen und unterstützt die NGO Lie Detectors in ihrem Bestreben, die Medienkompetenz von Schüler:innen zu stärken. Darüber hinaus ist er im Beirat von EDMO und GADMO, sowie Mitglied der von EDMO initiierten Sonderarbeitsgruppe zu Desinformation im Rahmen des Krieges in der Ukraine.

Im Vorhinein des mit ihm geführten Interviews hat Jochen Spangenberg nochmal hervorgehoben, dass er im Gespräch weder die DW ReCo oder eine der anderen Organisationen vertritt, denen er angehört, sondern für sich als Person mit seiner Erfahrung durch die Arbeit für die oben aufgeführten Institutionen spricht.

---

### *3.2.2 Interviewleitfäden als Erhebungsinstrument*

Wie bei Experteninterviews üblich, wurde zur besseren Steuerung und Strukturierung der Interviews mit einem Interviewleitfaden als **Erhebungsinstrument** gearbeitet. Dieser dient als Richtschnur, gibt jedoch weder exakte Formulierung noch Reihenfolge der Fragen vor und lässt zudem Raum für spontane Nachfragen. Auch den Interviewten wird ihre Antwort vollkommen freigestellt. Damit wird dem Charakter qualitativer Erhebungen Rechnung getragen, wie es beispielsweise der Sozialwissenschaftler Udo Kuckartz beschreibt, auf dessen Methodenliteratur auch die spätere Auswertung der Interviews beruht: „Ein Charakteristikum qualitativer Methoden ist der Anspruch, die Forschungsteilnehmenden selbst zu Wort kommen zu lassen, ihnen die Gelegenheit zu geben, Antworten in eigenen Worten zu formulieren und nicht lediglich aus einer Anzahl vorformulierter Antworten auszuwählen“ (Kuckartz 2018, 179; vgl. Gläser u. Laudel 2010, 41ff., 142; Lamnek u. Krell 2016, 689f.; Misoch 2019, 65f., 123).

Es handelt sich daher nicht um eine standardisierte Befragung, sondern je nach Autor:in um nicht-standardisierte Leitfadeninterviews (vgl. Gläser u. Laudel 2010, 41f.), teilstandardisierte Befragungen (vgl. Lamnek u. Krell 2016, 318ff.) oder auch halboffene bzw. halb-/semi-strukturierte Interviews (vgl. Misoch 2019, 13f.). Der **Standardisierungsgrad des Interviews** wurde gewählt, um die Vorteile nicht-standardisierter Befragungen in Bezug auf Informationstiefe und geringe Prädetermination mit praktischen Vorzügen bei der Durchführung und Auswertung standardisierter Befragungen zu kombinieren. Durch die leitfadengestützte Befragung soll ein möglichst optimierter Kompromiss aus Flexibilität, Informationsbreite/-tiefe, Zeitaufwand, Vergleichbarkeit und potenziellen Fehlerquellen erreicht werden, der in jedem Fall eine Gewichtung von Themen, Problemen und Lösungen durch die Interviewten ermöglicht: „Eine leitfadenorientierte Gesprächsführung wird dem thematisch begrenzten Interesse des Forschers und dem Expertenstatus des Befragten gerecht, da eine Konzentration auf das funktionsbedingte Sonderwissen erfolgt, gleichzeitig aber auch Freiräume für die spezifische Sichtweise des Experten und unerwartete Themendimensionierungen bestehen“ (Lamnek u. Krell 2016, 690, vgl. 319ff.).

Im Folgenden werden die Schritte der Leitfadenerstellung beschrieben, um die Operationalisierung nachvollziehbar zu machen. Letztlich geht es darum, dass das „aus der Untersuchungsfrage und den theoretischen Vorüberlegungen abgeleitete Informationsbedürfnis in Themen und Fragen des Leitfadens übersetzt wird“ (Gläser u. Laudel 2010, 115), Leitfragen also zum Leitfaden werden. Dafür muss das Erkenntnisinteresse in den Kontext des Erfahrungshintergrundes der Befragten übersetzt werden. Der Erfahrungshintergrund wurde an den recherchierten und bereits genannten Informationen zu den Interviewpartner:innen festgemacht, bei gegebenen Publikationen auch anhand derer. Wie empfohlen und aufgrund der unterschiedlichen Erfahrungshintergründe der Befragten auch kaum anders möglich, wurde jedes Interview individuell vorbereitet, um die Expertise der einzelnen Personen voll auszuschöpfen (vgl. Gläser u. Laudel 2010, 95, 116f., 142, 150-151).

Die Interviewleitfäden orientieren sich an den Anforderungen und Regeln, die mehr oder weniger ausführlich in allen drei genannten Publikationen aufgeführt werden. Dies betrifft zunächst den Umfang und die Struktur der Interviewleitfäden, sowie die Reihenfolge der Fragen. Die Fragen wurden entsprechend nach Themen gruppiert, um einen möglichst natürlichen Gesprächsverlauf zu ermöglichen. Darüber hinaus „muss ein hinreichend breites Spektrum von Problemen angesprochen werden, damit die Befragten eine möglichst große Chance haben, in nicht antizipierter Weise zu reagieren“ (Gläser u. Laudel 2010, 116), der Interviewleitfaden muss spezifisch sein und tiefgehende Darstellungen der Interviewten unterstützen. Das Maß an Offenheit der Fragen wurde versucht so auszubalancieren, dass die Formulierung der Fragen möglichst wenig Einfluss auf den Inhalt der Antworten ausübt und gleichzeitig ausreichend konkret ist, um von den interviewten Personen in der begrenzten Zeit alle gewünschten Informationen zu erhalten (vgl. Gläser u. Laudel 2010, 115f., 131, 144, 146; Lamnek u. Krell 2016, 334; Misoch 2019, 66ff.).

Als „Anwärmfrage“ (Gläser u. Laudel 2010, 147f.) wurde jeweils eine möglichst offene Frage gestellt, die als Erzählanregung verstanden werden kann, und gleichzeitig den Bezug der entsprechenden Person mit dem Forschungsthema überprüft und mit Details versieht. Anschließend folgen weitere Fakt- aber auch Meinungsfragen. Die Fakt-

fragen wurden in den Leitfäden, wie gefordert, möglichst als offene Fragen bzw. Erzählanregungen formuliert und ggf. durch Detailfragen ergänzt. Während die Faktfragen das Wissen der Expert:innen und ihre Erfahrungen eruieren, wurden Meinungsfragen genutzt, um subjektive Bewertungen und Theorien der Interviewpartner:innen zu erfahren (vgl. Gläser u. Laudel 2010, 145). Davon wurde vor allem Gebrauch gemacht, um die technologische Momentaufnahme<sup>25</sup> durch Einschätzungen zu der zukünftigen Entwicklung relevanter Technologien zu ergänzen. Außerdem wurden Meinungsfragen genutzt, um subjektive Beurteilungen zur Detektierbarkeit audiovisueller Manipulationen, des Risikos für und des Einflusses auf den Journalismus und notwendiger Handlungsfelder einzuholen. Alle diese Einschätzungen beruhen logischerweise auf dem spezifischen Wissen der Expert:innen und stellen damit einen Mehrwert dar, auch wenn sie nicht mit Fakten gleichgesetzt werden dürfen und in der Auswertung der Interviews entsprechend behandelt werden. Gläser und Laudel warnen davor, sich in Experteninterviews mit subjektiven Theorien der Interviewpartner:innen zufrieden zu geben, bekennen aber gleichzeitig, dass die Meinung der Interviewpartner:innen wichtige Informationen beinhalten kann. Sie dürfe bloß nicht die einzige Quelle für die Fakten sein, deren Analyse die Forschungsfrage beantworten soll (vgl. Gläser u. Laudel 2010, 113f.). Insofern erscheint die Kombination aus Faktfragen, zur direkten Beantwortung der Forschungsfrage, und Meinungsfragen, für einen Blick über den Teller- rand, dem vorliegenden Forschungsvorhaben angemessen.

Die konkreten Fragen schließlich wurden mithilfe des von Lamnek und Krell (vgl. 2016, 334) aufgegriffenen SPSS- bzw. S<sup>2</sup>PS<sup>2</sup>-Verfahren erschlossen: Bereits während der Erstellung des Theorieteils wurden möglichst viele Fragen gesammelt, anschließend nach Interviewpartner:innen sortiert, geprüft und reduziert, sinnvoll angeordnet und schließlich in den passenden Formulierungen zum jeweiligen Interviewleitfaden geformt. Es wurde darauf geachtet, dass die Angaben der Interviewten zu einigen Aspekten vergleichbar sind, indem einzelne Fragestellungen bei mehreren oder teilweise sogar allen Interviews angesprochen wurden (vgl. Gläser u. Laudel 2010, 152; Misoch 2019, 13). Ziel war eine möglichst unbeeinflusste Darstellung der Thematik durch die

---

<sup>25</sup> Welche eines der zentralen Forschungsanliegen war, das durch Faktfragen bearbeitet werden kann.

Interviewten. Zur Stärkung der Interviewoffenheit wurde am Ende (vgl. Gläser u. Laudel 2010, 148f.) zusätzlich gebeten, Aspekte zu ergänzen, die im Interview nach Meinung der Interviewten nicht genug angesprochen wurden.<sup>26</sup>

Die folgende Tabelle gibt einen Überblick über die Themen, aus denen sich die jeweiligen Interviewleitfäden zusammensetzen und greift auch bereits vor, welche Themen in den jeweiligen Interviews tatsächlich angesprochen worden sind (x bzw. (x)):

| Themen   | Interview  |            |            |            |            |            |            |
|--|------------|------------|------------|------------|------------|------------|------------|
|  | DD<br>(01) | JW<br>(02) | CR<br>(03) | JJ<br>(04) | AS<br>(05) | SZ<br>(06) | JS<br>(07) |
| <b>1 Manipulationsmöglichkeiten</b>                    | x          |            | (x)        | (x)        |            | (x)        |            |
| Deepfakes  | x          |            | x          | x          |            | x          |            |
| <b>2 Multimedia-Forensik</b>                           | x          | (x)        | x          | x          |            | x          |            |
| Systematisierung Multimedia-Forensik                   |            |            | x          |            |            | x          |            |
| Passive Forensik                                       |            |            | x          |            |            | x          |            |
| Detektion von Deepfakes                                | x          |            | x          | x          |            | x          |            |
| Aktive Forensik  | x          | x          | x          | x          |            | x          |            |
| <b>3 Praxis</b>  | (x)        | (x)        | (x)        | (x)        | (x)        | (x)        | (x)        |
| Vorkommende Manipulationsformen                        | x          | x          |            | x          | x          | x          | x          |
| Praxis IT-Forensik                                     | x          |            | x          |            |            | x          |            |
| Praxis in Redaktionen / Recherchenetzwerken            |            | x          | x          | x          | x          | x          | x          |
| Zusammenarbeit (akt.) zw. den Akteur:innen             |            | x          |            |            |            | x          |            |
| <b>4 Einschätzung Detektierbarkeit AV-Medien insg.</b> | (x)        | (x)        | (x)        | (x)        | (x)        | (x)        | (x)        |
| Aktuelle Detektierbarkeit AV-Medien                    | x          | x          | x          | x          |            | x          | x          |
| Zukünftige Detektierbarkeit AV-Medien                  | x          | x          | x          | x          | x          | x          | x          |
| <b>5 Einschätzung Einfluss und Risiken</b>             | x          | x          | x          | x          | x          | x          | x          |
| <b>6 Einschätzung notwendiger Handlungsfelder</b>      | x          | x          | x          | x          | x          | x          | x          |

Tabelle 2: Überblick über die in den Experteninterviews angesprochenen Themen (Grafik: Martha Peters)

<sup>26</sup> Zwei Dinge haben die Fragestellung insgesamt vereinfacht und die Gefahr minimiert, durch unausgereifte Formulierungen unbrauchbare Ergebnisse zu erhalten. Zum einen handelt es sich kaum um heikle Themen für die interviewten Personen, zu denen auch kaum sozial erwünschte Antworten präsent sind. Zum anderen sind besonders „Wissenschaftler [...] daran gewöhnt, in Diskussionen die Argumentation ihres Gesprächspartners zu analysieren und ihre eigene Position zu behaupten“ (Gläser u. Laudel 2010, 135) – Personen mit journalistischem Hintergrund nicht weniger. Es kann kritisiert werden, dass die der Alltagskommunikation nahen dichotomen Fragen nicht so konsequent vermieden wurden, wie es ideal gewesen wäre. Aufgrund des gerade beschriebenen Hintergrunds der Befragten stellte dies jedoch kein Problem dar. Zudem wurde im Interviewverlauf darauf geachtet, dass im Zweifelsfall eine offene Nachfrage ergänzt wurde und die dichotomen Fragen nicht zu einer negativen Kettenreaktion führen konnten (vgl. Gläser u. Laudel 2010, 131ff.).

Die Leitfäden wurden, wie gefordert, abschließend noch einmal kontrolliert (vgl. Gläser u. Laudel 2010, 145, 149). Die vollständigen Leitfäden für alle durchgeführten Experteninterviews sind im Anhang der Arbeit zu finden, wobei der jeweils genutzte ‚Vorspann‘ mit Notizen über Sachverhalte, die im Vorhinein des Interviews zu beachten oder zu klären waren, aus Platzgründen herausgenommen wurde, da er keinen inhaltlichen Mehrwert bietet.

### *3.2.3 Datenerhebung und -erfassung*

Die Interviews wurden im Zeitraum vom 12.04.2023 bis zum 09.05.2023 geführt, nachdem der erste Kontakt zu den Interviewpartner:innen über schriftliche Anfragen per Mail bereits Ende Februar / Anfang März stattfand. Alle Interviews, bei denen ein persönliches Treffen aufgrund der räumlichen Distanz umständlich gewesen wäre oder von den Gesprächspartner:innen nicht angestrebt wurde, fanden per Videokonferenzsystem statt. Ob das Interview vor Ort oder online durchgeführt wurde, wird im Rahmen der Auswertung nicht weiter beachtet, da es sich bei Videointerviews um eine „relativ reichhaltige und dem Face-to-face-Austausch ähnliche Kommunikationssituation“ (Misoch 2019, 177) handelt. Die meisten Argumente, die für Face-to-Face-Interviews angeführt werden, sind bei Videointerviews ebenfalls gegeben. Kritikpunkte gegenüber Telefoninterviews, die größtenteils auf dem fehlenden visuellen Kontext beruhen (vgl. Gläser u. Laudel 2010, 153f.), lassen sich somit nicht auf Videointerviews übertragen. Die Interviews wurden mit einem „weichen bis neutralen“ (Lamnek u. Krell 2016, 325f.) Kommunikationsstil durchgeführt. Es wurde eine Interviewlänge von 30-45 min angestrebt, die tatsächliche Interviewlänge variiert zw. 25 min und 1:20 h.

Alle Interviews wurden, wie bei Experteninterviews geboten (Gläser u. Laudel 2010, 157f.; vgl. Lamnek u. Krell 2016, 335), mithilfe einer Audioaufzeichnung protokolliert, welche bereits im Rahmen der E-Mailanfrage abgesprochen wurde. Im Anschluss an die Durchführung der einzelnen Interviews wurde zudem ein Interviewbericht als interviewbegleitende Dokumentation angefertigt, der den später erstellten Transkriptionen vorangestellt ist.



### 3.2.4 *Transkription und Interviewbericht*

Anhand der Audioaufzeichnungen wurde eine vollständige Transkription der verbalen Interviewdaten in Standardorthografie angefertigt. Die Entscheidung gegen eine selektive Transkription, wie sie mitunter für diese Interviewform vorgeschlagen wird (vgl. Misoch 2019, 123ff.), soll verhindern, dass der Interpretation vorgegriffen wird.

Für die Erstellung der Transkripte wurden die Audiodateien zunächst durch das Spracherkennungssystem Whisper<sup>27</sup> automatisiert transkribiert. Die Ergebnisse des Programms wurden durch einen manuellen Korrekturdurchgang kontrolliert und angepasst. Nichtverbale Äußerungen (Lachen, Räuspern etc.) wurden nur dann transkribiert, wenn sie den Sinn der Aussage verändern. Die genauen Transkriptionsregeln befinden sich zusammen mit den vollständigen Interviewtranskriptionen auf dem digitalen Datenträger zu dieser Arbeit. Sie sind das Ergebnis eines Abgleichs verschiedener Vorschläge aus der Methodenliteratur (vgl. Misoch 2019, 252f., 261; Kuckartz 2018, 166ff.; Gläser u. Laudel 2010, 192ff.). Da es sich um Experteninterviews handelt, bei denen die Interviewten als Wissensträger:innen und nicht als Individuen im Fokus stehen, wird dem nonverbalen Inhalt eine vergleichsweise geringe Funktion zugeschrieben. Vor diesem Hintergrund wurde aufgrund der Einfachheit und des geringeren Störfaktors (vgl. Lamnek u. Krell 2016, 368) auf eine Videoaufzeichnung und eine literarische Transkription verzichtet. Da der Hintergrund der Expert:innen für die Einordnung und Gewichtung ihrer Aussagen relevant ist, wurde mit ihnen abgesprochen, auf eine Anonymisierung zu verzichten.

## 3.3 **Eine qualitative Inhaltsanalyse nach Udo Kuckartz**

Die angefertigten Transkriptionen bilden die Grundlage für eine qualitative Inhaltsanalyse, die hier in Form einer inhaltlich-strukturierenden Inhaltsanalyse durchgeführt worden ist. Die inhaltlich-strukturierende Inhaltsanalyse gilt als zentrale Variante der qualitativen Inhaltsanalyse. Sie ermöglicht die systematische Beschreibung und Zusammenfassung der Interviewinhalte, so dass das Wissen und die Einschätzungen der Expert:innen zur Beantwortung der Forschungsfrage genutzt werden können: „Kern

---

<sup>27</sup> Ein KI-basiertes Open-Source-Programm der Firma OpenAI, das lokal betrieben werden kann

der inhaltlich-strukturierenden Vorgehensweise ist es, am Material ausgewählte inhaltliche Aspekte zu identifizieren, zu konzeptualisieren und das Material im Hinblick auf solche Aspekte systematisch zu beschreiben“ (Schreier 2014). Für die Konzeptualisierung wird anschließend an die initiierende Textarbeit ein Kategoriensystem erstellt, das die relevanten Aspekte der Erhebung darstellt und ermöglicht, die Äußerungen der Interviewpartner:innen Kategorien zuzuordnen und damit thematisch zu sortieren (vgl. ebd; vgl. Kuckartz 2018, 97ff.).

Der erste bewusst als Methodenlehrbuch geschriebene Text über eine *qualitative Inhaltsanalyse* dem beschriebenen Sinne nach stammt von Philipp Mayring, weshalb der Begriff im deutschsprachigen Raum besonders mit den von Mayring vorgestellten Analyseformen gleichgesetzt wird (vgl. Kuckartz 2018, 23, 26). Allerdings wird kritisiert, dass Mayring „zwar die Entwicklung des Kategoriensystems ‚öffnet‘, sich aber ansonsten sehr stark an der quantitativen Inhaltsanalyse orientiert“ (Gläser u. Laudel 2010, 198). Tatsächlich geht Mayring davon aus, dass die Ergebnisse der Analysen meist quantitativ z. B. durch das Auszählen von Kategorienhäufigkeiten weiterverarbeitet werden, womit sie auf eine statistische Analyse hinauslaufen (vgl. Mayring 2015, 8; Kuckartz 2018, 6, 27). So würden durch „das Mayringsche Verfahren letztlich Häufigkeiten analysiert, anstatt Informationen zu extrahieren“ (Gläser u. Laudel 2010, 199).

In der vorliegenden Arbeit wird stattdessen vorwiegend auf die von Udo Kuckartz beschriebene Form der Inhaltsanalyse zurückgegriffen, die an Mayrings Ansatz anknüpft. Nicht allein aufgrund der stärkeren qualitativen und hermeneutischen Orientierung (vgl. Kuckartz 2018, 6), die dem Forschungsinteresse und der Anzahl der Interviewpartner:innen angemessen erscheint, sondern auch, um in der Analyse möglichst nah an den originalen Äußerungen der Gesprächspartner:innen zu bleiben. Besonders wegen der erhobenen Meinungsfragen sind die wörtlichen Äußerungen der Interviewpartner:innen durchaus von Interesse. Während Mayrings Analyse auf Paraphrasen aufbaut (vgl. Mayring 2015, 103) und auch die durch Gläser und Laudel beschriebene Form der qualitativen Inhaltsanalyse „sich frühzeitig und konsequent vom Ursprungstext trennt“ (Gläser u. Laudel 2010, 200; vgl. Schreier 2014), erlaubt Kuckartz die Analyse am Wortlaut der Interviewten. Zudem bezieht Kuckartz (vgl. 2018) die

Möglichkeit der computerunterstützten Analyse sehr ausführlich mit ein. In dieser Arbeit wurde die qualitative Inhaltsanalyse mithilfe des Programms MAXQDA durchgeführt. Anders als von Kuckartz vorgeschlagen, wurden jedoch aufgrund der übersichtlichen Anzahl an Interviewten und ihrer Funktion als Expert:innen keine Fallzusammenfassungen angefertigt. Zudem war es möglich, nicht nur alle Hauptkategorien, sondern auch einen Großteil der Subkategorien deduktiv anhand der Leitfragen bzw. der Interviewleitfäden zu bestimmen. Dadurch erübrigten sich die unterschiedlichen Phasen für die Erstellung und Codierung von Haupt- und Subkategorien, welche zu einem Codiervorgang zusammengefasst wurden (vgl. Kuckartz 2018, 110). Um die Passgenauigkeit des Kategoriensystems sicherzustellen und überraschend angesprochenen Aspekten gerecht zu werden, wurde das Kategoriensystem im Rahmen des Codierprozesses induktiv durch weitere Subkategorien ergänzt und bestehende Kategorien ausdifferenziert. Es handelt sich daher um ein deduktiv-induktives Verfahren (vgl. Schreier 2014). Die Codierung der betroffenen Hauptkategorien wurde anschließend durch einen zweiten Materialdurchlauf kontrolliert. Denn im Rahmen des Codierprozesses werden, wie gefordert, die gesamten Interviewdaten analysiert. Textstellen, die für die Forschungsfrage irrelevant sind, bleiben aber uncodiert (vgl. Kuckartz 2018, 26, 53, 102).

Alternativ lässt sich die durchgeführte Inhaltsanalyse auch losgelöst vom Autor Udo Kuckartz allein anhand ihrer Spezifikationen definieren. So kritisiert die Expertin für qualitative Methodenlehre Margrit Schreier die existierenden Varianten qualitativer Inhaltsanalyse als „willkürlich und keineswegs erschöpfend“ (Schreier 2014) und schlägt stattdessen ein „Konzept des Werkzeugkastens“ (ebd.) vor. Nach diesem Werkzeugkasten lassen sich die zentralen Aspekte der durchgeführten Inhaltsanalyse auch als Extraktion des Wissens und der Einschätzungen von Expert:innen aus Interviews anhand thematischer Kategorien beschreiben, bei der die größtenteils deduktiv erstellten Kategorien induktiv durch Subsumtion ergänzt werden. Anstelle einer Probecodierung im Vorfeld des tatsächlichen Codierprozesses wurde die Zuordnung bei den im Rahmen der Codierung induktiv ergänzten Kategorien durch einen anschließenden Kontrolldurchgang überprüft. Als Analyse- und auch Kontexteinheiten sind die einzelnen Interviews zu verstehen, als Codiereinheiten werden Sinneinheiten gewählt (vgl. Kuckartz 2018, 30f., 41ff.).

Insgesamt wird die qualitative Inhaltsanalyse beschrieben als „kategorienbasierte Methoden zur systematischen Analyse qualitativer Daten“ (Kuckartz 2018, 6). Das Vorgehen sei „*systematisch, regelgeleitet*, an den Gütekriterien der *Validität* und der *Reliabilität* orientiert. Die Bedeutung des Reliabilitätskriteriums zeigt sich darin, dass meist ein intersubjektiv-konsensuales Textverständnis angestrebt wird [...]. Die Bedeutung der Validität spiegelt sich in der Anforderung, das Kategoriensystem so zu erstellen, dass es in der Lage ist, wesentliche Bedeutungsaspekte des Materials zu erfassen“ (Schreier 2014). Für den letzten Punkt hebt Schreier (vgl. ebd.) die Bedeutung eines zumindest teilweise induktiven Vorgehens hervor.

Als „Herzstück“ (Schreier 2014; vgl. Kuckartz 2018, 40) der qualitativen Inhaltsanalyse gilt das Kategoriensystem. Eine übersichtliche Darstellung der Struktur des hierarchischen Kategoriensystems (vgl. Kuckartz 2018, 38f.) liefert die folgende Grafik:

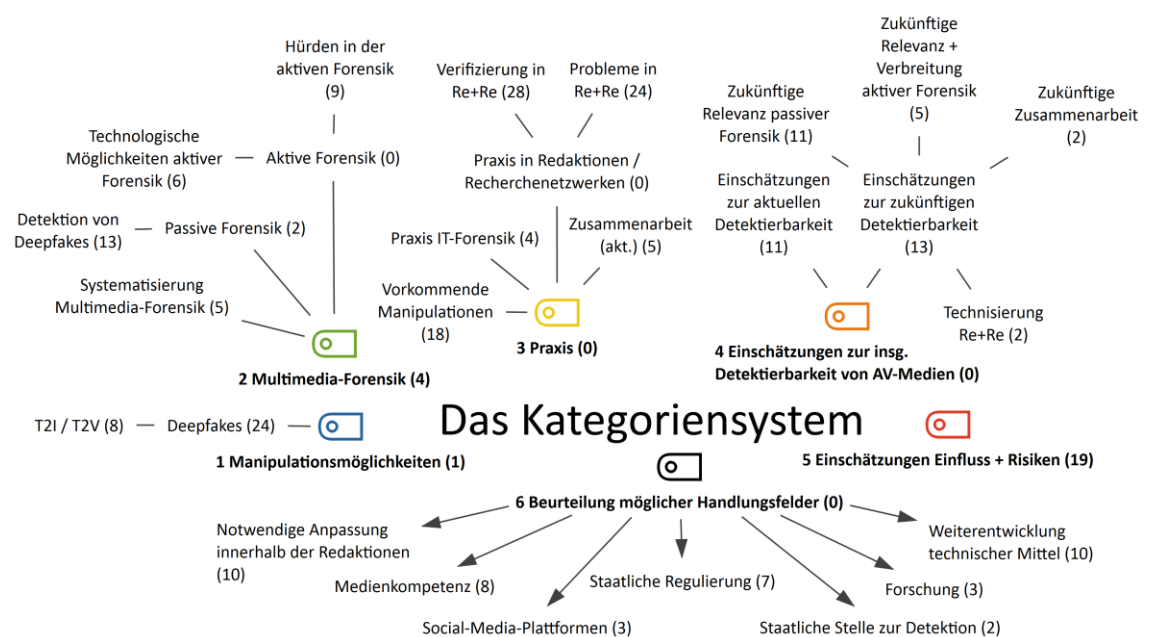


Abbildung 2: Das Kategoriensystem (Grafik: Martha Peters)

Das vollständige Kategorienhandbuch mit den inhaltlichen Beschreibungen, allen (trennscharfen) Kategoriendefinitionen und Anwendungsbeispielen befindet sich im Anhang.

Eine Zusammenstellung aller mit der gleichen Kategorie codierten Textstellen (Text-Retrieval) ermöglicht schließlich einen Überblick über das gemeinsame Wissen und die Einschätzungen der Expert:innen zu den relevanten Fragestellungen.

### **3.4 Probleme und Anmerkungen**

Die durchgeführten Experteninterviews mitsamt der Inhaltsanalyse unterliegen als interaktives Verfahren qualitativer Sozialforschung bereits während der Interviews einer subjektiven Interpretation durch mich als Forschende, was auch bei Zuhilfenahme eines Interviewleitfadens und einer systematischen Auswertung nicht gänzlich zu verhindern ist (vgl. Lamnek u. Krell 2016, 192). Die Darlegung des Forschungs- und Sachstandes vermittelt hierbei gleichzeitig einen Überblick über meine Motivation und mein Vorverständnis für diese Arbeit.

Rückblickend lässt sich insgesamt festhalten, dass der Themenbereich für den Umfang der Arbeit sehr groß gewählt war, wodurch weniger absolute Antworten möglich waren, als es bei einer noch stärkeren Fokussierung der Fall gewesen wäre.

## 4 Ergebnisse: Die Einschätzungen der Expert:innen

### 4.1 Die technische Entwicklung audiovisueller Manipulationen

Wie bereits in Kapitel 2.4 beschrieben, sorgt gerade **die technologische Entwicklung synthetischer Medien** für ein sehr dynamisches Forschungsfeld. Für den IT-Forensiker Dr. Sascha Zmudzinski (vgl. Int. 126ff.) bergen ML-basierte Verfahren in der Manipulationslandschaft auch das größte Potenzial. Eine Beschreibung der derzeitigen technologischen Manipulationsmöglichkeiten ist daher nur eine Momentaufnahme. Sie erlaubt jedoch nicht nur, den Forschungsstand zu ergänzen, sondern auch Schlussfolgerungen zu der Geschwindigkeit zu ziehen, mit der sich die technologischen Möglichkeiten weiterentwickeln. Einen tatsächlichen Blick in die Zukunft über die Einschätzungen der Expert:innen liefern dagegen die Kapitel 4.4 bis 4.6.

Der Deepfake-Experte des BSI Dr. Dominique Dresen beschreibt, dass synthetische Medien die Manipulationslandschaft besonders deshalb stark verändern würden, weil sie Manipulationen in einem Bereich ermöglichen, bei dem klassische Methoden an ihre Grenzen stoßen, dem menschlichen Gesicht. Als die ersten Deepfake-Beispiele 2017 in der Öffentlichkeit kursierten, seien sie noch recht einfach zu erkennen gewesen, jetzt würden sie zunehmend authentischer. Gerade seit ein, zwei Jahren werde es schwierig, sie noch bei einfacher Betrachtung als solche zu erkennen. Mit genügend personellen und technischen Ressourcen seien synthetische Medien daher besonders gut geeignet, gewollte Botschaften zu verbreiten und die potenziell perfekteste Manipulation (vgl. Int. Dresen 28ff., 162f., 392ff., 700ff., 742ff.; Jahnel 102ff.).

Für ein sehr gutes Ergebnis brauche es gut passendes Trainingsmaterial und derzeit noch Zeit für eine Nachbearbeitung, in der die Übergänge des synthetisch generierten Gesichts dem Rest des Inhalts angeglichen werden können. Wenn man auch das Ausgangsmaterial selbst produzieren könne, in welches die Manipulation eingesetzt wird, ermögliche einem das zudem noch, bestehende Problemsituationen (Kapitel 2.3.2) zu umgehen. Darunter fallen für Dresen (vgl. Int. 219ff.) vor allem starke Close-Ups und besonders Perspektiven, welche die betroffene Person im Profil zeigen.

Dr. Jutta Jahnel vom KIT hebt hervor, dass es Manipulationen schon immer gegeben habe. Deepfake-Technologien hätten die Situation in erster Linie verstärkt und dadurch mit einer höheren Relevanz versehen, indem sie nicht nur authentischere Manipulationen ermöglichen, sondern vor allem deren Erstellung beschleunigen, vereinfachen und kostengünstiger machen. Auch Dresen beschreibt, dass man bei der Produktion synthetischer Medien auch ohne Ressourcen einer professionellen Filmproduktion inzwischen sehr weit komme, insbesondere, wenn man keine höhere Auflösung als etwa 720p anstrebt, die gewünschte Mimik keine starken Grimassen aufweist und die Mimik-vorgebende Person der Zielperson in ihren äußerlichen Merkmalen ähnelt. Gerade bei Politiker:innen und anderen Prominenten sei dafür in der Regel immer genug audiovisuelles Material vorhanden. Für ein gutes Ergebnis brauche eine App wie *DeepFaceLive* aktuell etwa einen Tag Trainingszeit, um die Gesichter der Ausgangs- und der Zielperson zu lernen. Es gebe auch Modelle mit etwas schlechteren Resultaten, die bei einer vorgegebenen Zielperson ohne Trainingszeit auskommen oder anhand von ein, zwei Samples ein relativ passables Ergebnis einer selbst gewählten Zielperson erzeugen (vgl. Int. Jahnel 85ff.; Int. Dresen 261ff., 416ff., 732f.).

Eine äußere Ähnlichkeit zwischen Ausgangs- und Zielperson ist u. a. deshalb relevant, weil in den für Laien angebotenen Programmen wie *DeepFaceLive* bisher nur das Gesicht ausgetauscht werde. Grundsätzlich sei es aber genauso möglich, bspw. Ohren ebenfalls auszutauschen. Deshalb dürfen passende Ohren allein nicht als Beweis für die Authentizität einer Aufnahme gesehen werden. In Zukunft würden Ohren vermutlich mit manipuliert, wenn man jetzt wisse, dass darauf geachtet wurde (vgl. Int. Dresen 274ff., 735f.). Schlussendlich sei es eine Ressourcenfrage. *DeepFaceLive* sei mit Qualitätsabstrichen auch bereits echtzeitfähig und erzeuge lediglich etwa 100 ms Delay für die Verarbeitungsberechnung. Wenn die Auflösung nicht zu hoch sein muss, schaffe man so bereits mit einer normalen Gamer-Grafikkarte Echtzeitmaterial mit 10-20 fps. Schwieriger sei die Audiospur, weil es zunächst ein paar zusammenhängende Töne brauche, bevor mit der Manipulation begonnen werden kann. Dadurch habe man auf der Audiospur noch eine Verzögerung von etwa einer Sekunde. Ein Programm, das Bild- und Tonmanipulation in einem ermöglicht, ist Dresen (vgl. Int. 296ff.) bisher nicht bekannt. Er sieht es jedoch nicht als technische Hürde an.

Für andere Bereiche oder ganze Körper hingegen brauche „man vielleicht gar nicht diese Deepfake-Methoden. Da muss man auch mal überlegen, ob man mit den klassischen Methoden nicht vielleicht schon sehr weit kommt und dann nur noch das Gesicht draufsetzt“ (Int. Dresen 388ff., vgl. 341ff.). In bisherigen Videobeispielen, die sich der synthetischen Erzeugung ganzer Körper annähern, sei die Qualität vergleichsweise schlecht. Zum einen, weil es aufwendig sei, den Detailreichtum eines Gesichts auf einen ganzen Körper zu übertragen. Zum anderen könnten sich Körperhaltungen stärker als Gesichter unterscheiden, wodurch man noch mehr Trainingsmaterial brauche. Möglich sei die Synthetisierung ganzer Körper in Zukunft theoretisch, aber vielleicht gar nicht die einfachste Variante. Sascha Zmudzinski sieht die synthetische Generierung ganzer Körper dagegen als logischen nächsten Schritt:

„Also das wird kommen. [...] Mit immer höheren Auflösungen [...] ist natürlich auch zu erwarten, [...] dass es dann nicht nur um diesen Bereich geht, sondern eben den ganzen Kopf oder den ganzen Oberkörper, so wie es jetzt in diesem Videotelefonat hier auch ist [...] Wie lange das dann dauert, bis es weit verfügbar ist... Die Entwicklungen dort sind so schnell. Ich würde jetzt nicht sagen Monate oder so, aber vielleicht haben wir da nächstes Jahr schon was viel Besseres als dieses Jahr. Vielleicht dauert es auch zwei Jahre, vielleicht drei, aber das ist da auch zu erwarten.“ (Int. Zmudzinski 202ff.)

Als „zusätzliche Bedrohung“ und „ganz starke Herausforderung“ (Int. Zmudzinski 136ff.) oder auch „massive Änderung“ und „wirklich umwälzend“ (Int. Riess 196f. 205) sehen Zmudzinski und Dr. Christian Riess die inzwischen ausgereifte und verbreitete Möglichkeit (vgl. Zmudzinski 136ff.), Bilder vollständig synthetisch anhand einer Textbeschreibung zu erzeugen (**Text-to-Image**):

„Wir haben ja diese klassische Vorstellung: Ein Bild ist eine physikalische Messung im Prinzip von Photonen, die in Richtung meiner Kamera unterwegs waren. Und das kann man sich abschminken. Das gibt es nicht mehr.“ (Int. Riess 202ff.)

Auch wenn die T2I-Generatoren derzeit teilweise noch Schwachsinn generieren würden, schätzt Dresen (vgl. Int. 1004ff.), dass sie in den nächsten Jahren sehr gut werden. Im Moment sei besonders die Darstellung von Händen für die Generatoren noch schwierig, ebenso wie in der traditionellen Malerei. Riess (Int. 403ff.) stimmt ihm zu: „Diese offensichtlichen visuellen Fehler, die werden sicher irgendwann zu Ende sein.“ Gleichzeitig müsse man sehen, dass die Szenarien der generierten Bilder bisher noch



stark eingeschränkt seien und damit auch der Raum für semantische Fehler. Das sei bei komplexeren Szenen anders.

Zusätzlich zu T2I-Generatoren sieht Riess auf der Bildebene eine weitere Herausforderung darin, dass nächstes Jahr JPEG-AI standardisiert werde, ein ML-basiertes Kompressionsverfahren, das bei seiner Anwendung „locker einen Faktor von zehn in der Kompressionseffizienz“ (Int. Riess 246f.) bringe. Mit dem Preis, dass Bildmaterial jeglichen Ursprungs in erster Instanz synthetisiert werde. Womit es nicht oder nur sehr indirekt auf das ursprüngliche Bild zurückzuführen sei:

„Das heißt im Prinzip, ein neuronales Netz schaut dieses Bild an und überlegt sich dann eine kurze Beschreibung dafür [...] aus der dann sozusagen ein dazu passendes neuronales Netz das Bild rekonstruieren kann. [...] Und das hat Auswirkungen, die, denke ich, wichtig sind [...] Was ist eigentlich ein echtes Bild? Was ist kein echtes Bild? Gibt es einen Unterschied zwischen echten und nicht-echten Bildern?' Also das hat, glaube ich, sehr starken Einfluss“ (Int. Riess 220ff.).

Besonders deshalb habe es Einfluss, weil die Multimedia-Forensik und ihre statistischen Modelle von dem Vorwissen und den Rahmenbedingungen leben, dass Bildern eine physikalische Messung zugrunde liegt (vgl. Int. Riess 205ff.).

Um hingegen authentische Videoinhalte komplett synthetisch generieren zu können (**Text-to-Video**), seien deutlich mehr Trainingsdaten und Rechenleistung nötig:

„Diese zeitliche Abfolge muss ja irgendwie gelernt werden. Man braucht nicht nur 'Ich erzeuge jetzt ein Bild, sondern ich erzeuge viele Bilder, die zeitlich irgendwie zusammenhängend oder kohärent so stimmig sind, wo halt das Haar sich korrekt bewegt, die ganze Mimik' ...Also ich denke, das ist schon ein sehr viel komplizierteres Problem“ (Int. Dresen 612ff.)

Auch Zmudsinski (Int. 247ff.) ist der Meinung, dass es dafür in erster Linie genügend Rechenleistung brauche: „Letztlich ist es dann auch eine finanzielle Frage, die entsprechenden Modelle so zu trainieren, dass es auch im Videobereich gut funktioniert. Wie lange sowas dauert, kann ich schlecht einschätzen, aber... Also letztlich ist es dann so auch zu erwarten, klar.“ Dominique Dresen (vgl. Int. 610ff.) schätzt, dass es frühestens in fünf bis zehn Jahren so weit ist.

Rechtlich gesehen sei die Verbreitung synthetischer Medien bisher nur dann rechtswidrig, wenn es sich um strafrechtlich relevante Inhalte wie Hassrede oder pornografische Inhalte handelt: „Alle anderen, wie zum Beispiel Desinformation, sind im Prinzip

nicht staatlich reguliert“ (Int. Jahnel 234f.). Allerdings gebe es eine Co-Regulierung in Form des *Digital Service Act*, der Social-Media-Plattformen in die Verantwortung nehme, legale Formen synthetischer Inhalte zu kennzeichnen und rechtswidrige zu löschen – sofern sie darauf aufmerksam gemacht werden. Zusätzlich müssten die ganz großen Plattformen evaluieren, ob es systemische Risiken in Form betroffener demokratischer Werte gibt. Mit der bisher angedachten Form einer KI-Verordnung auf EU-Ebene würden Deepfakes in Zukunft zwar als geringes Risiko eingestuft, müssten aber zumindest von der erstellenden Person gekennzeichnet werden (vgl. Int. Jahnel 212ff.).<sup>28</sup>

Dresen (vgl. Int. 705ff.) weist insgesamt darauf hin, dass es für die Täuschung besonders bei Journalist:innen in der Regel nicht reiche, ein Video zu fälschen, sondern zusätzliche Ressourcen nötig seien, um das manipulierte Material mit einem entsprechenden Kontext zu einer **glaubwürdigen Kampagne** zusammenzustellen.

Auch Jutta Jahnel hält eine für sich genommene Betrachtung der einzelnen Manipulationsformen aus einem anderen Grund nicht für sinnvoll, obwohl es von Methode zu Methode andere Schwächen gebe: „Letztendlich ist nämlich eh das Problem, dass man hinten die Wirkung und die Schäden reguliert und eigentlich gar nicht die Technologie, die verfließen alle“ (Int. Jahnel 198ff., vgl. 469). In Zukunft werde es zudem immer schwieriger, Manipulationsformen einzeln zu betrachten:

„Man kann ja hingehen und Texte mit Chat-GPT produzieren, aus denen dann mittels Text-to-Speech quasi die Sprache machen und mit Lip-Sync dann wieder ein Video. Man kann diese ganzen Technologien irgendwie verschmelzen [...] In Zukunft wird es dann immer schwieriger sogar Deepfakes alleine zu betrachten. Das wird irgendwie nach allen Seiten größer werden, schwammiger. Das wird ein Riesenproblem sein.“ (Int. Jahnel 479ff.)

---

<sup>28</sup>Ein entsprechender Gesetzestext ist am 14.06.23 durch das EU-Parlament verabschiedet worden, muss nun aber mit der EU-Kommission und den EU-Mitgliedsstaaten weiter verhandelt werden, bevor er in Kraft treten kann.

## 4.2 Möglichkeiten und Grenzen der Multimedia-Forensik

Bereits in Kapitel 2.1 wurde angesprochen, dass es sich bei der digitalen Multimedia-Forensik um ein recht junges Forschungsgebiet handelt. Das zeigt sich auch dadurch, dass es bisher keine spezifischen Professuren für Multimedia-Forensik gebe und auch keine saubere bzw. einheitliche Definition (vgl. Int. Riess 52ff, 77f.).

### 4.2.1 *Definition und Systematisierung multimedia-forensischer Bereiche*

Gerade semantische Verfahren wie die inverse Bildersuche werden in der IT-Forensik bei der Beschreibung multimedia-forensischer Methoden häufig vernachlässigt. Laut Riess (vgl. Int. 132ff.) liegt das vor allem daran, dass sich semantische Verfahren nur schwierig auf ihre Wirksamkeit testen lassen. Trotzdem handele es sich bei der inversen Bildersuche um eine forensische Methode, bei der es im Kern darum gehe, „dass man intern eine möglichst kompakte Repräsentation von dem Bild hat, so dass man dann schnell [...] und effizient abgleichen kann, ob es so ein ähnliches Bild schon mal gibt“ (Int. Riess 93ff.). Salopp könne man das Verfahren der inversen Bildersuche auch mit einer besonders aggressiven Bildkompression gleichsetzen. Die inverse Bildersuche sei daher ein technischer Vorgang (vgl. Int. Riess 96ff.). Die Zuordnung der inversen Bildersuche als Teil der Multimedia-Forensik wird auch von Zmudzinski unterstützt. Weitergehende Recherchen, die den Kontext audiovisueller Medien miteinbeziehen, sieht er hingegen nicht als Bestandteil im engeren Sinne, da es in der Multimedia-Forensik darum gehe, Inhalte ohne umfangreiches Kontextwissen in erster Linie aufgrund des vorliegenden Datenmaterials technisch auszuwerten (vgl. Int. Zmudzinski 23ff.). Zu bedenken ist bei dieser Differenzierung allerdings, dass eine inverse Bildersuche auf dem Abgleich mit bereits veröffentlichten Medien beruht, die in gewisser Weise auch als Kontext gesehen werden können.

Die folgende Tabelle gibt einen systematischen Überblick über die relevanten Ansätze, audiovisuelles Material zu verifizieren. Der Überblick beruht sowohl auf der im Theorierteil verwendeten Literatur, als auch auf den Sichtweisen von Christian Riess (vgl. Int. 83ff., 391ff., 465ff.) und Sascha Zmudzinski (vgl. Int. 101ff., 224ff.), die sich nicht widersprechen, aber durchaus ergänzen.

| Verfahren, die ohne spezifische Software durchführbar sind                                  | Verfahren, die auf spezifische Software angewiesen sind   |  |  |                        |
|---|---|--|--|------------------------|
| <b>Klassische Recherche</b><br>Analyse des Publikationskontextes<br>Urheber:in kontaktieren | <b>Passive Multimedia-Forensik</b>  |  |  | <b>Aktive Forensik</b> |
|   | <b>Semantische Verfahren (u. a. physikbasierte Verfahren)</b>   | <b>Formatbasierte Verfahren</b>  | <b>Pixelbasierte Verfahren</b>   |                        |
|   | Inverse Bildersuche<br>Analyse vorhandener Logos und anderer Quellenangaben<br>Geolokalisierung + anschl. Chronolokalisierung durch die Analyse + den Abgleich von <ul style="list-style-type: none"> <li>• Schriftinhalten / Schildern</li> <li>• Sprache / Dialekt</li> <li>• Objekten / landschaftlichen Merkmalen</li> <li>• jahreszeitlichen Änderungen / Witterung</li> <li>• Sonnenstand / Schattenwurf</li> </ul> Analyse auf physikalische Unstimmigkeiten (Perspektive + Lichtsituation)<br>Analyse des Verhaltens vorkommender Personen + physiologischer Merkmale<br>Einzelbildanalyse zur Detektion hinzugefügter oder entfernter Frames (nur bei Video) | Analyse der Metadaten<br>Analyse der Dateistruktur<br>Analyse + Abgleich von ENF-Signalen, also örtlichen Schwankungen in der Netzfrequenz (hauptsächlich bei Audio) | Analyse + Abgleich des Gerätemodells bzw. des spezifischen Geräts, von dem die Aufnahme stammt, durch die Identifizierung von Artefakten und Sensorrauschen<br>Analyse der Kompressionsmuster und daraus die Anzahl erfolgter Kompressionsvorgänge<br>Suche nach identischen Stellen innerhalb einer Aufnahme<br>Analyse lokaler Inkonsistenzen auf Pixelebene durch (mehrfache) Komprimierung im Bildrauschen, der Bildschärfe oder Spuren des Farbfilter-Arrays des Gerätes<br>Analyse auf GAN-spezifische oder andere vorgegebene Spuren, die auf eine synthetische Erstellung hindeuten<br>ML-basierte Analyse ohne Vorgabe definierter Merkmale |                        |

Tabelle 3: Überblick über Verifizierungsverfahren (Grafik: Martha Peters)

Am Ende von Kapitel 4.1 wurde bereits die zunehmende Schwierigkeit angesprochen, einzelne Manipulationsformen voneinander abzugrenzen. Anknüpfend daran sprechen gleich drei Expert:innen zum Ende ihres Interviews an, dass es wichtig sei, die Authentifizierung von Medien über verschiedene Medientypen hinweg **multimodal** zu denken. Begonnen damit, bei einer Zeitungsmeldung Text und Bild zu betrachten, hin dazu, bei Deepfakes sowohl den Faktor Gesicht als auch den Faktor Stimme zu betrachten. Genauso könne es helfen, durch die Verwendung gleicher Manipulationen oder eines ähnlichen Schreibstils auf Zusammenhänge zwischen Social-Media-Accounts und deren Identität zu schließen (vgl. Int. Dresen 1099ff.; Jahnel 476ff.; Zmudzinski 766ff.).

#### 4.2.2 Die ML-basierte Detektion von Deepfakes

Wie bereits in der obenstehenden Tabelle dargestellt, werden nicht nur zur Erstellung von Manipulationen, sondern auch zu ihrer Detektion ML-basierte Verfahren eingesetzt – in erster Linie mit dem Ziel, synthetische Medien von authentischen Inhalten zu unterscheiden. Das passiert bisher noch mit bescheidenem Erfolg (siehe Kapitel 2.5), wie auch die Expert:innen bestätigen. Dresen bezeichnet die aktuelle Leistung automatisierter Systeme als „nah am Münzwurf“ (Int. Dresen 459f.; vgl. 444ff.; Jahnel 27ff.).

Gründe dafür seien vor allem unzureichendes Trainingsmaterial, welches dem Detektionssystem Unterscheidungsmerkmale nahelege, die gar nichts mit der Unterscheidung zwischen authentischem und synthetischem Material zu tun haben oder aber zu spezifisch sei, als dass der Detektor über die trainierten Inhalte hinaus einsetzbar wäre.

GAN-spezifische Spuren bspw. seien nicht in allen Deepfakes zu finden, die bei der Generierung entstehenden Artefakte durchaus unterschiedlich. Entsprechend müssten auch die Detektionsmaßnahmen spezifisch für die verschiedenen Varianten synthetischer Medien sein, um richtig erfolgreich zu sein. Eine mögliche Lösung sei, viele verschiedene Detektoren parallel laufen zu lassen. Allerdings seien auch, wie bereits im Theorieteil beschrieben, die Kompressionsverluste besonders durch Social-Media-Uploadsysteme ein Problem. Zudem würden bisher bekannte (Audio-)Detektoren nur Wahrscheinlichkeitswerte ausgeben (vgl. Int. Dresen 466ff., 530ff., 676ff., 826ff.; Jahnel 28ff.) und das nicht zuverlässig.<sup>29</sup> Sascha Zmudzinski (vgl. Int. 152ff., 224ff.) hingegen ist optimistischer und erläutert, dass das Fraunhofer SIT bereits maschinelles Lernen als Werkzeug nutze, um Bildmanipulationen zu erkennen – aber auch als Untersuchungsobjekt, um die Resilienz ML-basierter Verfahren zu prüfen. Beide – manuelle und ML-basierte Vorgehensweisen – seien sinnvoll und gerade das Thema KI und ML werde jetzt auch in der Forensik sehr wichtig. Auch er sieht den entscheidenden Faktor im richtigen Trainingsdatensatz, aber teilweise auch in Vorgaben, welche Merkmale sich der Detektor überhaupt anschauen soll.

---

<sup>29</sup> Ein Beispiel für einen solchen Detektor synthetischer Audioinhalte vom Fraunhofer AISEC findet sich unter [deepfake-total.com](https://deepfake-total.com), zuletzt geprüft am 22.06.23.

### 4.2.3 Möglichkeiten und Hürden aktiver Forensik

Wie bereits am Ende von Kapitel 2.4 beschrieben, besteht eine weitere Möglichkeit zur Sicherstellung der Integrität audiovisueller Inhalte darin, zum Zeitpunkt der Aufnahme Informationen zu erstellen, die es später ermöglichen zu kontrollieren, dass die Datei seit diesem Moment nicht verändert worden ist. Dazu gibt es verschiedene Ansätze, von denen bisher keiner großflächig eingesetzt wird. Auf den folgenden Seiten werden die Antworten der Expert:innen zu den verschiedenen Umsetzungsmöglichkeiten und möglichen Problemen zusammengefasst.

Eine erste Möglichkeit besteht darin, direkt in die Aufnahme ein digitales Wasserzeichen zu integrieren, was tatsächlich das Bild- oder Tonmaterial verändert und je nach Robustheit sogar bei einem Sprung in die analoge Welt erhalten bleibe, bspw. wenn man den ursprünglichen Inhalt vom Bildschirm abfilmt. Je robuster ein Wasserzeichen sei, umso eher spiele allerdings auch der folgende Qualitätsverlust eine Rolle. Eine weitere Variante ist, direkt im Aufnahmegerät einen Signaturschlüssel zu erstellen und in die Aufnahme-Datei hineinzuschreiben. Auch hier seien frühere Ansätze, genauso wie die Forschung zu digitalen Wasserzeichen, bereits in den frühen 2000er Jahren zu sehen gewesen, hätten sich aber nicht durchgesetzt (vgl. Int. Zmudzinski 259ff., 286ff., 339ff.). Digitale Signaturen würden durchaus weiter von der Wirtschaft angestrebt. Besonders hervorgehoben wurde sowohl von Dresen (vgl. Int. 887ff.) als auch von Jahnel (vgl. Int. 370ff.) in dem Zug die *Content Authenticity Initiative* von Adobe.

Von einer Umsetzung über Blockchain-Technologie werde gerade durch Informatiker:innen immer wieder abgeraten, sagt Jahnel (vgl. Int. 372ff.). Das bestätigt sich durch die Aussage Zmudzinskis (vgl. Int. 390ff.), der schlicht keine Notwendigkeit für den Einsatz der Technologie sieht: Mit einer kryptografisch sicheren Signatur, einem Hash, der auf abgesicherten Webseiten öffentlich zur Verfügung gestellt wird, erlange man bereits die Möglichkeit, das einem selbst vorliegende Material zu verifizieren, indem man die daraus entstehende Signatur mit der öffentlichen Signatur abgleicht. Weder sei vom Hash ein Rückschluss auf das originale Bild möglich, noch sei mit heutigen Methoden die Erstellung eines manipulierten Bildes mit demselben Hashwert möglich – eine Blockchain damit nicht nötig. Dresen (vgl. Int. 902ff.) gibt zudem zu be-

denken, dass es sich wenn überhaupt um eine öffentliche Blockchain handeln müsse, um sicher zu sein und für jede Aktion im Rahmen der Blockchain dadurch Kosten anfallen würden: „Das heißt, man müsste dann für jedes Bild und für jedes Video ein paar Cent in die Hand nehmen. Das würde sich dann schnell läppern“ (ebd. 946ff.). Zudem sei, genau wie bei digitalen Wasserzeichen oder Signaturen, damit nicht automatisch auch die Vertrauenswürdigkeit der Quelle sichergestellt.

Dass man weiterhin dem Ursprung des Wasserzeichens oder der Signatur vertrauen muss, sehen auch Zmudzinski (vgl. Int. 264ff.) und Riess als Stolperstein. Für Riess (vgl. Int. 495ff.) ist es deshalb wichtig, ein solches Verfahren direkt im Chip des Aufnahmeapparates zu verankern. Eine weitere Herausforderung sei, die Verifikationsmöglichkeit bei legitimen Änderungen zu erhalten, dass die digitale Signatur sich also bspw. gegenüber Kompressionsschritten robust zeigt oder alle Bearbeitungsschritte ebenfalls manipulationssicher vermerkt werden (vgl. Int. Dresen 893ff.; Int. Zmudzinski 284ff.).<sup>30</sup> Bei digitalen Wasserzeichen sieht Zmudzinski (Int. 300ff., vgl. 365ff.) dafür gerade bei nutzergenerierten Inhalten wenig Chancen: „Vielleicht im Profibereich, Broadcastbereich oder so kann ich mir das vorstellen. Bei digital umcodierten Telegram-Videos oder so wäre ich auch nach heutigem Stand skeptisch, wenn es um echte Integritätswasserzeichen geht. Eine Quellenkennung oder so, das kann ich mir noch vorstellen.“

Abseits der technischen Hürden nennen die Expert:innen weitere Herausforderungen: So müsse man aufpassen, dass ein solches Verfahren auch mit Datenschutz und Persönlichkeitsrechten zu vereinbaren ist – und nicht in jedem Fall sei es von Vorteil, wenn die Quelle von Inhalten für alle öffentlich sichtbar ist. Zudem müsse man überlegen, ob ein solches Verfahren auch für alle zugänglich sei: „Es kann auch ein Problem sein, dass solche komplexen Methoden für Länder, für Entwicklungsländer oder vielleicht auch Barrieren für einfache Leute sind oder auch für Journalisten, die gar nicht so hoch aufgerüstet sind“ (Int. Janel 341ff., vgl. 378ff.).

Und selbst bei einem Verfahren, das für alle zugänglich ist, sei man auf die Bereitschaft angewiesen, es einzusetzen: „Das Problem sind wirklich die Nutzer im Moment [...] Es funktioniert leider nicht so wirklich. Ja, die Leute nutzen es einfach nicht“ (Int. Wild

---

<sup>30</sup> Die *Content Authenticity Initiative* scheint genau das zu ermöglichen.

255; vgl. Int. Zmudzinski 266ff.). Darüber hinaus erfülle gerade das Verhältnis zwischen bekannten Akteuren wie erprobten Fotograf:innen und Redaktionen innerhalb fester Strukturen häufig schon eine ähnliche Funktion, wie man sie durch Signaturen auf technischer Ebene anstrebt: „Ich habe ja oft auf Verfahrensprotokollebene in den Verlagen schon so eine Art Trust-Chain implementiert“ (Int. Riess 490ff.).

### **4.3 Die Praxis**

#### *4.3.1 Vorkommende Manipulationsformen*

Auch wenn synthetische Medien besonders überzeugende Ergebnisse liefern können, spielen sie laut den Interviewten zur tatsächlichen Meinungsmanipulation bisher eine untergeordnete Rolle. Zwar sei der Einsatz synthetischer Medien bereits vor der letzten US-Wahl erwartet und davor gewarnt worden, dann aber gar nicht zu beobachten gewesen (vgl. Int. Dresen 686ff, 702ff., 762ff.). Insgesamt gebe es bisher nur wenige Beispiele, wo synthetische Medien überhaupt Auswirkungen gehabt hätten und nicht mit einer kurzen journalistischen Recherche zu erkennen gewesen seien (vgl. Int. Spangenberg 528ff.). Erste ernsthafte Einsätze auf gesellschaftlicher Ebene seien inzwischen im Rahmen des Krieges in der Ukraine zu beobachten, wo Deepfakes von Wolodymyr Selenskyj und Wladimir Putin zu kriegstaktischen und Propagandazwecken verbreitet worden sind (vgl. Int. Jahnel 56ff.).

Sowohl die Bellingcat-Mitarbeiterin Johanna Wild als auch Andrea Sauerbier vom SPIEGEL sind der Meinung, dass Deepfakes in ihrem Redaktionsalltag bisher kein Problem darstellen:

„Es gibt ja eine riesengroße Diskussion über Deepfakes im Moment und das sehen wir noch sehr, sehr viel weniger. Also das spielt bis jetzt in unserer Recherche noch keine allzu große Rolle.“ (Int. Wild 15ff.)

Auf einen Deepfake sei Johanna Wild (vgl. Int. 19ff.) entsprechend in keiner ihrer bisherigen Recherchen gestoßen. Sie beschäftige sich aber auch kaum mit Videos öffentlicher Personen wie Politiker:innen, sondern vielmehr mit nutzergenerierten Inhalten.



Für die Printausgabe des SPIEGELs seien Manipulationen insgesamt kaum ein Thema. An vorsätzliche Manipulationen könne Andrea Sauerbier sich in ihrem Arbeitsleben kaum erinnern:

„Das spielt in unserem Arbeitsalltag ganz ehrlich gesagt - und ich mache das jetzt ja auch schon lange - so gut wie keine Rolle. Das ist tatsächlich so. Also die Fehler, die bei uns auftreten, die haben ganz andere Ursachen. Aber mit Manipulationen haben wir so gut wie gar nichts zu tun und deswegen spielt so dieser forensische Bereich auch bei uns so gut wie gar keine Rolle. Fehler passieren tatsächlich eher von menschlicher Seite durch Unachtsamkeit.“ (Int. Sauerbier 219ff., vgl. 248ff.)

Bellingcat hingegen greift viel auf Onlineinhalte zurück und ist entsprechend stark mit audiovisuellen Manipulationen konfrontiert. Am häufigsten stoße man auf Material, das nicht unbedingt bearbeitet sei, sondern schlicht dekontextualisiert verwendet werde. Ab und zu habe man auch mit bearbeiteten Inhalten zu tun, die aber oft sehr plump manipuliert und nicht schwierig zu erkennen seien. Entsprechend bewertet Johanna Wild (vgl. Int. 6ff.) auch die Herausforderungen, mit denen Bellingcat durch audiovisuelle Manipulationen konfrontiert ist:

„Also ich denke wirklich, falscher Kontext ist die größte Herausforderung insgesamt, viel mehr als Manipulationen bis jetzt. [...] Was man ja viel in den Medien hört, dass es ganz schwierig ist, gefälschte Bilder und Videos zu erkennen, das ist gar nicht so sehr unser Problem. Es ist tatsächlich dieser Kontext.“ (Int. Wild 30ff.)

Das deckt sich mit den bereits im Theorieteil aufgegriffenen Aussagen journalistischer Akteur:innen und den Einschätzungen der anderen Interviewpartner:innen (vgl. Int. Dresen 762ff.). Aus aktuellem Anlass seien beispielsweise viele Bilder im Umlauf, die vorgeben aus der Ukraine zu kommen, tatsächlich aber bereits älter sind und aus anderen Kriegen z. B. in Syrien stammen (vgl. Int. Spangenberg 309ff.).

Unterstützt werden diese Erfahrungswerte durch ein aktuelles Projekt des Fraunhofer SIT mit dem Namen *DisCo – Disinformation in Conflicts*, das die Erstellung und Verbreitung von Desinformation im Rahmen aktueller Konflikte, insbesondere der Covid-19-Pandemie, zum Thema hat. Im Rahmen des Projekts wurden die Formen visueller Manipulationen und deren Häufigkeit auf bekannten Internetplattformen betrachtet:

„Da haben wir so vier verschiedene Dinge gesehen - oder vier verschiedene Arten mit Bildern Desinformation zu betreiben. Das Erste ist, wo man vielleicht so als erstes dran denken würde bei Bildmanipulationen, dass tatsächlich die Integrität des Bildes verletzt wurde. Also dass innerhalb des Bildes irgendwie Pixelinforma-

tionen verändert wurden. Die zweite Art ist, ein Bild aus dem Zusammenhang zu reißen. Also ein Bild aus einer früheren Nachrichtenmeldung zu verwenden, um jetzt im Kontext von Corona etwas scheinbar zu belegen. Die dritte Art ist, dass man das Bild, die Aufnahmesituation nachstellt. Also dass man ein Foto so gesehen komplett inszeniert. Und das vierte, gar nicht mal so unwichtig, ist, dass das Bild selbst erstmal gar keine Falschaussage enthält, sondern dass es zu Illustrationszwecken benutzt wird, um so gesehen die Falschnachricht, die dann irgendwie im Text steht [...] optisch anzureichern“ (Int. Zmudzinski 584ff.)

Die Beobachtungen im Rahmen des Projekts decken sich mit den bereits in Kapitel 2.3 vorgestellten Manipulationsformen, vernachlässigen aber die Kategorie synthetischer Medien, welche keine Rolle gespielt hätten (vgl. ebd. 643f.). Das unterstreicht die Aussagen zum bislang geringen Einsatz synthetischer Medien zu Manipulationszwecken. Stattdessen weist Zmudzinski (vgl. ebd. 593ff.) auf eine weitere Manipulationsmöglichkeit in Form beigefügter Symbolfotos hin: Das Manipulationspotenzial von Symbolfotos gegenüber Laien sei nicht zu unterschätzen,<sup>31</sup> wird hier jedoch vernachlässigt, da es um die Auswirkungen audiovisueller Manipulationen auf Redaktionen geht. Genauso wird die von Zmudzinski (vgl. ebd. 626ff.) angesprochene Verbreitung schriftlicher Falschaussagen in Form von Bilddateien hier vernachlässigt, da die Inhalte weder die klassische Funktion audiovisueller Inhalte erfüllen, noch als solche zu verifizieren sind. Insgesamt spiegeln die Ergebnisse des Projekts größtenteils die bereits ausgeführten Eindrücke der anderen Interviewten wieder, legen aber auch eine höhere Relevanz inszenierter Inhalte nahe als Literatur und die redaktionellen Interviewten bisher vermuten ließen:

„Da waren tatsächlich die aus dem Kontext gerissenen Bilder und gestellte Aufnahmen [...] sehr wichtig. Die Bildmanipulationen, also die Manipulationen im Bild, Integritätsverletzungen im Bild, spielten fast keine Rolle.“  
(Int. Zmudzinski 657ff.)

Sascha Zmudzinski (vgl. ebd. 701ff.) und Christian Riess (vgl. Int. 327ff.) ergänzen, dass gerade bei wissenschaftlichen Ergebnissen Bilder häufig eine zentrale Beweisfunktion einnehmen würden. Unter dem Stichwort ‚Academic Integrity‘ sei die Überprüfung von Bildmaterial daher schon seit mehr als zehn Jahren auf dem Radar wissenschaftlicher Verlage und würde durch Faktenprüfer:innen umgesetzt.

---

<sup>31</sup> Siehe dazu bspw. Peters, Martha (2020): Symbolfotos im Technik- und Wissenschaftsjournalismus. Intention und Wirkung.

Schlussendlich sei die Wahl der Manipulationsform eine Kosten-Nutzen-Frage und offensichtlich reiche die Leistung von Shallowfakes in Form von einfachen Anpassungen einer Tonspur oder der Nutzung alter Inhalte in einem neuen Kontext aus, um Meinungen zu manipulieren. Synthetische Medien seien (noch) einfach mehr Aufwand (vgl. Int. Dresen 686ff., 762ff.). Sie seien bisher vor allem in Form pornografischer Inhalte in Umlauf, womit sie eher ein individuelles Problem als von gesellschaftlicher Relevanz sind (vgl. Int. Jahnel 14ff.).

#### *4.3.2 Praxis und Probleme in Redaktionen und Recherchenetzwerken*

Natürlich spiegeln sich die Erfahrungen mit audiovisuellen Manipulationen im Umgang mit ihnen wieder. Sascha Zmudzinski (vgl. Int. 426ff.) sieht das Thema Echtheitsprüfung in allen großen Medienhäusern und Rundfunkanstalten auf dem Radar. Das werde sowohl durch die Durchführung von Schulungen wie auch Hintergrundgespräche deutlich. Allerdings sei der Umgang mit dem Thema sehr unterschiedlich, meint Jochen Spangenberg:

„Man muss hier von Redaktion zu Redaktion differenzieren. Man muss womöglich auch von Land zu Land unterscheiden. Weil es ist sehr, sehr, sehr unterschiedlich, also wirklich sehr, sehr, sehr unterschiedlich, wie Redaktionen da aufgestellt sind, wie sie das machen: Haben sie eine eigene Factchecking-Abteilung oder ist das Teil des ganz normalen journalistischen Alltags? Wie ist das integriert oder nicht? Wie ist die Expertise? Meiner Meinung nach, das ist meine persönliche Meinung, gehört das zum Handwerkszeug eines jeden, einer jeden Journalistin, aber es ist halt noch nicht im Repertoire einer jeden und eines jeden Journalisten, dass man halt relativ sicher mit diesen digitalen Tools arbeitet. Manche brauchen das ja auch häufig nicht. Aber nichtsdestotrotz denke ich, jeder/jede sollte wirklich so zumindest Basiskenntnisse haben im Umgang mit der Verifizierung oder Überprüfung von digitalem Material“ (Int. Spangenberg 361ff.)

In Deutschland habe man, im Vergleich zur britischen BBC, erst spät begonnen, die Verifizierung nutzergenerierter Inhalte als integren Bestandteil des Nachrichtengeschäfts zu sehen (vgl. Int. Spangenberg 375ff.).

Der Überblick über die Verifizierungspraxis in Redaktionen aus Kapitel 2.5 wird auf den folgenden Seiten besonders durch einen detaillierten Blick auf die Arbeitsweise der Dokumentationsabteilung des deutschen Magazins SPIEGEL und des internationalen Recherchenetzwerks Bellingcat ergänzt. Während der SPIEGEL ein etabliertes, klassisch journalistisches und zuallererst wöchentlich erscheinendes Printprodukt ist, handelt es

sich bei Bellingcat um ein Recherchenetzwerk, das als Vorreiter im Bereich der OSINT-Recherchen gilt, ausschließlich online publiziert und keinen festen Publikationsrhythmus besitzt. Insofern wird durch die beiden Redaktionen ein möglichst großes Spektrum unterschiedlicher Aspekte abgedeckt. Ein besonderer Fokus liegt dabei auf Problemen bei der Verifizierung audiovisueller Inhalte.

Zu beachten ist, dass die Dokumentationsabteilung des SPIEGEL in erster Linie für die Verifizierung der Printausgabe zuständig ist und es sowohl für die Online-Inhalte wie auch für Investigativ- und OSINT-Recherchen weitere davon weitgehend unabhängige Abteilungen/Teams im SPIEGEL gibt, die potenziell mit der Verifizierung audiovisueller Inhalte zu tun haben. In der Dok.-Abteilung des SPIEGEL arbeiten insgesamt fast 70 Dokumentar:innen, die nach ihren Fachgebieten sortiert alle Inhalte für die Printausgabe des Magazins prüfen, einschließlich jeder Bildunterschrift. Bis 2022 gab es eine eigene Bilddokumentationsabteilung, so dass Text- und Bildinhalte getrennt überprüft wurden. Seit diesem Jahr werden stattdessen alle Beiträge gemeinsam mit den dazugehörigen Bildinhalten überprüft. Grund für die Zusammenlegung sei allerdings nicht etwa ein Stellenabbau, vielmehr gehe es um eine Optimierung der Arbeitsprozesse. Die neue Struktur bewähre sich, weil die Bildverifikation für den SPIEGEL sehr stark im Kontext des Beitrages stattfindet und so alle notwendigen Informationen in einer Hand lägen (vgl. Int. Sauerbier 6ff., 283ff.).

Die Überprüfung des Bildmaterials besteht für die Printausgabe des SPIEGEL in erster Linie darin, zu kontrollieren, ob Beitrag, Bildinhalt und Bildunterschrift zusammenpassen. Besonders die Abbildung unprominenter Personen, Orte, Strukturen (wie Militär, Polizei oder Sicherheitskräfte – gerade im fremdsprachigen Raum), von Motiven aus der Tier- und Pflanzenwelt, technischer Inhalte oder auch historischer Fotos sei fehleranfällig. Zudem sei es wichtig, die Zuverlässigkeit der einzelnen Agenturen einschätzen zu können. Eine tatsächliche Echtheitsprüfung findet hingegen kaum statt. Dies sei deshalb möglich, weil die Bildinhalte für den SPIEGEL im Regelfall ausschließlich aus vertrauenswürdigen Quellen wie Agenturen und von eigens beauftragten Fotograf:innen stammen (vgl. ebd. 160ff., 320ff.):

„Grundsätzlich gucken wir auf viele Bilder mittlerweile nicht mehr so genau drauf, wo wir wissen, da muss man auch nicht so genau gucken. [...] Im Großen und Ganzen vertrauen wir darauf, dass das Bild echt ist, wenn es von einer guten Quelle kommt, wie zum Beispiel dpa, Reuters, AFP und wie sie alle heißen. Wenn ich eine schlechte Quelle habe [...] dann gucke ich mir das natürlich schon mal genauer an. Gucke mir vielleicht mal an, wo ist das schon mal veröffentlicht worden? Man guckt mit so einer Art – das lernt man mit den Jahren – gesundem Menschenverstand darauf [...] Aber nun haben wir eben auch, wie ich schon sagte, gar nicht so seltsames Bildmaterial bei uns.“ (Int. Sauerbier 312f., 380ff.)

Die Verwendung nutzergenerierter Inhalte komme auch mal vor, sei aber die Ausnahme. Die Möglichkeit, mehr nutzergeneriertes Material verwenden zu können, fehlt Sauerbier nicht. Im Zweifelsfall behelfe man sich mit Illustrationen oder Symbolbildern. Dies sei auch die Konsequenz einer Absprache im Zusammenhang mit der Auflösung der Bilddokumentationsabteilung: Dass man mit der Bildauswahl von vorneherein vorsichtig sein und sich überwiegend aus zuverlässigen Quellen bedienen müsse, weil nicht alles genau geprüft werden könne (vgl. ebd. 191ff., 266ff., 298ff.):

„Wir sind halt keine Forensiker, also von der technischen Seite können wir das nicht großartig prüfen. Und die Zeit ist auch nicht da. Wie ich schon sagte: Wenn so am Donnerstagabend oft erst spät Bilder kommen, da kann sich keiner noch mal sechs, sieben Stunden damit beschäftigen 'Ist das jetzt zuverlässig oder nicht?'“ (Int. Sauerbier 207ff.)

Grund für den Zeitdruck sei zum einen, dass es den fertigen Text brauche, um über die Bebilderung zu entscheiden, zum anderen würden viele Leute mitreden und auch für das Gesamtbild des Heftes wenige Stunden vor Redaktionsschluss noch Änderungen vornehmen (vgl. ebd. 136ff.).

Zu hinterfragen ist der unterschiedliche Umgang mit Print- und Online-Inhalten beim SPIEGEL. Dass alle Print-Inhalte – Texte und Bilder – durch die Dok.-Abteilung verifiziert werden, jedoch längst nicht alle Online-Inhalte, sei auf die gewachsene Struktur rund um die Produktion der Online-Inhalte zurückzuführen. Die Online-Produktion habe als „Spielwiese“ mit einem ganz kleinen Team angefangen, dessen Beiträge lange Zeit zudem ausschließlich kostenlos verfügbar waren. Entsprechend sei das Heft als Premium-Produkt angesehen und dort auch mehr Energie in die Verifizierung investiert worden. Bei den Online-Inhalten hingegen „wurde gar nicht groß darüber nachgedacht, ob das auch jemand verifizieren muss“ (ebd. 430, 441). Inzwischen verifiziere die Dok.-Abteilung zumindest von den kostenpflichtigen Spiegel+-Artikeln textseitig so viele wie

möglich. Alle zu verifizieren sei jedoch zeitlich unmöglich und die Verifizierung der zugehörigen Bildinhalte weiter die Ausnahme: „Da sind tatsächlich die Redakteure selbst verantwortlich für ihr Bild- und Videomaterial, also das zu prüfen“ (ebd. 21ff.). Bei der Menge und Taktung von Online-Artikeln sei es auch gar nicht anders möglich: „Die Artikel müssen immer schnell auf die Seite und wenn ein Dokumentar das alles noch prüfen soll... das ist gar nicht zeitlich zu schaffen“ (ebd. 448f.). Bisher scheine das aber kaum ein Problem gewesen zu sein und gut zu funktionieren. Trotzdem sei gerade in letzter Zeit intern ein Thema, wie man den Leser:innen die unterschiedlichen Verfahrensweisen mitteilen und erklären könne (vgl. ebd. 426ff.). Insgesamt zeigt die Vorgehensweise des SPIEGEL, dass die Redaktion einen Weg gefunden hat, mit den gegebenen redaktionellen Möglichkeiten das Risiko im Umgang mit manipulierten Inhalten zu minimieren, aber sie zeigt auch die Bewegung im Umgang mit den veränderten Rahmenbedingungen durch die Digitalisierung.

Im Gegensatz dazu ist Bellingcat deutlich stärker auf die daraus neu entstandenen Möglichkeiten fokussiert und arbeitet viel mit nutzergeneriertem Material, um gesellschaftlich relevante Sachverhalte zu belegen. Dabei greife das Recherchenetzwerk vorwiegend auf semantische Verfahren (siehe Kapitel 4.2.1) und die Verifizierung des Inhalts durch den Abgleich mit weiteren Quellen zurück:

„Wir setzen uns Stunden oder auch Tage oder Wochen hin und schauen: Finden wir andere Fotos oder Videos, die dasselbe zeigen, die uns diesen Ort zeigen? Und so verifiziert man eben ein einziges Video oft dadurch, dass man es in dem Kontext sieht und schaut: Was unterstützt das, was man da sieht, denn sonst noch so? Was für andere Aufnahmen finde ich?“ (Int. Wild 291ff.)

Der Hauptgrund, weshalb sowohl pixelbasierte als auch formatbasierte Methoden bei Bellingcat nicht im Fokus der Verifizierungsprozesse stünden, liege darin, dass sie häufig mit verwackelten Aufnahmen von geringer Auflösung arbeiten würden, die keine Metadaten besitzen:

„Also wenn das in einem Foto drin ist, schauen wir uns das auch an. Aber das findet man tatsächlich eher selten, wenn man einfach nur online recherchiert. [...] Wir finden unsere Materialien oft in den sozialen Netzwerken und die sozialen Netzwerke entfernen die Metadaten, bevor sie sie uns zur Verfügung stellen. Das heißt, damit können wir gar nicht arbeiten. Das heißt, bei uns geht es wirklich in ganz vielen Fällen darum zu schauen: Was ist denn sichtbar in so einem Foto? Natürlich schauen wir uns auch so Dinge an wie Error Level Analysis. Also es gibt etli-

che Tools, auch frei verfügbare, die so ein bisschen schauen: Okay, wenn man ein Foto komprimiert, dann verändert sich das Foto in einer gewissen Art und Weise und wenn man dann zwei zusammenpackt, dann sieht man da Unregelmäßigkeiten manchmal. Also sowas schauen wir uns schon auch an. Aber es ist in den seltensten Fällen wirklich der ausschlaggebende Punkt bei uns.“

(Int. Wild 62ff.; vgl. 46ff.)

Auch Johanna Wild spricht wie Andrea Sauerbier den Faktor Zeit an. Sie sieht große Vorteile gegenüber klassischen Redaktionen wie dem SPIEGEL darin, dass Bellingcat keinem Veröffentlichungsdruck ausgesetzt sei:

„In einer normalen, traditionellen Redaktion hat man einfach mehr Druck, mehr zu veröffentlichen, regelmäßiger zu veröffentlichen. Und das haben wir eben nicht. Wir können uns so viel Zeit nehmen, wie wir wollen. Und wir haben auch den ganzen Tag Zeit, uns eben nur mit dieser Online-Recherche zu beschäftigen. Und dadurch gewinnt man viel Erfahrung, viele Kenntnisse, die man eben einfach langsamer gewinnt, wenn man auch noch Interviews macht, wenn man auch noch zu Terminen geht. [...] Andere Journalisten, auch investigative Journalisten, haben eben viel mehr Aufgaben, die sie gleichzeitig bewältigen müssen, und da ist es schwieriger so eine Spezialisierung aufzubauen.“ (Int. Wild 108ff.)

In dem Zusammenhang sei eine gute Entwicklung, dass es wieder mehr und mehr Investigativeinheiten in deutschen Redaktionen gebe (vgl. ebd. 106ff.). Zeit sei auch wichtiger als der Ausbau technischer Hilfsmittel:

„Hilfsmittel helfen uns überhaupt nicht, wenn wir keine Zeit haben uns damit zu beschäftigen. Und das ist oft der Knackpunkt, würde ich sagen, noch in Deutschland.“ (Int. Wild 147ff.)

Die gleiche Prioritätensetzung wird auch bei Jochen Spangenberg deutlich, der das Gehirn als wichtigstes Verifizierungsinstrument benennt: „Vieles kann man wirklich mit dem genauen Hinschauen, mit dem genauen Zuhören und mit dem Innehalten, mit Nachdenken und mit dem Einschalten des Gehirnes schon mal – nicht gleich verifizieren, aber zumindest kann man schon mal so einen ersten Reality-Check damit machen“ (Int. Spangenberg 221ff.; vgl. 214ff., 558ff.). Manchmal sei es gar nicht die Technik, die die Lösung gibt, sondern ganz klassische Dinge wie anzurufen, hinzufahren oder jemanden hinzuschicken.

Auch Riess (vgl. Int. 556f.) und Zmudzinski (vgl. Int. 528) sprechen Zeit als entscheidenden Faktor an, für IT-Forensiker:innen genauso wie für Redaktionen. Darum sei die Erwartungshaltung gegenüber Journalist:innen, Inhalte sehr schnell aufzubereiten, ein

Problem: „Das ist natürlich im direkten Widerspruch dazu, dass man Sachen prüft“ (Int. Riess 560). Daher sieht Riess Wochenmagazine wie SPIEGEL, ZEIT oder auch den amerikanischen New Yorker in einer besseren Ausgangssituation für den Umgang mit manipulierten Inhalten als enger getaktete Redaktionen. Und da es schlussendlich auch eine Kosten-Nutzen-Frage sei, sei zudem das öffentlich-rechtliche Rundfunksystem an der Stelle wirklich vorteilhaft und in einer vergleichsweise komfortablen Situation: „Wie es mit privaten Redaktionen aussieht, na ja, schwierig. Die Nürnberger Zeitung zum Beispiel oder so, die hat jetzt hier nicht die Möglichkeiten, da groß nachzufassen. Das geht dann wirklich nur bei so ganz Großen“ (Int. Riess 542ff.; vgl. 3, 522ff.).

Auch wenn der Faktor Zeit am wichtigsten sei, nennen die Interviewten technische Hilfsmittel, die besonders hilfreich seien: „Ohne Satellitenbilder würde es kaum gehen“, meint Johanna Wild (Int. 181f.; vgl. 183f.). Außerdem helfe alles, was dazu beiträgt, Social-Media-Plattformen zu durchsuchen. Jochen Spangenberg hebt die inverse Bildersuche hervor, die von sehr vielen Journalist:innen als Einstieg in die Verifizierung genutzt werde, und betont die Bedeutung der unterschiedlichen Anbieter: „Als Journalist, Journalistin finde ich sollte man unbedingt wissen, dass es mehr als Google gibt. Man sollte als erstes überhaupt erst mal wissen, dass es Bilderrückwärtssuchmaschinen gibt. Das wissen auch viele noch nicht. Und dann, dass es nicht nur Google gibt, sondern auch Bing oder Yandex oder TinEye“ (Int. Spangenberg 233ff.; vgl. 228ff.). Aber auch die selbst entwickelte Browser-Erweiterung, das *InVID-WeVerify-vera.ai-Plugin*, habe inzwischen über 90.000 zweiwöchentliche Nutzer:innen und werde immer wieder als das de-facto-Tool genannt. Es ermöglicht u. a. die gebündelte Nutzung mehrerer inverser Bildersuchen. Plattformen wie das Plugin, welches mehrere Hilfsmittel gebündelt zur Verfügung stellt, seien wichtig, um die Hemmschwelle für Journalist:innen zu senken, sich mit der Masse an technischen Hilfsmitteln auseinanderzusetzen. Es sei gerade zu Beginn sehr schwierig, dort einen Überblick zu gewinnen, zu behalten und zu verstehen, welches Werkzeug in welchem Fall hilfreich sei, was auch Johanna Wild als Problem sieht:

„Es gibt nicht das eine Tool, das alle miteinander verbindet. Das ist tatsächlich was, was sich die meisten, die anfangen in dem Bereich, wünschen. [...] Aber das gibt es leider nicht. Also dieses Feld der Tools im Bereich der Online-Recherche ist



tatsächlich extrem fragmentiert. Und das macht es sehr herausfordernd für Leute, die anfangen.“ (Int. Wild 169ff.; vgl. Int. Spangenberg 116ff.)

Häufig fehle im Umgang mit technischen Hilfsmitteln noch die Erfahrung und gerade ältere Journalisten hätten nie gelernt, überhaupt mit solchen Werkzeugen umzugehen:

„Das ist eindeutig eine Herausforderung, dass viele Journalisten in den Journalistenschulen, die sie besucht haben, noch vor einiger Zeit im Studium, eigentlich nie mit solchen Recherchemethoden zu tun hatten. Das ändert sich jetzt gerade, [...] dass viel mehr, ja... Fortbildung angeboten wird in dem Bereich. [...] Aber gerade die Generation davor, die das noch nicht hatte in ihrer Ausbildung, für die ist das natürlich schwierig, weil die nie mit solchen Verifizierungsmethoden oder Recherchemethoden in Berührung gekommen sind“ (Int. Wild 124ff.)

Zudem handele es sich um ein sehr dynamisches Feld, das sich immer wieder neuen Gegebenheiten – wie veränderten Strukturen der Social-Media-Plattformen – anpassen müsse und häufig aus Zwischenlösungen bestehe, die nicht besonders robust seien und ein genaues Verständnis der Methoden voraussetzen würden, auf denen sie beruhen (vgl. Int. Wild 187ff.; Int. Riess 532ff.). Auch Zmudzinski sieht eine Hürde in der richtigen Bedienung vorhandener Werkzeuge. Zudem sei bei online verfügbaren Tools, bei denen man sein Material zur Überprüfung hochladen muss, mitunter auch Daten- bzw. Quellenschutz ein Problem. Es gebe durchaus kommerzielle Software, die dieses Problem umgehe und zudem alle notwendigen Möglichkeiten audiovisuelles Material forensisch zu analysieren, bereitstelle. Die seien mit Kosten von mehreren tausend Dollar für eine einzelne Lizenz aber ziemlich teuer und seien auch nicht leicht zu bedienen. Möglicherweise sind sowohl die Kosten wie auch die anspruchsvolle Bedienung Gründe dafür, dass solche Software laut Zmudzinski (vgl. Int. 445ff.) bisher nicht von Redaktionen genutzt wird.

In jedem Fall seien Journalist:innen keine IT-Forensiker:innen und eine forensische Analyse könne man auch nicht von ihnen erwarten (vgl. Int. Riess 527ff.; Int. Spangenberg, 736f.). Eine mögliche Lösung könnte der Ausbau einer Zusammenarbeit zwischen Redaktionen wie dem SPIEGEL, Recherchenetzwerken wie Bellingcat und IT-forensischen Spezialist:innen wie dem Fraunhofer SIT sein, wie sie bereits teilweise stattfindet, als bestehende Möglichkeit allerdings nicht überall bewusst wahrgenommen wird: „Das kann sehr, sehr hilfreich sein für Journalisten und Journalistinnen. Weil das... Ich glaube diese Expertise selbst im Haus aufzubauen, also ein forensisches

Team, das wie ein Team der Fraunhofer Researcher arbeitet, das will man sich nicht leisten. [...] Weil so viele Anforderungsfälle, glaube ich, wird es da nicht geben“ (Int. Spangenberg 755ff; vgl. Int. Wild 82ff.; Int. Zmudzinski 49ff., 494ff.).

#### **4.4 Einschätzung der tatsächlichen Detektierbarkeit audiovisueller Medien – heute und zukünftig**

Die restliche Ergebnisdarstellung fasst Einschätzungen der Interviewpartner:innen zusammen. Es geht also in den folgenden Kapiteln um die Meinungen der Beteiligten. Diese werden miteinander verglichen – zunächst in Bezug auf die tatsächliche Detektierbarkeit audiovisueller Manipulationen. Die Interviewten wurden dafür alle bis auf Andrea Sauerbier, für die intentionale Manipulationen im Redaktionsalltag keine Rolle spielen (vgl. Int. Sauerbier 218ff.), gefragt, ob sie denken, dass es aktuell möglich ist, jede Manipulation zu erkennen. Es ist interessant zu sehen, dass sich die Interviewten hierbei nicht einig sind, dass dies jedoch im Großteil auf die unterschiedlichen Blickrichtungen und Spezialisierungen zurückgeführt werden kann.

Eine technische Möglichkeit, alle Manipulationen als solche zu detektieren, sieht aktuell nur der Deepfake-Experte Dominique Dresen. Aufgrund der aktuell noch vorhandenen Fehler und Artefakte sei es mit genügend Aufwand bisher möglich, synthetische Medien als solche zu erkennen. Es gebe allerdings zunehmend Randfälle, wo es bereits sehr schwierig werde und der Moment, wo eine sichere Zuordnung nicht mehr möglich ist, komme bestimmt. Er sehe keine physikalische Grenze, welche der Erstellung perfekter Manipulationen in Jahrzehnten widerspreche. Dann müsse man den Ansatz, wie man verifiziert, neu denken, Kontext und andere Quellen mit beachten oder andere Wege finden, authentisches Material als authentisch darzustellen. Eine eindeutige Antwort, ob ein audiovisuelles Dokument authentisch ist, könne es bei der Möglichkeit perfekter Manipulationen nicht mehr geben (vgl. Int. Dresen 204ff., 548ff., 992ff.).

Jutta Jahnel sieht den Moment hingegen schon gekommen, dass man nicht mehr eindeutig sagen könne, ob etwas echt sei oder synthetisch generiert:

„Es ist technisch ja schon schwierig, Deepfakes zu identifizieren. [...] Aber es ist noch schwieriger, mit Sicherheit technisch zu sagen: 'Das ist kein Deepfake'. [...] Man müsste eigentlich immer weiter suchen.“ (Int. Jahnel 163ff.)

Und wenn etwas die Detektion synthetischer Medien in Zukunft ermögliche, dann seien es ML-basierte Verfahren (vgl. Int. Jahnel 330ff.). Auch Sascha Zmudzinski und Christian Riess sehen bereits heute keine absolute Sicherheit in forensischen Verfahren. Während Sascha Zmudzinski (vgl. Int. 119ff., 748ff.) sich optimistisch zeigt, dass man heute wie in Zukunft auf Detektionsseite trotzdem viel tun können wird, empfindet Christian Riess es auch heute schon als relativ einfach, Manipulationen zu erzeugen, die plausibel seien und bei denen alle Detektoren fehlschlagen: „Ich denke, da gibt es mehr Lücken als Lösungen tatsächlich“ (Int. Riess 169; vgl. 447ff.). Zumal die nach Riess besonders große Errungenschaft der Multimedia-Forensik, Fotos einer spezifischen Kamera aufgrund ihres einzigartigen Sensorabdrucks zuzuordnen, mit der Existenz synthetischer Medien an Bedeutung verliere. Er ist genauso wie Johanna Wild der Meinung, dass für die Verifizierung audiovisueller Inhalte bereits heute zusätzliche Recherchen über das eigentliche Material hinaus nötig sind, unterstützt durch semantische Verfahren (vgl. Int. Riess 178ff., 347ff.) und/oder auch den Abgleich mit anderen Quellen (vgl. Int. Wild 277ff.):

„Wir würden uns in den seltensten Fällen auf ein einziges Video verlassen. [...] Wenn wir zehn Aufnahmen von demselben Ort finden, von demselben Ereignis, von verschiedenen Nutzern online gepostet, dann ist die Wahrscheinlichkeit sehr gering, dass da einer irgendwas... dass alles manipuliert wurde“ (Int. Wild 287ff.)

Unter der gleichen Prämisse ist Jochen Spangenberg (vgl. Int. 766ff., 831ff.) überzeugt, dass aktuell jede Manipulation mit genügend Zeit und Aufwand erkannt werden kann und glaubt auch kaum, dass sich das ändern wird. Obwohl es in einzelnen Fällen sehr schwierig werden könne. Ein Gedanke, den auch Andrea Sauerbier teilt:

„Man arbeitet ja sehr viel mit Vergleichsbildern. [...] Vielleicht kann mir das auch in Zukunft noch helfen, damit ich weiß 'Hat dieses Ereignis überhaupt stattgefunden, was ist denn da vor Ort gewesen?' Aber es wird natürlich auch zunehmend Situation geben, da ist kein anderer gewesen, da haben wir wahrscheinlich keine anderen Belege oder Möglichkeiten irgendetwas zu prüfen. Ich glaube schon, dass da große Herausforderungen auf einen zukommen, wie man denn sowas überhaupt gegenchecken will, wenn etwas am anderen Ende der Welt stattgefunden hat, zum Beispiel. Manche Sachen lassen sich vielleicht über Plausibilität erschließen.“ (Int. Sauerbier 574ff.)

Aber nur weil es Grenzfälle und möglicherweise auch perfekte Manipulationen geben wird, werden nicht alle Manipulationen perfekt sein, betont Riess. Ein Potenzial, Fehler

zu machen, bleibe immer erhalten. Entsprechend würden klassische modellbasierte Methoden der Multimedia-Forensik, manuelle Verfahren wie physikbasierte Methoden, weiter relevant bleiben (vgl. Int. Riess 377ff.; Int. Zmudzinski 141ff.). Weder Dresen noch Riess sind hingegen „überzeugt“ (Int. Dresen 482ff.) bzw. „Glaubensanhänger von diesen vollautomatischen Sachen“ (Int. Riess 191f.). Zmudzinski hingegen benennt ML-basierte Detektionsverfahren als die forensische Methode mit dem größten Potenzial. Er zeigt sich optimistisch, dass automatisierte Detektionsmethoden in Zukunft auch zuverlässig auf unbekanntes Datenmaterial anzuwenden sein werden. Es hänge an ausreichend reichhaltigen und universellen Trainingsdaten. Ablösen werden sie die traditionellen Verfahren aber auch nach Zmudzinski (vgl. Int. 126ff.) nicht.

Eine weitere technische Möglichkeit, die Authentizität audiovisueller Inhalte sicherzustellen, sind Methoden der aktiven Forensik. Dominique Dresen (vgl. Int. 974ff.) ist allerdings skeptisch, dass sich eine fälschungssichere Markierung audiovisueller Inhalte durchsetzt. Und auch Johanna Wild (vgl. Int. 253ff.) sieht nur dann eine Chance für die Verbreitung eines solchen Verfahrens, wenn die Grundlage dafür als Standardfunktion auf jedem Gerät installiert ist. Vorstellbar findet Christian Riess (vgl. Int. 290ff.) hingegen, dass synthetische Fotos bei einer überschaubaren Anzahl beteiligter Akteure von diesen durch ein unsichtbares Wasserzeichen markiert werden. Damit könne allerdings nicht verhindert werden, dass böswillige Akteure durch eigene Systeme weiter Manipulationen ohne Kennzeichnung verbreiten und auch für die in Kapitel 3.3.3 zusammengetragenen Probleme müssten noch Lösungen gefunden werden (vgl. Int. Zmudzinski 266ff.). Insgesamt könnten technische Hilfsmittel u. a. bei zeitaufwendigen Aufgaben eine sehr starke Unterstützung sein, Journalist:innen ihre Verantwortung aber nicht abnehmen:

„Das heißt, der Algorithmus oder die Technik, wie auch immer, unterstützt und kann super, super tolle Hilfestellungen bieten, aber letztendlich die ultimative Entscheidung, publiziere ich jetzt etwas und mit einem bestimmten Label, nämlich das stimmt oder ich sage jetzt, das war, ja, das war eine Rakete von, abgefeuert von X auf Ort Y... Diese ultimative Entscheidung wird, glaube ich, in absehbarer Zeit immer bei einem Menschen liegen, also bei, in unserem Fall, Journalistinnen und Journalisten aus den genannten Gründen.“ (Int. Spangenberg 501ff.)

#### 4.5 Einschätzungen zu Einfluss u. Risiken audiovisueller Manipulationen

In Bezug auf die tatsächlichen Auswirkungen, die aus der veränderten Manipulationslandschaft folgen, waren die Expert:innen recht vorsichtig mit ihren Einschätzungen. Sascha Zmudzinski sieht, dass audiovisuelle Manipulationen in Redaktionen als Bedrohung für die Vertrauenswürdigkeit der Medien und auch ihrer eigenen Existenz wahrgenommen würden. Gerade für Agenturen sei es wichtig, garantieren zu können, dass ihr audiovisuelles Material und die damit verbundenen Informationen echt sind. Andernfalls verlören sie ihre Bedeutung: „Das ist die Information, mit der sie handeln und eben auch mit der Vertrauenswürdigkeit darin“ (Int. Zmudzinski 435f.). Gerade mit manipuliertem Videomaterial könne man besonders großen Schaden anrichten, weil man intuitiv immer noch ein hohes Vertrauen darin habe und Deepfakes für viele noch eine sehr abstrakte Gefahr seien (vgl. Int. Dresen 744ff.). Jutta Jahnel betont, dass das Problem audiovisueller Manipulationen schon vorher dagewesen, aber nicht so im Bewusstsein gewesen sei. Zudem hätten synthetische Medien die Problematik verstärkt, weil es einfacher und kostengünstiger geworden sei, zugleich noch authentischere Manipulationen zu erstellen. Jahnel (vgl. Int. 92ff., 453f.) sieht genauso wie Zmudzinski (vgl. Int. 525ff.) eine sehr reale Gefahr darin, dass Medienschaffende sich durch Deepfakes täuschen lassen. Schließlich erfordere der richtige Umgang mit Deepfakes ein ständiges Wissen über deren aktuelle Schwachstellen, zum anderen stelle der Zeitmangel im Journalismus ein Risiko dar. Auch Jochen Spangenberg (vgl. Int. 524ff.) sieht in Deepfakes zumindest das Potenzial für größere Auswirkungen auf den Journalismus, auch wenn es dafür bisher wenige Beispiele gebe. Ähnlich äußert sich Johanna Wild (vgl. Int 30f., 308ff.), die synthetische Medien ebenfalls als Herausforderung sieht, auch wenn sie es derzeit für sehr unwahrscheinlich hält, von Deepfakes getäuscht zu werden und insgesamt die Verbreitung dekontextualisierter Medien zurzeit als größte Herausforderung wahrnimmt. Andrea Sauerbier bewertet das Risiko durch audiovisuelle Manipulationen in ihrem Redaktionsalltag hingegen insgesamt als eher gering. Für manipuliertes Material gebe es kaum Gelegenheit:

„Eben weil wir uns auch von Agenturen bedienen und weil wir aber auch Themen haben, wo ich gar nicht wüsste, wo man da mit einer Manipulation ankommen sollte. Also wenn wir jetzt so Bilder haben: Pistorius auf dem Truppenübungsplatz

oder Merkel all die Jahre, die sich mit irgendwelchen anderen Politikergrößen trifft. Was sollte da manipuliert sein?“ (Int. Sauerbier 250ff.; vgl. 235ff.)

Gleichzeitig seien die Möglichkeiten künstlicher Intelligenz beim SPIEGEL durchaus Thema. Gerade in der Woche des Interviews habe es in der Redaktion eine Vollversammlung zu den Chancen und Risiken durch KI-Technologien für den SPIEGEL gegeben (vgl. ebd. 556ff.). Christian Riess (vgl. Int. 254ff.) schätzt, dass man sich schnell an die künstliche Generierung audiovisueller Medien gewöhnen und die neuen Möglichkeiten auch nutzen werde, um zum Beispiel bisher übliche Gerichtszeichnungen zu ersetzen. Eine potenzielle Vereinfachung, die auch Andrea Sauerbier (vgl. Int. 611ff.) anspricht: Vielleicht könnten synthetische Medien in etlichen Bereichen auch eine Hilfe sein, selbst wenn es viele Fotograf:innen sicherlich ihre Jobs kosten werde. Zumindest im Rahmen politischer Fotografie, Krisen und Konflikte sei man weiter auf authentisches Bildmaterial angewiesen. Auch Jutta Jähnel (vgl. Int. 420ff.) betont, dass es für synthetische Medien auch positive Einsatzzwecke gebe, zum Beispiel, um Taubstummen eine Stimme zu geben.

Problematisch ist für Christian Riess allerdings eine mögliche Machtkonzentration auf wenige Akteure, wie sie auch in anderen Bereichen der Informatik zu beobachten sei: „Wie viele von diesen Systemen wird es tatsächlich geben auf der Welt? Also diesen Foundation-Models sozusagen, die entsprechend aufwendig trainiert sind, um diese Technik anzubieten. Ist da dann vielleicht auch so eine Konzentration, wie man es in anderen Informatikbereichen sieht? Also in anderen Informatikbereichen ist es ja immer so: Man hat Google, Facebook, Apple und Microsoft“ (Int. Riess 282ff.). Gleichzeitig erhöhe eine geringe Anzahl beteiligter Akteure aber die Chance, synthetische Medien recht flächendeckend mit Wasserzeichen oder Signaturen als solche zu kennzeichnen.

Ein Risiko, das von mehreren Expert:innen über die verschiedenen Interviews hinweg angesprochen worden ist und teilweise für gefährlicher als das Manipulationspotenzial an sich gehalten wird, findet sich in der Literatur als *Liars Dividend*. Der Begriff umschreibt, dass durch die Möglichkeit authentisch wirkender Manipulationen potenziell alles manipuliert und damit auch authentische Inhalte angezweifelt werden können.

Damit einhergehend würden Narrative wie die russische Reaktion auf das Bild- und Videomaterial aus Butscha Zweifel säen, in den Köpfen zurückbleiben und seien schlussendlich eine Gefahr für Demokratie und Gesellschaft (vgl. Int. Dresen 206ff.; Int. Jahnel 136ff., 458f.; Int. Spangenberg 572ff.):

„Diese Verunsicherung in der Gesellschaft, dass man alles irgendwie als potenziell falsch oder manipuliert betrachtet, das ist, glaube ich, ein großes Problem für die Gesellschaft. [...] Dass das dann dazu führt, dass man gar nichts mehr glaubt oder allen gegenüber skeptisch ist. Dass das Vertrauen auch in Instanzen wie den Journalismus noch mehr schwindet, was schlimm wäre. Und das könnte passieren, da Akteure auf allen möglichen Ebenen ein Interesse daran haben, Gesellschaften zu destabilisieren und aktiv die Narrative beeinflussen.“ (Int. Spangenberg 572ff.)

#### 4.6 Was muss getan werden? Mögliche Handlungsfelder

Auch bei möglichen Lösungsansätzen oder Handlungsfeldern waren die Interviewten eher zurückhaltend. Im Prinzip müsse man an jeder Ecke gleich viel tun, sagt Jutta Jahnel (vgl. Int. 305ff.) – in technischer und regulatorischer Hinsicht genauso wie bei der Medienkompetenz der Rezipient:innen. Eine einfache Lösung gebe es nicht.

**Im Bereich der technischen Verifizierungsinstrumente** sei eher die Unübersichtlichkeit als das Fehlen von Werkzeugen ein Problem, wie bereits in Kapitel 4.3.2 angeklungen ist (vgl. Int. Wild 169ff.; Int. Spangenberg 116ff.). Projekte und allgemein Anstrengungen, die sich nicht mit der Entwicklung technischer Hilfsmittel, sondern vielmehr mit deren übersichtlicher Darstellung samt ihrer Einsatzbereiche und Möglichkeiten beschäftigen, wurden im Theorieteil der Arbeit vernachlässigt, existieren aber<sup>32</sup>. Vielleicht kann mit dem weiteren Ausbau übersichtlicher Zusammenstellungen die Problematik etwas entschärft werden. Zumindest in absehbarer Zeit wird das Feld technischer Verifizierungsinstrumente aber wohl aufgrund der dynamischen Entwicklung des Anwendungsbereichs weiter fragmentiert bleiben (müssen). Entsprechend ergebe sich auch der Bedarf technischer Hilfsmittel aus den jeweils aktuellen Herausforderungen. Grundsätzlich sei alles sehr hilfreich, was die Arbeit mit Satellitenbildern vereinfacht und kostengünstiger macht. Auch helfe alles, was dabei unterstützt, Social-Media-

---

<sup>32</sup> Bspw. im Rahmen der Projekte KID und KID2 („Künstliche Intelligenz gegen Desinformation“) der DW-Innovationsabteilung wurden Webseiten mit diesem Ziel erstellt (<https://www.howtoverify.info/>, zuletzt geprüft am 27.06.23). Auch Bellingcat stellt seit Jahren eine Excel-Liste mit aktuell verfügbaren Tools zur Verfügung ([bit.ly/bcattools](https://bit.ly/bcattools), zuletzt geprüft am 27.06.23).

Plattformen zu durchsuchen. Gerade sei eine große Sorge, inwiefern man zukünftig noch Twitterdaten sammeln könne – was bisher einfach gewesen sei, sich mit den aktuellen Entwicklungen bei Twitter aber gerade ändere. An „dem einen großen Tool“ (Int. Wild 209f.) bestehe zumindest bei Bellingcat kein Interesse. Zudem reiche es nicht, ein Werkzeug zu entwickeln. Man müsse anschließend auch die Kapazität haben, es zu betreuen und am Leben zu erhalten (vgl. ebd. 175ff.).

Jochen Spangenberg (vgl. Int. 66ff.) spricht die möglichst einfache Anwendung der inversen Bildersuche auf Videomaterial als Herausforderung an, für die noch keine vollumfängliche Lösung existiere. Zudem wünscht er sich Software-Werkzeuge, die für Redakteur:innen vorselektieren und ihnen zeitaufwendige Arbeiten abnehmen. Teilweise werden solche Tools in Bezug auf Textverifizierung derzeit bereits vom Fraunhofer SIT entwickelt (vgl. Int. Zmudzinski 545ff.). Besonders wichtig ist für Jochen Spangenberg (vgl. Int. 462ff.), dass die technischen Hilfsmittel so einfach und verständlich wie möglich entwickelt werden. Die Ergebnisse der Hilfsmittel müssten für Journalist:innen nachvollziehbar sein.

Dafür sei natürlich **auf journalistischer Seite** nötig, den Umgang mit audiovisuellen Manipulationen und den entsprechenden Verifizierungsinstrumenten zu lernen, was bisher noch nicht überall der Fall sei (vgl. Int. Spangenberg 367ff., 482ff.; Int. Sauerbier 595ff.). Johanna Wild (vgl. Int. 136ff.) wünscht sich in Redaktionen dabei mehr Unterstützung der Personen, die sich darin einarbeiten, länger mit einem Thema beschäftigen und in der Zeit nicht so viel veröffentlichen. Außerdem hofft sie, dass es irgendwann auch im Lokaljournalismus mehr investigative Einheiten geben wird, damit auch regional investigative Recherchen betrieben werden können.

In Bezug auf die konkrete Berichterstattung müsse man sich durch die Existenz synthetischer Medien neue Gedanken über Standards wie eine saubere Quellenkennzeichnung machen (vgl. Int. Riess 313ff.). Außerdem sei es wichtig, technisch korrekt über Manipulationen zu berichten, damit in der Bevölkerung kein falscher Eindruck von den Möglichkeiten und Grenzen audiovisueller Manipulationen entstehe, der auch die Glaubwürdigkeit authentischer Inhalte unnötig in Zweifel ziehe (siehe Kapitel 4.5, vgl. Int.



Dresen 194ff.).<sup>33</sup> Außerdem sei eine große und wichtige Frage für Redaktionen, wie man gerade die Menschen erreiche, die Verschwörungserzählungen aufsitzen und dadurch nicht so leicht zu erreichen sind (vgl. Int. Spangenberg 713ff.).

Insgesamt sei wichtig, die **Medienkompetenz** der Rezipient:innen zu stärken. Gegenüber den gesellschaftlichen Auswirkungen der Manipulationen sei Prävention durch die Sensibilisierung der Bevölkerung die erste Gegenmaßnahme (vgl. Int. Dresen 69ff., 329ff.). Jochen Spangenberg unterstreicht die Relevanz des Handlungsfeldes durch seine persönlichen Erfahrungen in Schulen als Mitglied der NGO Lie Detectors:

„Das kann ich, glaube ich, auch ohne Empirie sagen, dass da noch wirklich sehr viel Luft nach oben ist, was halt die Medienbildung von jungen Menschen angeht [...] Bei manchen würde es auch nicht schaden mal zu erklären, wie man... nicht nur, wie man Tools nutzt, das ist ja schon bald der Advanced Level, sondern auch so Dinge: Wie funktioniert denn Journalismus? Was ist denn überhaupt Journalismus? Wie funktionieren Medien? Was sind denn überhaupt öffentlich-rechtliche Medien? [...] Wie das Ganze funktioniert, wie das finanziert ist, da gibt es halt immer noch sehr viel... da gibt es einfach viel Unwissen. [...] Also junge Menschen, die können natürlich alle solche Geräte wunderbar bedienen, sehr intuitiv, können das auch sehr viel schneller als ich. Aber die Mechanismen, die dahinterstecken, wie zum Beispiel Geld verdient wird, wie Werbung funktioniert, wie so Mechanismen wie Clickbaiting funktionieren, also da ist es schon manchmal nicht so rosig.“ (Int. Spangenberg 617ff.)

Auch Jutta Jahnel (vgl. Int. 317ff.) sieht Handlungsbedarf im Bereich der Medienkompetenz. Gleichzeitig dürfe man sich mit dem Verweis auf die Medienkompetenz aber nicht aus der Verantwortung im Bereich der anderen Handlungsfelder stellen.

Dass **ein Verbot der Technologie** an sich keine Lösung sei, darin sind sich Dresen und Jahnel einig: „Wenn wir es nicht machen, dann macht es jemand anders“ (Int. Dresen 1048). Allerdings müsse man sie **regulieren** (vgl. Int. Jahnel 404ff.): Dresen (vgl. Int. 867ff., 1038ff.) meint, man könne hinterfragen, ob die Systeme für alle frei verfügbar sein sollten und ob es zumindest bei öffentlich erreichbaren Systemen eine Möglichkeit geben sollte, Produkte auf ihre:n Urheber:in zurückzuführen. Gäbe es gute Detektoren für die Erkennung synthetischer Inhalte, wäre er zudem für deren verpflichtenden Einsatz auf Social-Media-Plattformen. Auch Jahnel (vgl. Int. 271ff.) findet es wich-

---

<sup>33</sup> Das vermeintliche Telefonat von Franziska Giffey mit Vitali Klitschko wird bspw. immer wieder als Deepfake benannt – u. a. in der Talkshow von Markus Lanz am 04.04.23, als längst klar war, dass es sich höchstwahrscheinlich um altes Material handelt, das neu vertont wurde.

tig, bei der Regulierung synthetischer Medien nicht nur auf die Hersteller zu achten, sondern auch alle anderen Akteure wie die Nutzer:innen von Social-Media-Plattformen und die Plattformen selbst zu berücksichtigen. Der entworfene AI-Act auf EU-Ebene löse nur einen kleinen Teil des Problems. So habe Jahnels Arbeitsgruppe dem EU-Parlament bspw. empfohlen, Deepfakes auch namentlich im Digital Service Act<sup>34</sup> zu benennen. Das sei jedoch in einer Abstimmung gescheitert, so dass dort nur allgemein von illegalen Inhalten die Rede sei, womit nur der rechtswidrige Anteil synthetischer Medien (siehe Kapitel 4.1) erfasst werde.

Im Theorieteil dieser Arbeit und auch in den Interviews ist immer wieder der Verlust zugehöriger Informationen wie bspw. der Metadaten durch den Upload von Inhalten auf Social Media zur Sprache gekommen. Der Idee, Social-Media-Plattformen dazu anzuhalten, diese Daten stattdessen öffentlich zur Verfügung zu stellen, steht Johanna Wild (vgl. Int. 231ff.) aus Datenschutzgründen trotzdem zwiegespalten gegenüber. Schließlich seien sich nicht alle Nutzer:innen bewusst, dass ihre Inhalte überhaupt Metadaten enthalten. Helfen würde aber, wenn die Plattformen den Zugang zu ihren Daten für Researcher wie Universitäten (weiter) ermöglichen. Insgesamt tue man in Sachen Regulierung schon viel, aber das System dahinter sei zu langsam und fast überfordert (vgl. Int. Jahnelt 310ff.).

Zusätzlich zu der Regulierung audiovisueller Manipulationen durch staatliche Stellen, sieht Dominique Dresen (vgl. Int. 143ff.) aus der Perspektive des BSI auch die Notwendigkeit einer staatlichen Stelle zur Detektion. Bisher gebe es keine dezidierte Stelle, die dafür zuständig sei. So sei die Manipulation im Rahmen des vermeintlichen Videoanrufs von Vitali Klitschko bei Berlins damaliger regierender Bürgermeisterin, Franziska Giffey, beispielsweise durch Journalist:innen rekonstruiert worden – und nicht durch eine staatlich beauftragte Stelle.

---

<sup>34</sup> Verordnung, die Social-Media-Plattformen Haftungs- und Sicherheitsvorschriften vorgibt (siehe Kapitel 4.1)

## 5 Fazit: Zusammenfassung und Ausblick

### 5.1 Zusammenfassung

Insgesamt liefert die Auswertung der Experteninterviews in erster Linie einen Mehrwert durch die Bewertung und Priorisierung bereits im Forschungs- und Sachstand beschriebener Technologien und Probleme. Aus der Darstellung des Forschungsstandes und den Experteninterviews ergibt sich zunächst ein aktuelles Bild der derzeitigen Möglichkeiten, audiovisuelles Material zu manipulieren, wie auch der derzeit existierenden und genutzten Verifikationsmöglichkeiten. Sowohl für klassische Manipulationsformen wie auch zur Erkennung synthetischer Medien gibt es zahlreiche technische Ansätze, die bei der Verifizierung unterstützen können. Allerdings setzen sie größtenteils ein technisches Verständnis der dahinterstehenden Technologien voraus und müssen entsprechend der mutmaßlichen Manipulationsform ausgewählt werden. Synthetische Medien stellen zwar die potenziell perfektste Manipulationsform dar, spielen für den Redaktionsalltag jedoch bisher kaum eine Rolle. Als größte Herausforderung wird derzeit die Verbreitung dekontextualisierter Inhalte gesehen. Eine rein technische Lösung, die in der Lage ist alle audiovisuellen Manipulationen zu erkennen, sieht die Mehrheit der Interviewten nicht. Umso wichtiger sei genügend Zeit für die Durchführung semantischer Verfahren, eine kontextorientierte Recherche und den Abgleich mit anderen Quellen. Pixel- und formatbasierte Methoden hingegen sind für nutzergeneriertes Material durch die Veränderungen und den Informationsverlust im Rahmen der Uploadprozesse in der Praxis häufig weniger geeignet.

Insbesondere der Faktor Zeit sei in Redaktionen häufig ein Problem und stehe dem Druck, Inhalte sehr schnell aufzubereiten, konträr gegenüber. Gerade deshalb sei der Ausbau dezidierter Investigativ- bzw. Verifikationseinheiten in Redaktionen wichtig. Zudem könnte eine erweiterte Zusammenarbeit zwischen Redaktionen, Recherchenetzwerken und IT-Forensiker:innen hilfreich sein. Die Dokumentarabteilung des SPIEGEL zeigt zudem, dass auch eine Lösung sein kann, die Problematik audiovisueller Manipulationen zu minimieren, indem man im Redaktionsalltag fast ausschließlich auf zuverlässige Quellen zurückgreift. Für den zukünftigen Umgang mit audiovisuellen Ma-

nipulationen sei sowohl deren staatliche Regulierung, als auch die Stärkung der Medienkompetenz wichtig, aber auch eine Überprüfung der Arbeitsweise in Redaktionen. Teilweise wird das Risiko durch eine veränderte Manipulationslandschaft zukünftig alles in Frage stellen zu können, als größere Gefahr für Gesellschaft und Demokratie angesehen als die Manipulationen selbst.

### 5.2 Fazit

Klar ist, dass Journalismus gerade durch die Verunsicherung der Gesellschaft aufgrund einer zunehmenden Menge an Desinformation und immer authentischer wirkenden audiovisuellen Manipulationen wichtig ist, um zuverlässige Informationen bereitzustellen. Damit der Journalismus als Instanz nicht an Bedeutung verliert, müssen Redaktionen (auch weiter) authentische Inhalte bereitstellen und die Authentizität ihrer Inhalte garantieren können. Aber inwiefern ist und bleibt es nun möglich, die Echtheit audiovisuellen Materials in Redaktionen sicherzustellen?

Nach überwiegender Meinung der Interviewten ist es schon heute und auch zukünftig nicht möglich, allein auf technologischer Ebene sicherzustellen, dass audiovisuelles Material authentisch ist. Trotzdem werden technische Hilfsmittel – manuelle wie möglicherweise auch automatisierte Detektionssysteme – nötig sein, um Verifizierungsprozesse zu unterstützen, angefangen bei dem Prozess einer inversen Bildersuche.

Schlussendlich wird allerdings nicht die Authentizität des Materials zu beweisen sein, sondern die Plausibilität der enthaltenen Informationen. Damit einhergehend sind zurzeit nicht fehlende technische Hilfsmittel das Problem, sondern vielmehr genügend Zeit für die Beschäftigung mit der Thematik und einzelne Recherchen. Um auch in Zukunft die Echtheit audiovisuellen Materials sicherstellen zu können, müssen Redaktionen ihren Mitarbeiter:innen also genügend Zeit zur Verfügung stellen, zur tatsächlichen Recherche, aber auch um die technologische Entwicklung im Bereich der Manipulations- und Verifizierungsmöglichkeiten konstant zu verfolgen. In schwierigen Fällen kann eine Zusammenarbeit zwischen Redaktionen, Recherchenetzwerken und IT-Forensiker:innen helfen, wie sie bereits heute teilweise stattfindet. Ein möglicher Weg könnte sein, dass gar nicht mehr versucht wird, mit der Verbreitung von Inhalten im Internet Schritt zu halten, sondern ausschließlich über die Zuverlässigkeit und Einord-

nung der Informationen und absoluter Transparenz in Bezug auf die Quellenlage ein redaktioneller Mehrwert geschaffen wird.

### **5.3 Weitere mögliche Forschungsfelder**

Im Rahmen der Arbeit wurden deutlich mehr Redaktionen und Recherchenetzwerke, öffentlich-rechtliche Redaktionen genauso wie private, angeschrieben, als sich schlussendlich zu einem Interview bereiterklärt haben. Eine breitere Erhebung zum Umgang mit audiovisuellem Material in Redaktionen wäre jedoch nötig, um tatsächlich verallgemeinerbare Aussagen treffen zu können. Außerdem würde es sich anbieten, die Interviewpartner:innen gegenseitig mit ihren Aussagen, eventuell im Rahmen einer Gruppendiskussion, zu konfrontieren, um vermeintliche Widersprüche aufzulösen und Erklärungen zu finden, warum bspw. die vom IT-Forensiker Dr. Sascha Zmudzinski angesprochene kommerzielle Software in Redaktionen keine Anwendung findet.

Auch ein Forschungsprojekt, das sich ausschließlich mit den aktiven Methoden der Multimedia-Forensik beschäftigt und den von Adobe initiierten C2PA-Standard und damit verbundene Möglichkeiten für den Journalismus analysiert, könnte ein nächster Schritt sein.

Darüber hinaus muss in jedem Fall die tatsächliche Wirkung audiovisueller Manipulationen auf Rezipient:innen untersucht werden: Es mangelt derzeit an formalen und quantitativen Untersuchungen der wahrnehmungsbezogenen und psychologischen Faktoren, die einer Täuschung zugrunde liegen (vgl. Lyu 2022, 327).

## Literaturverzeichnis

- Adair, Bill; Stencel, Mark (2020): How We Identify Fact-Checkers - Duke Reporters' Lab. Online unter [reporterslab.org/how-we-identify-fact-checkers/](https://reporterslab.org/how-we-identify-fact-checkers/), zuletzt geprüft am 20.01.23.
- Adobe (04.05.23): Adobe Unveils Firefly, a Family of new Creative Generative AI. Online unter [news.adobe.com/news/news-details/2023/Adobe-Unveils-Firefly-a-Family-of-new-Creative-Generative-AI/default.aspx](https://news.adobe.com/news/news-details/2023/Adobe-Unveils-Firefly-a-Family-of-new-Creative-Generative-AI/default.aspx), zuletzt geprüft am 04.05.23.
- AFP (2023): Über uns. Online unter [www.afp.com/de/die-agentur/ueber-afp/ueber-uns](https://www.afp.com/de/die-agentur/ueber-afp/ueber-uns), zuletzt geprüft am 12.07.23.
- Apostolidis, Evlampios; Apostolidis, Konstatinos; Patras, Ioannis; Mezaris, Vasileios (2019): Video Fragmentation and Reverse Search on the Web. In: Vasileios Mezaris, Lyndon J. B. Nixon, Symeon Papadopoulos, Denis Teyssou und Lyndon Nixon (Hg.): Video Verification in the Fake News Era. Cham: Springer International Publishing, S. 53–90.
- Baig, Rachel (21.03.23): Wie wir bei der DW Fake News überprüfen. Deutsche Welle. Online unter [p.dw.com/p/4Og4X](https://p.dw.com/p/4Og4X), zuletzt geprüft am 18.04.23.
- Barni, Mauro; Li, Wenjie; Tondi, Benedetta; Zhang, Bowen (2022): Adversarial Examples in Image Forensics. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): Multimedia Forensics. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), S. 435–466.
- Bartlitz, Christine (2022): „Alles hat sich verdichtet“ – Kriegsbilder aus der Ukraine. Ein Interview mit Michael Pfister und Andreas Prost aus der Bildredaktion von „Zeit Online“. Visual History. Online unter [visual-history.de/2022/08/01/bartlitz-pfister-prost-alles-hatsich-verdichtet-kriegsbilder-aus-der-ukraine/](https://visual-history.de/2022/08/01/bartlitz-pfister-prost-alles-hatsich-verdichtet-kriegsbilder-aus-der-ukraine/), zuletzt geprüft am 01.08.22.
- Bayerische Akademie der Wissenschaften (2020): Das Bild in den Medien: Bildmanipulationen und Visualisierungen im Journalismus. Wie Redaktionen die Möglichkeiten der digitalen Bildbearbeitung nutzen: Prof. Dr. Christoph Neuberger (FU Berlin) im Gespräch mit Prof. Dr. Klaus Meier. FAKE UND FAKT IM BILD - Online-Gesprächsreihe über Macht und Ohnmacht der Bilder. Online unter [faktizitaet.badw.de/veranstaltungen/fake-und-fakt-im-bild.html](https://faktizitaet.badw.de/veranstaltungen/fake-und-fakt-im-bild.html), zuletzt geprüft am 20.01.22.
- Bendel, Oliver (2020): Definition: Was ist "Deepfake"? Gabler Wirtschaftslexikon. Online unter [wirtschaftslexikon.gabler.de/definition/deepfake-120960](https://wirtschaftslexikon.gabler.de/definition/deepfake-120960), zuletzt geprüft am 12.07.23.
- Bentele, Günter; Brosius, Hans-Bernd; Jarren, Otfried (Hg.) (2013): Lexikon Kommunikations- und Medienwissenschaft. Wiesbaden: Springer VS (Studienbücher zur Kommunikations- und Medienwissenschaft).
- Bischoff, Manon (2022a): Interview mit Forrest Rogers: "Fakten sind immer das erste Opfer eines Kriegs". In: *Spektrum der Wissenschaft* (7), S. 71–73.

- Bischoff, Manon (2022b): Open-Source-Intelligence. Auf digitaler Suprensuche. In: *Spektrum der Wissenschaft* (7), S. 62–70.
- BR24 (2023): #Faktenfuchs: Unser Team. Online unter [www.br.de/nachrichten/deutschland-welt/faktenfuchs-unser-team,SRAF9ZX](http://www.br.de/nachrichten/deutschland-welt/faktenfuchs-unser-team,SRAF9ZX), zuletzt geprüft am 21.04.23.
- BSI: Deepfakes - Gefahren und Gegenmaßnahmen. Online unter [www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes\\_node.html](http://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html), zuletzt geprüft am 03.04.23.
- BSI: Was sind Aufgaben des BSI? Online unter [www.bsi.bund.de/DE/Home/home\\_node.html](http://www.bsi.bund.de/DE/Home/home_node.html), zuletzt geprüft am 03.04.23.
- Büllesbach, Alfred (2008): Digitale Bildmanipulation und Ethik. Aktuelle Tendenzen im Fotojournalismus. In: Elke Grittmann, Irene Neverla und Ilona Ammann (Hg.): *Global, lokal, digital. Fotojournalismus heute*. Köln: Herbert von Halem Verlag, S. 108–136.
- Burkell, Jacquelyn; Gosse, Chandell (2019): Nothing new here: Emphasizing the social and cultural context of deepfakes. In: *First Monday* 24 (12).
- C2PA: Overview. Online unter [c2pa.org/](http://c2pa.org/), zuletzt geprüft am 09.07.23.
- C2PA: Specifications. Online unter [c2pa.org/specifications/specifications/1.3/index.html](http://c2pa.org/specifications/specifications/1.3/index.html), zuletzt geprüft am 09.07.23.
- CAI: Content Authenticity Initiative. Online unter [contentauthenticity.org/](http://contentauthenticity.org/), zuletzt geprüft am 09.07.23.
- Chung, Chin-Wan (Hg.) (2014): *Proceedings of the 23rd International Conference on World Wide Web*. Seoul, Korea, 7.4. - 11.04.2014. ACM Special Interest Group on Hypertext, Hypermedia, and Web. New York: ACM (ACM Digital Library).
- Cozzolino, Davide; Poggi, Giovanni; Verdoliva, Luisa (2022): Data-Driven Digital Integrity Verification. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 281-311.
- Diakopoulos, Nicholas (2018): Reporting in a Machine Reality: Deepfakes, misinformation, and what journalists can do about them. In: *Columbia Journalism Review*. Online unter [www.cjr.org/tow\\_center/reporting-machine-reality-deepfakes-diakopoulos-journalism.php](http://www.cjr.org/tow_center/reporting-machine-reality-deepfakes-diakopoulos-journalism.php), zuletzt geprüft am 22.07.20.
- Digitale Politikvermittlung. Chancen und Risiken interaktiver Medien (2014). Hg. v. Mike Friedrichsen und Roland A. Kohn. Wiesbaden: Springer VS.
- Diplaris, Sotiris; Papadopoulos, Symeon; Kompatsiaris, Ioannis; Heise, Nicolaus; Spangenberg, Jochen; Newman, Nic; Hacid, Hakim (2012): "Making sense of it all". In: Alain Mille, Fabien Gandon, Jacques Misselis, Michael Rabinovich und Steffen Staab (Hg.): *21st World Wide Web Conference 2012, April 16 - 20, 2012, Lyon*. New York: ACM, S. 1241.

- Duden (2020): Fake News | Rechtschreibung, Bedeutung, Definition, Herkunft. Online unter [www.duden.de/rechtschreibung/Fake\\_News](http://www.duden.de/rechtschreibung/Fake_News), zuletzt geprüft am 17.7.20.
- DW (28.09.22): DW setzt mit vera.ai den Kampf gegen Desinformation fort. Online unter [p.dw.com/p/4HNF1](http://p.dw.com/p/4HNF1), zuletzt geprüft am 04.05.23.
- DW Innovation: Research and Cooperations - About. Online unter [innovation.dw.com/about](http://innovation.dw.com/about), zuletzt geprüft am 04.05.23.
- Epp, Alexander; Höfner, Roman (24.05.22): Satellitenfotos, Geheimbefehle und Haftlingslisten: So haben wir die Xinjiang Police Files überprüft. SPIEGEL Ausland. Online unter [www.spiegel.de/ausland/xinjiang-police-files-so-haben-wir-das-datenleck-ueberprueft-a-0dc1d2e6-2771-48ab-af72-ded4aae6970b](http://www.spiegel.de/ausland/xinjiang-police-files-so-haben-wir-das-datenleck-ueberprueft-a-0dc1d2e6-2771-48ab-af72-ded4aae6970b), zuletzt geprüft am 21.04.23.
- First Draft: About. Online unter [firstdraftnews.org/about/](http://firstdraftnews.org/about/), zuletzt geprüft am 13.04.23.
- Forster, Klaus (2003): Rezeption von Bildmanipulationen. In: Thomas Knieper und Marion G. Müller (Hg.): Authentizität und Inszenierung von Bilderwelten. Köln: Herbert von Halem Verlag, S. 66-101.
- Fraunhofer SIT: DisCo - Disinformation in Conflicts. Online unter [www.sit.fraunhofer.de/de/disco/](http://www.sit.fraunhofer.de/de/disco/), zuletzt geprüft am 01.05.23.
- Fricke, Rolf; Thomsen, Jan (2019): Video Verification in the Newsroom. In: Vasileios Mezaris, Lyndon J. B. Nixon, Symeon Papadopoulos, Denis Teyssou und Lyndon Nixon (Hg.): Video Verification in the Fake News Era. Cham: Springer International Publishing, S. 301–335.
- Gerling, Winfried (2022): Bildforensik im Journalismus - Kontexte und Methoden. In: Elke Grittmann und Felix Koltermann (Hg.): Fotojournalismus im Umbruch. Hybrid, multimedial, prekär. Köln: Herbert von Halem Verlag, S. 296–317.
- Gläser, Jochen; Laudel, Grit (2010): Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. Wiesbaden: VS Verlag (Lehrbuch).
- Goertz, Lutz (1995): Wie interaktiv sind Medien? Auf dem Weg zu einer Definition von Interaktivität. In: *Rundfunk und Fernsehen* 43. Jg., S. 477–493.
- Graßl, Michael; Schützeneder, Jonas; Meier, Klaus (2022): Künstliche Intelligenz als Assistenz. In: *Journalistik* 5 (1), S. 3–27.
- Grimm, Imre (16.03.22): Bilder zum Krieg in der Ukraine: Was ist echt, was Fälschung? RedaktionsNetzwerk Deutschland (RND). Online unter [www.rnd.de/medien/bilder-zum-krieg-in-der-ukraine-was-ist-echt-was-faelschung-ARJX75LDS5GGFKQRNH7EXIDPXU.html](http://www.rnd.de/medien/bilder-zum-krieg-in-der-ukraine-was-ist-echt-was-faelschung-ARJX75LDS5GGFKQRNH7EXIDPXU.html), zuletzt geprüft am 17.04.23.
- Grittmann, Elke; Koltermann, Felix (Hg.) (2022): Fotojournalismus im Umbruch. Hybrid, multimedial, prekär. Köln: Herbert von Halem Verlag.
- Grittmann, Elke; Neverla, Irene; Ammann, Ilona (Hg.) (2008): Global, lokal, digital. Fotojournalismus heute. Köln: Herbert von Halem Verlag.



- Hajj-Ahmad, Adi; Wong, Chau-Wai; Choi, Jisoo; Wu, Min (2022): Power Signature for Multimedia Forensics. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): Multimedia Forensics. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 235-280.
- Halvani, Oren; Freifrau Heereman von Zuydtwyck, Wendy; Herfert, Michael; Kreutzer, Michael; Liu, Huajian; Fhom, Hervais-Clemence Simo et al. (2020): Automatisierte Erkennung von Desinformationen. In: Martin Steinebach, Katarina Bader, Lars Rinsdorf, Nicole Krämer und Alexander Roßnagel (Hg.): Desinformation aufdecken und bekämpfen. Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG (Schriften zum Medien- und Informationsrecht, 45), S. 101–148.
- Hartwig, Katrin; Reuter, Christian (2021): Fake News technisch begegnen – Detektions- und Behandlungsansätze zur Unterstützung von NutzerInnen. In: Peter Klimczak und Thomas Zoglauer (Hg.): Wahrheit und Fake im postfaktisch-digitalen Zeitalter. Distinktionen in den Geistes- und IT-Wissenschaften. Wiesbaden, Heidelberg: Springer Vieweg (ars digitalis), S. 133–149.
- Heine, Torben (2022): Hybride Kriegsführung: So kämpfen Faktenchecker gegen Desinformationen zum Ukraine-Krieg. MEEDIA. Online unter [www.meedia.de/medien/hybride-kriegsfuehrung-so-kaempfen-faktenchecker-gegen-desinformationen-zum-ukraine-krieg-efc1d63812c4de71f3165ef21f517903](http://www.meedia.de/medien/hybride-kriegsfuehrung-so-kaempfen-faktenchecker-gegen-desinformationen-zum-ukraine-krieg-efc1d63812c4de71f3165ef21f517903), zuletzt geprüft am 21.04.23.
- Hendrix, Justin; Morozoff, Dan (2022): Media Forensics in the Age of Disinformation. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): Multimedia Forensics. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), S. 7–40.
- Higgins, Eliot (2021): Digitale Jäger. Ein Insiderbericht aus dem Recherchenetzwerk Bellingcat: Quadriga Berlin.
- Hooffacker, Gabriele; Kenntemich, Wolfgang; Kulisch, Uwe (Hg.) (2018): Die neue Öffentlichkeit. Wie Bots, Bürger und Big Data den Journalismus verändern. Hochschule für Technik, Wirtschaft und Kultur Leipzig. Wiesbaden, Heidelberg: Springer VS.
- ICFJ (2019): The State of Technology in Global Newsrooms 2019. Online unter [www.icfj.org/sites/default/files/2019-10/2019%20Final%20Report.pdf](http://www.icfj.org/sites/default/files/2019-10/2019%20Final%20Report.pdf), zuletzt geprüft am 30.06.23.
- IKSŽ (Hg.) (2022): Changing Hearts and Minds? The Role of Fact-Checking. Panel-Beiträge im Rahmen der CEDMO-Konferenz "Europe Tackles Information Chaos" #CIC2022. Online unter [youtu.be/PQr556aXxII](https://youtu.be/PQr556aXxII), zuletzt geprüft am 27.12.22.
- Institute of Electrical and Electronics Engineers (2020): IEEE International Conference on Multimedia and Expo workshops (ICMEW). London, UK, 06.-10.07.20.
- Johnson, Bobbie (26.03.19): Deepfakes are solvable—but don't forget that "shallow-fakes" are already pervasive. In: *MIT Technology Review*, 26.03.19. Online unter [www.technologyreview.com/2019/03/25/136460/deepfakes-shallowfakes-human-rights/](http://www.technologyreview.com/2019/03/25/136460/deepfakes-shallowfakes-human-rights/), zuletzt geprüft am 08.04.23.

- Khan, Sohail Ahmed; Sheikhi, Ghazaal; Opdahl, Andreas L.; Rabbi, Fazle; Stoppel, Sergej; Trattner, Christoph; Dang-Nguyen, Duc-Tien (2023): Visual User-Generated Content Verification in Journalism: An Overview. In: *IEEE Access* 11, S. 6748–6769.
- Kietzmann, Jan; Lee, Linda W.; McCarthy, Ian P.; Kietzmann, Tim C. (2020): Deepfakes: Trick or treat? In: *Business Horizons* 63 (2), S. 135–146.
- Kirchner, Matthias (2022): Sensor Fingerprints: Camera Identification and Beyond. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), S. 65–88.
- Kirste, Moritz; Schürholz, Markus (2019): Einleitung: Entwicklungswege zur KI. In: Volker Wittpahl (Hg.): *Künstliche Intelligenz. Technologie | Anwendung | Gesellschaft*. Berlin, Heidelberg: Springer Vieweg, S. 21–35.
- Kleinstauber, Hans J. (2013): Audiovisuelle Medien. In: Günter Bentele, Hans-Bernd Brosius und Otfried Jarren (Hg.): *Lexikon Kommunikations- und Medienwissenschaft*. Wiesbaden: Springer VS (Studienbücher zur Kommunikations- und Medienwissenschaft), S. 21–22.
- Klimczak, Peter; Zoglauer, Thomas (Hg.) (2021): *Wahrheit und Fake im postfaktisch-digitalen Zeitalter. Distinktionen in den Geistes- und IT-Wissenschaften*. Wiesbaden, Heidelberg: Springer Vieweg (ars digitalis).
- Knieper, Thomas; Müller, Marion G. (Hg.) (2003): *Authentizität und Inszenierung von Bilderwelten*. Köln: Herbert von Halem Verlag.
- Köberer, Nina; Sehr, Marc (2017): Krisen- und Kriegsberichterstattung in der digitalisierten Welt: Eine medienethische Reflexion auf die Verifizierungspraxis von (Bewegt-)Bildern. In: Ingrid Stapf, Marlis Prinzing und Alexander Filipovic (Hg.): *Gesellschaft ohne Diskurs? Digitaler Wandel und Journalismus aus medienethischer Perspektive*. Unter Mitarbeit von Ingrid Stapf. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG (Kommunikations- und Medienethik, 5), S. 165–180.
- Köhler, Tanja (Hg.) (2020): *Fake News, Framing, Fact-Checking: Nachrichten im digitalen Zeitalter. Ein Handbuch*. Bielefeld: transcript Verlag.
- Koltermann, Felix; Grittmann, Elke (2022): Fotojournalismus im Umbruch: Hybrid, multimedial, prekär. In: Elke Grittmann und Felix Koltermann (Hg.): *Fotojournalismus im Umbruch. Hybrid, multimedial, prekär*. Köln: Herbert von Halem Verlag, S. 9–37.
- Koolwijk, Jürgen von (1974): Die Befragungsmethode. In: Jürgen von Koolwijk und Maria Wieken-Mayser (Hg.): *Bd. 4: Erhebungsmethoden: Die Befragung*: R. Oldenbourg Verlag München Wien (Techniken der empirischen Sozialforschung, 4), S. 9–23.
- Koolwijk, Jürgen von; Wieken-Mayser, Maria (Hg.) (1974): *Bd. 4: Erhebungsmethoden: Die Befragung*: R. Oldenbourg Verlag München Wien (Techniken der empirischen Sozialforschung, 4).
- Krake, Martin (2020): Tiefe Fälschung. In: *extradienst (APA)* (4-5), S. 102–112. Online unter [www.wiso-](http://www.wiso-)

- net.de/document/EXTR\_\_dd03540460f363e74bf976a0f500f3c42f925a09, zuletzt geprüft am 18.03.23.
- Kuckartz, Udo (2018): *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz.
- Lamnek, Siegfried; Krell, Claudia (2016): *Qualitative Sozialforschung*. Weinheim: Beltz.
- Long, Chengjiang; Basharat, Arslan; Hoogs, Anthony (2022): Video Frame Deletion and Duplication. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 333-362.
- Loosen, Wiebke; Solbach, Paul (2020): Künstliche Intelligenz im Journalismus? Was bedeutet Automatisierung für journalistisches Arbeiten? In: Tanja Köhler (Hg.): *Fake News, Framing, Fact-Checking: Nachrichten im digitalen Zeitalter. Ein Handbuch*. Bielefeld: transcript Verlag, S. 177–203.
- Lyu, Siwei (2022): DeepFake Detection. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 313-331.
- Maas, Julia; Ratering, Jörn (2022): Von Computerspielen und Doppelgängern. Die Verifizierung als Basis redaktioneller Entscheidungen im Ukraine-Krieg. In: *Communicatio Socialis (ComSoc)* 55 (4), S. 499–507.
- Maier, Anatol; Riess, Christian (2023): Zuverlässigkeit von KI in der Multimedia-Forensik. In: *Datenschutz und Datensicherheit - DuD* 47 (4), S. 215–219.
- Mandelli, Sara; Bonettini, Nicolò; Bestagini, Paolo (2022): Source Camera Model Identification. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 133-173.
- Mantzaris, Alexios (2015): Will verification kill fact-checking? Online unter [www.poynter.org/fact-checking/2015/will-verification-kill-fact-checking/](http://www.poynter.org/fact-checking/2015/will-verification-kill-fact-checking/), zuletzt geprüft am 11.04.23.
- Marinova, Zlatina; Spangenberg, Jochen; Teyssou, Denis; Papadopoulos, Symeon; Sarris, Nikos; Alaphilippe, Alexandre; Bontcheva, Kalina (2020): Weverify: Wider and Enhanced Verification for You. Project Overview and Tools. In: *IEEE International Conference on Multimedia and Expo workshops (ICMEW)*. London, UK, 06.-10.07.20. Institute of Electrical and Electronics Engineers, S. 1–4.
- Markatopoulou, Foteini; Zampoglou, Markos; Apostolidis, Evlampios; Papadopoulos, Symeon; Mezaris, Vasileios; Patras, Ioannis; Kompatsiaris, Ioannis (2019): Finding Semantically Related Videos in Closed Collections. In: Vasileios Mezaris, Lyndon J. B. Nixon, Symeon Papadopoulos, Denis Teyssou und Lyndon Nixon (Hg.): *Video Verification in the Fake News Era*. Cham: Springer International Publishing, 127-159.
- Mayring, Philipp (2015): *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. 12. Aufl. Weinheim: Beltz (Beltz Pädagogik).

- McCloskey, Scott (2022): Computational Imaging. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 41-62.
- Mercier, Grégoire; Markatopoulou, Foteini; Cozien, Roger; Zampoglou, Markos; Apostolidis, Evlampios; Metsai, Alexandros I. et al. (2019): Detecting Manipulations in Video. In: Vasileios Mezaris, Lyndon J. B. Nixon, Symeon Papadopoulos, Denis Teyssou und Lyndon Nixon (Hg.): *Video Verification in the Fake News Era*. Cham: Springer International Publishing, 161-89.
- Mezaris, Vasileios; Nixon, Lyndon J. B.; Papadopoulos, Symeon; Teyssou, Denis; Nixon, Lyndon (Hg.) (2019): *Video Verification in the Fake News Era*. Cham: Springer International Publishing.
- Mille, Alain; Gandon, Fabien; Misselis, Jacques; Rabinovich, Michael; Staab, Steffen (Hg.) (2012): *21st World Wide Web Conference 2012, April 16 - 20, 2012, Lyon*. New York: ACM.
- Misoch, Sabina (2019): *Qualitative Interviews*. Berlin, Boston: De Gruyter Oldenbourg.
- Möller, Christian (Hg.) (2019): *Was tun gegen Fake News und Hate Speech? Strategien und Ausblicke*. Fachhochschule Kiel.
- Neves, João C.; Tolosana, Ruben; Vera-Rodriguez, Ruben; Lopes, Vasco; Proença, Hugo; Fierrez, Julian (2022): GAN Fingerprints in Face Image Synthesis. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), S. 175–204.
- Nightingale, Sophie J.; Farid, Hany (2022): AI-synthesized faces are indistinguishable from real faces and more trustworthy. In: *Proceedings of the National Academy of Sciences of the United States of America* 119 (8).
- Nowotny, Ralf (2022): Kein echtes Video aus der Ukraine, sondern Hubschrauber-Szene aus "Arma 3". In: *Mimikama*, 04.04.2022. Online unter [www.mimikama.org/hubschrauber-szene-aus-arma-3/](http://www.mimikama.org/hubschrauber-szene-aus-arma-3/), zuletzt geprüft am 01.07.2023.
- NPP (2020): The News Provenance Project. Online unter [www.newsprovenanceproject.com/](http://www.newsprovenanceproject.com/), zuletzt geprüft am 05.04.23.
- Obermann, Gerrit (03.03.22): Wie ntv und RTL Fake-Videos entlarven, 03.03.22. Online unter [www.n-tv.de/mediathek/videos/politik/Wie-ntv-und-RTL-Fake-Videos-entlarven-article23170566.html](http://www.n-tv.de/mediathek/videos/politik/Wie-ntv-und-RTL-Fake-Videos-entlarven-article23170566.html), zuletzt geprüft am 21.04.23.
- Papadopoulou, Olga; Zampoglou, Markos; Papadopoulos, Symeon; Kompatsiaris, Ioannis (2019): Verification of Web Videos Through Analysis of Their Online Context. In: Vasileios Mezaris, Lyndon J. B. Nixon, Symeon Papadopoulos, Denis Teyssou und Lyndon Nixon (Hg.): *Video Verification in the Fake News Era*. Cham: Springer International Publishing, S. 191–222.
- Paus-Hasebrink, Ingrid et al. (2006): *Einführung in die audiovisuelle Kommunikation*. München, Wien: Oldenbourg (Lehr- und Handbücher der Kommunikationswissenschaft).

- Peters, Martha Rebekka (2020): Symbolfotos im Technik- und Wissenschaftsjournalismus. Intention und Wirkung. Bachelorarbeit. Hochschule Bonn/Rhein-Sieg, St. Augustin.
- Piva, Alessandro; Iuliani, Massimo (2022): Integrity Verification Through File Container Analysis. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 363-387.
- Pörzgen, Gemma (19.08.22): Auf zum Faktencheck! Ein Überblick über redaktionelle Faktenchecker in der Medienlandschaft. ReVue. Online unter [www.revue.org/beitrag/bild-in-zeit-gemma-poerzgen-auf-zum-faktencheck](http://www.revue.org/beitrag/bild-in-zeit-gemma-poerzgen-auf-zum-faktencheck), zuletzt geprüft am 17.04.23.
- PresseClub München (Hg.) (07.04.22): BJV-Podiumsdiskussion: Fotos und Videos: wahr oder falsch? – Bilder im Röntgenblick. Podiumsdiskussion mit Dennis Amour, Janina Lückhoff, Dr.-Ing. Christian Riess, Lea Weinmann. Online unter [youtu.be/XfeoOCc0bg](https://youtu.be/XfeoOCc0bg), zuletzt geprüft am 07.06.23.
- Primbs, Stefan (2018): Was Verifikations-Einheiten tun. In: Gabriele Hooffacker, Wolfgang Kenntemich und Uwe Kulisch (Hg.): *Die neue Öffentlichkeit. Wie Bots, Bürger und Big Data den Journalismus verändern*. Wiesbaden, Heidelberg: Springer VS, S. 115–124.
- Ramesh, Aditya; Pavlov, Mikhail; Goh, Gabriel; Gray, Scott (05.01.21): DALL·E: Creating images from text. OpenAI. Online unter [openai.com/research/dall-e#fn-11](https://openai.com/research/dall-e#fn-11), zuletzt geprüft am 04.05.23.
- Ratering, Jörn (2022): Zwischen Geotags und Gesichtserkennung: KI in der Verifikation. In: *Communicatio Socialis (ComSoc)* 55 (3), S. 360–366.
- Reuters (11.01.21): Fact check: Donald Trump concession video not a ‘confirmed deepfake’. Online unter [www.reuters.com/article/uk-factcheck-trump-concession-video-deep/fact-check-donald-trump-concession-video-not-a-confirmed-deepfake-idUSKBN29G2NL](https://www.reuters.com/article/uk-factcheck-trump-concession-video-deep/fact-check-donald-trump-concession-video-not-a-confirmed-deepfake-idUSKBN29G2NL), zuletzt geprüft am 10.04.23.
- Reuters: Identifizierung und Auseinandersetzung mit manipulierten Medien. Ein Kurs der Nachrichtenagentur Reuters. Online unter [www.reuters.com/manipulatedmedia](https://www.reuters.com/manipulatedmedia), zuletzt geprüft am 07.04.23.
- Riess, Christian (2021): Möglichkeiten und Grenzen bei der Ermittlung von Wahrheit mit Maschinellem Lernen. In: Christian Schicha, Ingrid Stapf und Saskia Sell (Hg.): *Medien und Wahrheit. Medienethische Perspektiven auf Desinformation, Lügen und „Fake News“*. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG (Kommunikations- und Medienethik, 15), S. 343–357.
- Riess, Christian (2022): Physical Integrity. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 207-234.
- Sängerlaub, Alexander (2020): Im Zeitalter von Fake News. Warum sich der (Nachrichten-)Journalismus neu erfinden muss. In: Tanja Köhler (Hg.): *Fake News, Framing,*

- Fact-Checking: Nachrichten im digitalen Zeitalter. Ein Handbuch. Bielefeld: transcript Verlag, S. 99–117.
- Schade, Frederic (2019): Deep Fakes - Produktion und Erkennung von Maschinengenerierten Fake Videos. In: Christian Möller (Hg.): Was tun gegen Fake News und Hate Speech? Strategien und Ausblicke, S. 95–107.
- Schicha, Christian; Stapf, Ingrid; Sell, Saskia (Hg.) (2021): Medien und Wahrheit. Medienethische Perspektiven auf Desinformation, Lügen und „Fake News“. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG (Kommunikations- und Medienethik, 15).
- Schreier, Margrit (2014): Varianten qualitativer Inhaltsanalyse: Ein Wegweiser im Dickicht der Begrifflichkeiten. In: *Forum Qualitative Sozialforschung / Forum: Qualitative Social* 15 (1), Art. 18.
- Sell, Saskia; Oswald, Bernd (2021): Verifikation von Online-Inhalten im Journalismus. In: Christian Schicha, Ingrid Stapf und Saskia Sell (Hg.): Medien und Wahrheit. Medienethische Perspektiven auf Desinformation, Lügen und „Fake News“. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG (Kommunikations- und Medienethik, 15), S. 241–262.
- Sencar, Husrev Taha (2022): Source Camera Attribution from Videos. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), S. 89–115.
- Sencar, Husrev Taha; Verdoliva, Luisa; Memon, Nasir (Hg.) (2022): *Multimedia Forensics*. Singapore: Springer (Advances in Computer Vision and Pattern Recognition).
- Silverman, Craig; Landesanstalt für Medien NRW (Hg.) (2020): *Verification Handbook - Das Handbuch zur Überprüfung von Desinformation und Medien-Manipulation*. Düsseldorf.
- Spangenberg, Jochen: vera.ai - Project Summary. Online unter [www.veraai.eu/project-summary](http://www.veraai.eu/project-summary), zuletzt geprüft am 08.05.23.
- Spangenberg, Jochen (2014): Soziale Medien und journalistische Berichterstattung: Digitale Politikvermittlung. Chancen und Risiken interaktiver Medien. Hg. v. Mike Friedrichsen und Roland A. Kohn. Wiesbaden: Springer VS, S. 105–126.
- Spangenberg, Jochen; Heise, Nicolaus (2014): News from the crowd. In: Chin-Wan Chung (Hg.): *Proceedings of the 23rd International Conference on World Wide Web*. Seoul, Korea, 7.4. - 11.04.2014. ACM Special Interest Group on Hypertext, Hypermedia, and Web. New York: ACM (ACM Digital Library), S. 765–768.
- Spangenberg, Jochen; Teyssou, Denis (2019): Video Verification: Motivation and Requirements. In: Vasileios Mezaris, Lyndon J. B. Nixon, Symeon Papadopoulos, Denis Teyssou und Lyndon Nixon (Hg.): *Video Verification in the Fake News Era*. Cham: Springer International Publishing, S. 3–14.
- Stamm, Matthew C.; Zhao, Xinwei (2022): Anti-Forensic Attacks Using Generative Adversarial Networks. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.):

- Multimedia Forensics. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), 467-490.
- Stapf, Ingrid; Prinzing, Marlis; Filipovic, Alexander (Hg.) (2017): Gesellschaft ohne Diskurs? Digitaler Wandel und Journalismus aus medienethischer Perspektive. Unter Mitarbeit von Ingrid Stapf. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG (Kommunikations- und Medienethik, 5).
- Stark, Holger; Zimmermann, Fritz (20.02.23): Desinformationsfirma "Team Jorge": Sie manipulieren die Welt. In: *DIE ZEIT*, 20.02.23. Online unter [www.zeit.de/2023/08/desinformation-team-jorge-social-media-storykillers](http://www.zeit.de/2023/08/desinformation-team-jorge-social-media-storykillers), zuletzt geprüft am 30.06.23.
- Steinebach, Martin; Bader, Katarina; Rinsdorf, Lars; Krämer, Nicole; Roßnagel, Alexander (Hg.) (2020): Desinformation aufdecken und bekämpfen. Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität. Baden-Baden: Nomos Verlagsgesellschaft mbH & Co. KG (Schriften zum Medien- und Informationsrecht, 45).
- Stencel, Mark; Ryan, Erica; Luther, Joel (21.06.23): Misinformation spreads, but fact-checking has leveled off - Duke Reporters' Lab. Online unter [reporters-lab.org/misinformation-spreads-but-fact-checking-has-leveled-off/](http://reporters-lab.org/misinformation-spreads-but-fact-checking-has-leveled-off/), zuletzt geprüft am 30.06.23.
- Stern, Jenny (2020): Fact-Checking und Verifikation. Neue Formen des Faktenprüfens im Nachrichtenjournalismus. In: Tanja Köhler (Hg.): Fake News, Framing, Fact-Checking: Nachrichten im digitalen Zeitalter. Ein Handbuch. Bielefeld: transcript Verlag, S. 119–149.
- Sterz, Christoph (29.10.16): Falschmeldungen im Netz - Den Fakes auf der Spur. Deutschlandfunk. Online unter [www.deutschlandfunk.de/falschmeldungen-im-netz-den-fakes-auf-der-spur-100.html](http://www.deutschlandfunk.de/falschmeldungen-im-netz-den-fakes-auf-der-spur-100.html), zuletzt geprüft am 19.04.23.
- Taspinar, Samet; Memon, Nasir (2022): Camera Identification at Large Scale. In: Husrev Taha Sencar, Luisa Verdoliva und Nasir Memon (Hg.): Multimedia Forensics. Singapore: Springer (Advances in Computer Vision and Pattern Recognition), S. 117–132.
- Tedros Adhanom Ghebreyesus (2020): Munich Security Conference. Online unter [www.who.int/director-general/speeches/detail/munich-security-conference](http://www.who.int/director-general/speeches/detail/munich-security-conference), zuletzt geprüft am 29.05.23.
- Teyssou, Denis (2019): Applying Design Thinking Methodology: The InVID Verification Plugin. In: Vasileios Mezaris, Lyndon J. B. Nixon, Symeon Papadopoulos, Denis Teyssou und Lyndon Nixon (Hg.): Video Verification in the Fake News Era. Cham: Springer International Publishing, S. 263–279.
- Truepic (2023): Truepic Lens - Build your apps on content authenticity. Online unter [truepic.com/truepic-lens/](http://truepic.com/truepic-lens/), zuletzt geprüft am 09.07.23.
- Truly Media (09.06.2019): ZDF trusts Truly Media to tackle disinformation. Online unter [www.truly.media/zdf-trusts-truly-media-to-tackle-disinformation/](http://www.truly.media/zdf-trusts-truly-media-to-tackle-disinformation/), zuletzt geprüft am 01.07.23.

- Urbani, Shaydanay (22.09.22): Verifying Online Information. First Draft. Online unter [firstdraftnews.org/long-form-article/verifying-online-information/](https://firstdraftnews.org/long-form-article/verifying-online-information/), zuletzt geprüft am 13.04.23.
- VAUNET (03.02.22): Mediennutzung 2021: Audio- und audiovisuelle Mediennutzung steigt in Deutschland erstmals auf 10 Stunden pro Tag. Online unter [vau.net/presse-meldungen/mediennutzung-2021-audio-und-audiovisuelle-mediennutzung-steigt-in-deutschland-erstmals-auf-10-stunden-pro-tag/](https://vau.net/presse-meldungen/mediennutzung-2021-audio-und-audiovisuelle-mediennutzung-steigt-in-deutschland-erstmals-auf-10-stunden-pro-tag/), zuletzt geprüft am 30.05.23.
- Verdoliva, Luisa (2020): Media Forensics and DeepFakes: an overview. In: *IEEE Journal of selected topics in signal processing*, S. 910–932.
- Verhoeven, Tim (2020): Digitalisierung im Recruiting. Wie sich Recruiting durch künstliche Intelligenz, Algorithmen und Bots verändert. Wiesbaden: Springer Gabler.
- Weinmann, Lea (02.03.22): SZ-Transparenz-Blog: Wie prüft die SZ Bilder aus den sozialen Medien? Süddeutsche Zeitung. Online unter [www.sueddeutsche.de/projekte/artikel/politik/sz-transparenz-blog-wie-prueft-man-bilder-aus-sozialen-medien-e462455/](https://www.sueddeutsche.de/projekte/artikel/politik/sz-transparenz-blog-wie-prueft-man-bilder-aus-sozialen-medien-e462455/), zuletzt geprüft am 18.04.23.
- Welchering, Peter (2020): Journalistische Praxis: Digitale Recherche. Verifikation und Fact Checking. Wiesbaden, Heidelberg: Springer VS.
- WeVerify: Verification plugin - WeVerify. Online unter [weverify.eu/verification-plugin/](https://weverify.eu/verification-plugin/), zuletzt geprüft am 02.04.23.
- Wiesel, Matthias (28.09.21): So funktioniert der Faktencheck bei ZDFheute, 28.09.21. Online unter [www.zdf.de/nachrichten/heute-sendungen/faktencheck-recherche-zdf-video-100.html](https://www.zdf.de/nachrichten/heute-sendungen/faktencheck-recherche-zdf-video-100.html), zuletzt geprüft am 19.04.23.
- Wittpahl, Volker (Hg.) (2019): Künstliche Intelligenz. Technologie | Anwendung | Gesellschaft. Institut für Innovation und Technik. Berlin, Heidelberg: Springer Vieweg.
- YouTube (2023): YouTube by the numbers. Online unter [blog.youtube/press/](https://blog.youtube/press/), zuletzt geprüft am 21.03.23.
- ZDF (15.02.23): So läuft der Faktencheck bei ZDFheute, 15.02.23. Online unter [www.zdf.de/nachrichten/heute-sendungen/faktencheck-nachrichten-video-100.html](https://www.zdf.de/nachrichten/heute-sendungen/faktencheck-nachrichten-video-100.html), zuletzt geprüft am 19.04.23.



**INDEX**

|   |    |
|---|----|
| Aktive Forensik .....   | 44 |
| Analyse von ENF-Signalen (Electrical Network Frequency) .....       | 38 |
| Analyse von Kompressionsspuren (Error Level Analysis) .....         | 36 |
| Andrea Sauerbier .....  | 59 |
| Audiovisuelle Medien .....  | 4  |
| Authentizität .....   | 12 |
| Bearbeitete Medien .....  | 13 |
| Bellingcat .....  | 58 |
| Bild .....  | 32 |
| Bildrauschen .....  | 32 |
| Bundesamt für Sicherheit in der Informationstechnik (BSI) .....     | 57 |
| Chronolokalisierung .....   | 30 |
| Coalition for Content Provenance and Authenticity (C2PA) .....      | 45 |
| Computer Generated Imagery (CGI) .....                              | 16 |
| Content Authenticity Initiative (CAI) .....                         | 45 |
| Context Aggregation And Analysis (CAA) Service .....                | 27 |
| Deep Learning (DL) .....  | 17 |
| Dekontextualisierte Medien .....                                    | 13 |
| Desinformation .....  | 8  |
| Digital Service Act .....   | 75 |
| Digitale Signaturen .....   | 79 |
| Digitale Wasserzeichen .....  | 79 |
| Dr. Christian Riess .....   | 58 |
| Dr. Dominique Dresen .....  | 57 |
| Dr. Jutta Jahnel .....  | 59 |
| Dr. Sascha Zmudzinski .....   | 59 |
| Einzelbildanalyse .....   | 37 |
| Expert:innen .....  | 55 |
| Fact-Checking .....   | 22 |
| Formatbasierte Verfahren .....                                      | 37 |
| Fraunhofer-Institut für Sichere Informationstechnologie (SIT) ..... | 59 |
| Geolokalisierung .....  | 30 |
| Graswurzel-Journalismus .....                                       | 48 |
| Inszenierte Medien .....  | 13 |
| Inverse Bildersuche (Image reverse search) .....                    | 26 |
| InVID .....   | 23 |
| Jochen Spangenberg .....  | 60 |
| Johanna Wild .....  | 58 |

## INDEX

---

|   |    |
|---|----|
| JPEG-AI.....  | 74 |
| Karlsruher Institut für Technologie (KIT).....                | 59 |
| KI-Verordnung auf EU-Ebene.....                               | 75 |
| Kollaborativer Journalismus .....                             | 48 |
| Künstliche Intelligenz (KI) .....                             | 17 |
| Liars Dividend .....  | 20 |
| Manipulation audiovisuellen Materials .....                   | 12 |
| Maschinelles Lernen (ML).....                                 | 17 |
| Multimedia-Forensik .....                                     | 22 |
| Open Source Intelligence (OSINT).....                         | 48 |
| Photo Response Non-Uniformity (PRNU) .....                    | 33 |
| Physikbasierte Verfahren.....                                 | 30 |
| Pixelbasierte Verfahren .....                                 | 32 |
| Projekt DisCo – Disinformation in Conflicts .....             | 82 |
| Shallowfakes / Cheapfakes .....                               | 13 |
| Synthetische Medien / Deepfakes .....                         | 15 |
| Text-to-Image- bzw. Text-to-Video-Generatoren (T2I, T2V)..... | 15 |
| Verifikation .....  | 22 |

## ANHANG

### IN DER GEDRUCKTEN / VERÖFFENTLICHTEN AUSGABE

#### Das empirische Erhebungsinstrument: Die Interviewleitfäden für

- Dr. Dominique Dresen (01) 117
- Johanna Wild (02) 119
- Dr. Christian Riess (03) 120
- Dr. Jutta Jahnel (04) 122
- Andrea Sauerbier (05) 124
- Dr. Sascha Zmudzinski (06) 125
- Jochen Spangenberg (07) 129

Das empirische Auswertungsinstrument: Das Kategorienhandbuch 129

Erklärung zur Master-Thesis 133

### AUF DEM DATENTRÄGER befindet / befinden sich zusätzlich

- eine digitale Fassung der Arbeit
- die codierten Transkripte der Interviews im Wortlaut

## Interviewleitfaden Dr. Dominique Dresen (01)<sup>35</sup>

**Funktion der Person:** Referent beim BSI im Referat Bewertungsverfahren für eID-Technologien in der Digitalisierung, Experte für synthetische Medien

**Themen:** Synthetische Medien + aktive Forensik

|   |  |
|---|--|
| <p><b>① Technischer Stand + Entwicklung (der Detektion) von Deepfakes</b></p>                 | <p><b>Inwiefern ist das BSI mit Deepfakes konfrontiert?</b><br/>→ Verifiziert das BSI selbst Material?</p> <p><b>Wie einfach und wie realistisch sind Deepfakes inzwischen zu erstellen?</b><br/>→ Kann man inzwischen mehr als das Gesicht austauschen?<br/>→ Wie ist es mit Stirn, Haaren, Ohren, wie mit Brillenträgern?<br/>→ Sind ganze Körper schon möglich?<br/>→ Gibt es inzwischen Software, die Gesicht und Stimme in einem manipulieren kann?<br/>→ Wie gut sind Live-Manipulationen schon möglich?<br/>→ Deepfake-Apps für die Hosentasche, für jeden mit wenigen Klicks einsetzbar = realistisch?</p> <p><b>In der Literatur wird die Erkennung von Deepfakes als einfach beschrieben, gleichzeitig wird bei der state-of-the-art Lösung eine durchschnittliche Genauigkeit von 65% bei unbekanntem Material (Lyu 2022, 316) angegeben, widerspricht sich das nicht?</b></p> <p><b>Glauben Sie, dass Deepfakes irgendwann so gut sein werden, dass sie nur noch mithilfe KI-basierter Methoden zu erkennen sind?</b> → Wann?</p> <p><b>Kann man Deepfakes und vollständig computergenerierte Inhalte (Text-to-Image, Text-to-Video) mit den gleichen Techniken detektieren?</b><br/>→ Sollte man zwischen beiden Erscheinungsformen unterscheiden?</p> <p><b>Wird die Entwicklung von Text-to-Video ähnlich aussehen, wie sie bei Text-to-Image zu sehen ist?</b></p> |
| <p><b>② Auswirkung von audiovisuellen Manipulationen / Deepfakes für den Journalismus</b></p> | <p><b>Glauben Sie, dass die Technologie hinter Deepfakes längerfristig die Methode mit den überzeugendsten Manipulationsergebnissen sein wird?</b></p> <p><b>Sehen Sie Deepfakes als das größte Risiko im Bereich der audiovisuellen Manipulationen? Können Deepfakes für die Gesellschaft und insbesondere den Journalismus zum Problem werden?</b></p> <p><b>Welche Methoden zur Detektion von Deepfakes können in Redaktionen eingesetzt werden?</b><br/>→ nennenswerte Softwarelösungen / Projekte?</p> <p><b>Können / müssen Social-Media-Plattformen mehr tun, um die Verifikation von audiovisuellem Material zu erleichtern?</b><br/>→ Wenn ja: Was können sie tun?</p>  |

<sup>35</sup> Nicht alle Fragen aus den folgenden Leitfäden wurden in den Interviews auch tatsächlich gestellt. Manche Fragen erübrigten sich bereits durch vorherige Antworten der Interviewten, andere fielen aus Zeitgründen weg. Als besonders wichtig erachtete Fragen wurden **fett** markiert.

|  |   |
|--|---|
| <p>③ <b>Aktive Forensik</b></p>                  | <p><b>Was sind die größten Hürden auf dem Weg, die Integrität audiovisuellen Materials durch die Einbettung von digitalen Signaturen sicherzustellen?</b></p> <p><b>Ist die Verwendung einer Blockchain in dem Kontext sinnvoll oder mehr ein Buzzword?</b> → evtl. Bezug zu News Provenance Project oder Apps wie Truepic Lens / Capture, falls bekannt</p>  |
| <p>④ <b>Allgemeiner Blick in die Zukunft</b></p> | <p><b>Denken Sie, dass aktuell jede Manipulation erkannt werden kann?</b></p> <p><b>Glauben Sie, dass der Moment kommen wird, wo nicht mehr sicher festzustellen ist, ob eine Bild- oder Tonaufnahme echt oder manipuliert ist?</b></p> <p><b>Glauben Sie, dass der technologische Abstand zwischen Manipulations- und Detektionsmöglichkeiten immer in etwa so bleiben wird, wie er jetzt ist oder verschiebt er sich?</b></p> |

Gibt es von Ihrer Seite noch Dinge, die wir im Interview nicht ausführlich genug angesprochen haben, die Sie aber gerne noch hinzufügen möchten?

## Interviewleitfaden Johanna Wild (02)

**Funktion der Person:** Ehem. Journalistin u. a. für die SZ, den BR, ARTE und weitere Medien, arbeitet inzwischen für das OSINT-Recherchenetzwerk Bellingcat und beschäftigt sich dort mit der Entwicklung von Technologien + Tools für digitale Ermittlungen.

**Themen:** Verifikationsmethoden/-prozesse bei Bellingcat + in Redaktionen + Einschätzung der Gefährlichkeit audiovisueller Manipulationen

|  |   |
|--|---|
| <p>① <b>Bellingcat + Die Verifizierung audiovisueller Inhalte bei Bellingcat</b></p>               | <p>Inwiefern ist Bellingcat mit manipuliertem Audiomaterial, Bildmaterial und Videomaterial konfrontiert?</p> <p><b>Wie würden Sie das Verhältnis oder die Zusammenarbeit zwischen Bellingcat und herkömmlichen journalistischen Redaktionen beschreiben?</b></p> <p><b>Die meisten anschaulichen OSINT-Verifizierungsinstrumente wie die Geolokalisierung anhand von Satellitenbildern oder ein Abgleich mit den Wetterdaten zum vermeintlichen Aufnahmezeitpunkt arbeiten ja mit dem Bildinhalt. Inwiefern arbeitet Bellingcat auch mit Werkzeugen, die Artefakte auf Pixelebene analysieren?</b></p> <p><b>Was bekommen Sie von den Umfragen unter Journalist:innen oder Workshops mit, wo Redaktionen bei der OSINT-Recherche stehen (im Vergleich zu Bellingcat)?</b></p> <p>→ <b>Was werden sich dort für Hilfsmittel gewünscht?</b></p> <p>→ <b>Wo liegen die Hürden für die Nutzung existierender Möglichkeiten?</b></p> <p>Welche Manipulationen sind für Bellingcat am schwersten zu verifizieren?</p> <p>Gibt es abgesehen von den vielen einzelnen OSINT-Tools technische Hilfsmittel, die Sie nutzen, die besonders herausstechen, z. B. weil sie viele einzelne Tools miteinander verbinden?</p> <p>Welche zusätzlichen technischen Hilfsmittel würden Sie sich wünschen?</p> |
| <p>② <b>(Einschätzung der) Auswirkungen audiovisueller Manipulationen für den Journalismus</b></p> | <p>Sind Deepfakes für Sie inzwischen ein reales Problem?</p> <p>Was sehen Sie als größtes Risiko im Bereich der audiovisuellen Manipulationen für den Journalismus?</p> <p>Glauben Sie, dass digitale, fälschungssichere Signaturen, die im Moment der Aufnahme einem Bild- oder Tondokument hinzugefügt werden und dessen Authentizität belegen, das Potenzial haben, einen manuellen Verifizierungsprozess irgendwann überflüssig zu machen?</p> <p>Inwiefern können/müssen Social-Media-Plattformen mehr tun, um die Verifikation audiovisuellen Materials zu erleichtern? Was können sie tun?</p> <p><b>Denken Sie, dass aktuell jede Manipulation erkannt werden kann?</b></p> <p><b>Glauben Sie, dass es immer möglich sein wird, die Echtheit oder Manipulation von audiovisuellem Material zu beweisen?</b></p>   |

Gibt es von Ihrer Seite noch Dinge, die wir im Interview nicht ausführlich genug angesprochen haben, die Sie aber gerne noch hinzufügen möchten?

## Interviewleitfaden Dr. Christian Riess (03)

**Funktion der Person:** Gruppenleiter für Multimediasticherheit am Lehrstuhl für IT-Sicherheitsinfrastrukturen der Friedrich-Alexander-Universität Erlangen-Nürnberg

**Themen:** Technische Verifizierungsmöglichkeiten + Praxis und Möglichkeiten für Redaktionen

|   |   |
|---|---|
| <p>① <b>Multimedia-Forensik allgemein + Einschätzungen zu den Potenzialen der Multimedia-Forensik</b></p> | <p>Inwiefern beschäftigt sich die Uni Erlangen-Nürnberg mit audiovisuellen Manipulationen?</p> <p><b>Wie definiert sich für Sie Multimedia-Forensik? Also zählt beispielsweise die Bilder-Rückwärtssuche für Sie auch dazu?</b></p> <p><b>Es gibt ja unfassbar viele versch. Ansätze, audiovisuelles Material zu verifizieren. Denken Sie, dass aktuell jede Manipulation erkannt werden kann?</b></p> <p><b>Glauben Sie, dass der Moment kommen wird, wo nicht mehr sicher festzustellen ist, ob eine Bild- oder Tonaufnahme echt oder manipuliert ist?</b></p> <p><b>Welche technologischen Entwicklungen auf Manipulations- wie auf Verifikationsseite haben Ihrer Meinung nach das größte Potenzial und warum?</b></p> <p>→ Deepfakes angesprochen?</p> <p>→ Bisher größte Errungenschaft für die Verifizierung?</p>  |
| <p>② <b>Details + aktive Forensik</b></p>   | <p>Glauben Sie, dass die Entwicklung von Text-to-Image- oder Text-to-Video-Generatoren den Stellenwert und die Verwendung von Bildmaterial in der Gesellschaft allgemein, aber auch im Journalismus verändern wird?</p> <p>Glauben Sie, dass physikbasierte Methoden der Multimedia-Forensik trotz zunehmender Bildverarbeitungsverfahren in Smartphones und Computergenerierungen Bestand haben werden?</p> <p><b>Werden manuelle Analysen in Zukunft überhaupt noch eine Rolle spielen oder werden KI-basierte oder sogar vollständig automat. Detektionsmethoden ihnen den Rang ablaufen?</b></p> <p><b>Glauben Sie, dass Deepfakes irgendwann so gut sein werden, dass sie nur noch mithilfe KI-basierter Methoden zu erkennen sind? → Wann?</b></p> <p>→ Fehleranfälligkeit automatisierter Systeme angesprochen?</p> <p>→ Arbeiten DL-Systeme auf Pixelebene?</p> <p><b>Für wie realistisch halten Sie es, dass die Einbettung fälschungssicherer digitaler Signaturen in Aufnahmen so weit Verbreitung findet, dass die Signaturen zum Kriterium in Verifizierungsprozessen werden können?</b></p> |
| <p>③ <b>Verifizierung in Redaktionen</b></p>  | <p>Was für einen Eindruck haben Sie von den aktuellen Verifizierungsprozessen in Redaktionen?</p> <p>→ Inwiefern wird bisher das technische Potenzial für die Verifizierung audiovisueller Inhalte in Redaktionen ausgereizt? Wo liegen die Hürden?</p> <p>Welche Detektionsmethoden können in Redaktionen eingesetzt werden?</p> <p>Was muss dafür getan werden?</p>   |

|   |  |
|---|--|
|   | → nennenswerte Softwarelösungen / Projekte?  |
| <b>④ Notwendiges Handeln + Blick in die Zukunft</b> | <p>Können / müssen Social-Media-Plattformen mehr tun, um die Verifikation von audiovisuellem Material zu erleichtern? Was können sie tun?</p> <p>Glauben Sie, dass der technologische Abstand zwischen Manipulations- und Detektionsmöglichkeiten immer in etwa so bleiben wird, wie er jetzt ist oder verschiebt er sich?</p> |

Gibt es von Ihrer Seite noch Dinge, die wir im Interview nicht ausführlich genug angesprochen haben, die Sie aber gerne noch hinzufügen möchten?



## Interviewleitfaden Dr. Jutta Jahnel (04)

**Funktion der Person:** Wissenschaftliche Mitarbeiterin am Institut für Technikfolgenabschätzung und Systemanalyse (ITAS) des Karlsruher Instituts für Technologie (KIT)

**Themen:** Gefahren durch synthetische Medien für die Gesellschaft und insbesondere die Medienlandschaft – Was kann und muss getan werden?

|  |   |
|--|---|
| <p>① <b>Einschätzung von Deepfakes allgemein + Einfluss auf die Gesellschaft</b></p> | <p><b>Inwiefern hat das KIT mit Deepfakes zu tun?</b></p> <p>Zu welchem Zweck werden die meisten Deepfakes bisher erstellt und mit welchem Zweck werden sie verbreitet?</p> <p><b>Seit wann haben Deepfakes tatsächlich einen Stellenwert in der gesellschaftlichen Meinungsbildung?</b></p> <p><b>Sehen Sie Deepfakes als ein größeres Risiko im Bereich der audiovisuellen Manipulationen als andere Manipulationsformen?</b><br/>→ <b>Worin sehen Sie Risiken für die Gesellschaft und den Journalismus?</b></p> <p><b>Ändern Text-to-Image-Generatoren bzw. inzwischen sogar Text-to-Video-Generatoren Ihre Einschätzung, was den Einfluss von Deepfakes auf die Gesellschaft betrifft?</b><br/>→ <b>Sollte man zwischen klassischen Deepfakes und vollständig computer-generierten Inhalten unterscheiden?</b></p> <p><b>Wie real ist Ihrer Meinung nach die Gefahr, dass Medienschaffende durch Deepfakes getäuscht werden?</b><br/>→ <b>Deepfakes selbst größere Gefahr oder dass durch die Existenz von guten Manipulationen alles an audiovisuellem Material in Zweifel gezogen werden könnte?</b></p> |
| <p>② <b>Rechtlicher Rahmen / Was muss getan werden?</b></p>                          | <p><b>Wann ist ein Deepfake aktuell rechtswidrig?</b></p> <p><b>Ist die Kennzeichnung von Deepfakes schon jetzt Pflicht oder kommt das erst – vermutlich – mit dem AI-Act?</b></p> <p><b>Was sind für Sie alles notwendige Bestandteile einer rechtlichen Regulierung von Deepfakes? → Wie weit ist man da?</b></p> <p><b>Können / müssen Social-Media-Plattformen mehr tun, um eine Unterscheidung zwischen manipuliertem audiovisuellem Material und Authentischem zu erleichtern?</b><br/>→ <b>Was können sie tun?</b></p> <p><b>Wo gibt es Ihrer Meinung nach am meisten zu tun, bei den technischen Verifizierungsinstrumenten, der (staatlichen) Regulierung oder der Medienkompetenz der Menschen?</b></p> <p>Sie sind ja gerade in Wien. Sehen Sie Unterschiede im Umgang mit audiovisuellen Manipulationen zwischen Deutschland und Österreich?</p>  |
| <p>③ <b>Allgemeine Einschätzungen + Blick in die Zukunft</b></p>                     | <p>Glauben Sie, dass sich die technische Entwicklung bei Bild- und Tonmaterial positiv oder negativ auf die Glaubwürdigkeit von Kommunikation auswirken wird?</p>   |

|  |  |
|--|--|
|  | <p><b>Denken Sie, dass aktuell jede Manipulation erkannt werden kann?</b></p> <p><b>Glauben Sie, dass Deepfakes irgendwann so gut sein werden, dass sie nur noch mithilfe KI-basierter Methoden zu erkennen sind? → Wann?</b></p> <p><b>Glauben Sie, dass der Moment kommen wird, wo nicht mehr sicher festzustellen ist, ob eine Bild- oder Tonaufnahme echt oder manipuliert ist?</b></p> <p><b>Halten Sie es für realistisch, dass die Einbettung fälschungssicherer digitaler Signaturen bzw. Wasserzeichen in Aufnahmen oder auch computergenerierten Inhalten so weit Verbreitung findet, dass die Signaturen zu einem Kriterium in Verifizierungsprozessen werden können?</b></p> |
|--|--|

Gibt es von Ihrer Seite noch Dinge, die wir im Interview nicht ausführlich genug angesprochen haben, die Sie aber gerne noch hinzufügen möchten?

## Interviewleitfaden Andrea Sauerbier (05)

**Funktion der Person:** Dokumentarin bei der Zeitschrift SPIEGEL, bis 2022 Teil der dann aufgelösten Abteilung Bilddokumentation

**Themen:** Verifikationsmethoden/-prozesse beim SPIEGEL + Einschätzung der Gefährlichkeit audiovisueller Manipulationen

|  |   |
|--|---|
| <p><b>① Einführung Andrea Sauerbier + Verteilung der Zuständigkeiten beim SPIEGEL</b></p>                                      | <p>Was haben Sie mit der Verifizierung von Bildmaterial beim SPIEGEL zu tun?</p> <p>Wie teilen sich die Zuständigkeiten für die Verifizierung von audiovisuellem Material beim SPIEGEL auf und wie hat sich das in den letzten Jahren verändert?</p>  |
| <p><b>② Die Verifizierung von Inhalten beim SPIEGEL: Prozess, Prüfkriterien + Aufgabenbereich der Dokumentar-Abteilung</b></p> | <p>Wie läuft so ein Verifikationsprozess für das Bildmaterial, die Bildunterschrift und inzwischen dann auch den Textinhalt eines Beitrags bei Ihnen ab?</p> <p>Nach welchen Kriterien werden die Bilder und Bildunterschriften bei Ihnen geprüft?</p> <p><b>Haben sich Ablauf und Kriterien in den letzten Jahren verändert?</b></p> <p><b>Stoßen Sie bei der Verifikation von Bildinhalten häufig auf manipuliertes oder sogar künstlich generiertes Material?</b></p> <p>2017 ist in Artikeln noch von 70 Dokumentar:innen die Rede, dann von 60, 2022 von 50 ... wurde da nur unterschiedlich gezählt oder wurde die Dokumentarabteilung tatsächlich verkleinert?</p> <p>Warum wird zwischen der Verifizierung von Inhalten für die Print-Aufgabe des SPIEGELS und SPIEGEL Online unterschieden?</p> <p>→ Wie kommt es, dass die Spiegel+Artikel von Textseite verifiziert werden, von Bildseite her aber nicht?</p> <p>→ Glauben Sie, dass die Leser zwischen der Sorgfalt hinter den Printausgaben des SPIEGELS und der Onlineinhalte unterscheiden kann?</p> <p>Haben Sie den Eindruck, dass es sinnvoll ist, den Textinhalt einer Geschichte von der gleichen Person verifizieren zu lassen, die das zugehörige Bildmaterial verifiziert?</p> |
| <p><b>④ Einschätzung + Blick in die Zukunft</b></p>  | <p>Denken Sie, dass aktuell jede Manipulation erkannt werden kann?</p> <p>Sehen Sie die zunehmenden technischen Möglichkeiten audiovisuelles Material zu manipulieren oder komplett synthetisch zu generieren als Problem für Redaktionen?</p> <p>Glauben Sie, dass es immer möglich sein wird, die Echtheit oder Manipulation von audiovisuellem Material sicherzustellen?</p>   |

Gibt es von Ihrer Seite noch Dinge, die wir im Interview nicht ausführlich genug angesprochen haben, die Sie aber gerne noch hinzufügen möchten?

## Interviewleitfaden Dr. Sascha Zmudzinski (06)

**Funktion der Person:** Wissenschaftlicher Mitarbeiter des Forschungsbereichs Multimedia-Sicherheit und IT-Forensik am Fraunhofer-Institut SIT sowie Referent für Schulungen von IT-Forensiker:innen, Ermittler:innen und Journalist:innen

**Themen:** theoretische technische Möglichkeiten + praktische Lösungen für die Detektion audiovisueller Manipulationen, kurzer Umriss des Projekts DisCo

|   |   |
|---|---|
| <p>① <b>Multimedia-Forensik allgemein + Verifizierung durch das Fraunhofer SIT + Einschätzungen des technologischen Hintergrundes</b></p> | <p>Inwiefern beschäftigt sich das Fraunhofer SIT mit audiovisuellen Manipulationen?</p> <p><b>Wie definiert sich für Sie Multimedia-Forensik? Also zählt beispielsweise die Bilder-Rückwärtssuche für Sie auch dazu?</b></p> <p><b>Sie forschen ja nicht nur, sondern verifizieren auch selbst audiovisuelles Material, u. a. für Redaktionen. Wie gehen Sie dabei vor, auf welche Methoden greifen Sie zurück?</b></p> <p><b>Denken Sie, dass aktuell jede Manipulation erkannt werden kann?</b></p> <p><b>Welche technologischen Entwicklungen auf Manipulations- wie auf Verifikationsseite haben Ihrer Meinung nach das größte Potenzial und warum?</b></p> <p>→ Deepfakes angesprochen, schon reales Problem?<br/>→ Bisher größte Errungenschaft für die Verifizierung?</p> <p>Glauben Sie, dass Deepfakes irgendwann so gut sein werden, dass sie nur noch mithilfe KI-basierter Methoden zu erkennen sind? → Wann?</p> <p>Werden manuelle Analysen in Zukunft überhaupt noch eine Rolle spielen oder werden KI-basierte oder sogar vollständig automatisierte Detektionsmethoden ihnen den Rang ablaufen?</p> <p>(Wann) wird es möglich sein nicht nur Gesichter, sondern ganze Körper im Rahmen von Deepfakes zu synthetisieren?</p> <p>Kann man Deepfakes und vollständig computergenerierte Inhalte (Text-to-Image) mit den gleichen Techniken detektieren? → Sollte man zwischen beiden Erscheinungsformen unterscheiden?</p> <p>Wird die Entwicklung von Text-to-Video ähnlich schnell voran gehen, wie sie bei Text-to-Image zu sehen ist?</p> |
| <p>② <b>Aktive Forensik</b></p>   | <p><b>Was sind die größten Hürden auf dem Weg, die Integrität audiovisuellen Materials durch die Einbettung digitaler Signaturen sicherzustellen?</b><br/>→ <b>realistisch, dass digitale Signaturen so weit Verbreitung finden, dass sie zum Kriterium in Verifizierungsprozessen werden können?</b></p> <p>Wie kann die Technologie dahinter aussehen?<br/>→ Ist die Verwendung einer Blockchain in dem Kontext sinnvoll oder mehr ein Buzzword? Evtl. Bezug zu Sony, News Provenance Project oder Apps wie Truepic Lens / Capture, falls bekannt + zugrunde liegender Technologien</p>   |

|  |  |
|--|--|
| <p>③ <b>Verifizierung in Redaktionen: Erfahrungen aus Workshops / Zusammenarbeit mit Redaktionen</b></p> | <p><b>Was bekommen Sie von Journalist:innen z. B. durch Ihre Seminare mit, wo sie und Redaktionen allgemein bei der Verifizierung audiovisueller Inhalte stehen?</b><br/>         → <b>Wird bisher das technische Potenzial für die Verifizierung audiovisueller Inhalte in Redaktionen ausgereizt? Wo liegen die Hürden? Was braucht es?</b></p> <p><b>In welchen Fällen wird die Verifizierung von Material an externe Stellen wie das Fraunhofer SIT abgegeben?</b><br/>         → <b>Wird das Ihrer Meinung nach bei schwierigeren Fällen immer notwendig oder sinnvoll sein?</b></p> <p><b>Wie real ist Ihrer Meinung nach die Gefahr, dass Medienschaffende durch Deepfakes getäuscht werden?</b></p> <p>Gibt es schon erste Erkenntnisse aus dem Projekt DisCo, inwiefern manipuliertes Ton- und Bildmaterial in der Corona-Pandemie eine Rolle gespielt hat?</p> |
| <p>④ <b>Blick in die Zukunft</b></p>   | <p>Glauben Sie, dass der technologische Abstand zwischen Manipulations- und Detektionsmöglichkeiten immer in etwa so bleiben wird, wie er jetzt ist oder verschiebt er sich?</p> <p><b>Glauben Sie, dass der Moment kommen wird, wo nicht mehr sicher festzustellen ist, ob eine Bild- oder Tonaufnahme echt oder manipuliert ist? Wann?</b></p>   |

Gibt es von Ihrer Seite noch Dinge, die wir im Interview nicht ausführlich genug angesprochen haben, die Sie aber gerne noch hinzufügen möchten?

## Interviewleitfaden Jochen Spangenberg (07)

**Funktion der Person:** Stellv. Leiter der Abteilung *DW ReCo*, Dozent im Fachbereich Medien- und Kommunikationswissenschaften, Unterstützer der NGO *Lie Detectors*, im Beirat von *EDMO + GADMO*  
**Ergänzung:** spricht nicht für eine der Institutionen, sondern für sich als Person

**Themen:** Sozialwissenschaftliche Perspektive auf DW-Projekte, welche Hilfsmittel Journalist:innen benötigen, auf die Auswirkung der zunehmenden Manipulationsmöglichkeiten für die Bedeutung von audiovisuellem Material im Journalismus

|  |  |
|--|--|
| <p>① <b>Die DW ReCo und ihre Projekte</b></p>                                | <p>Inwiefern beschäftigt sich die Innovationsabteilung der DW mit audiovisuellen Manipulationen?</p> <p><b>Können Sie skizzieren, wie sich der Forschungsfokus der DW ReCo seit ihrem ersten Projekt in die Verifizierungsrichtung 2012 bis zu dem aktuellen Projekt vera.ai verändert hat?</b></p>  |
| <p>② <b>Der Umgang mit audiovisuellen Manipulationen im Journalismus</b></p> | <p>Inwiefern sind die DW oder auch andere Redaktionen, in die Sie durch Konferenzen oder einen persönlichen Austausch Einblick haben, überhaupt mit manipuliertem Bild-, Ton- oder Videomaterial konfrontiert?</p> <p><b>Vielleicht ist die nächste Frage von mir an Sie gerichtet ein bisschen blauäugig, aber möglicherweise haben Sie ja auch eine überraschende Antwort für mich: Was halten Sie aktuell für die wichtigsten Verifizierungsinstrumente in Bezug auf Bild- und Videomaterial für Journalist:innen?</b></p> <p>→ Wenn WeVerify + Truly Media: Hat man damit für den Moment alles gebündelt, was Journalist:innen für die Verifizierung audiovisueller Inhalte brauchen bzw. wo stoßen die Tools an ihre Grenzen?</p> <p>→ Was für technische Hilfsmittel fehlen noch?</p> <p><b>Sind es hauptsächlich technische Hilfsmittel, die fehlen, oder wo liegen aktuell insgesamt die größten Probleme für Redaktionen, audiovisuelles Material zu verifizieren?</b></p> <p><b>Wie technisch glauben Sie kann und muss der Verifizierungsprozess in Redaktionen überhaupt sein oder werden?</b></p> |
| <p>③ <b>Technische Möglichkeiten und Risiken</b></p>                         | <p><b>Im Rahmen des AI4media-Projektes habe ich auf der DW-Seite einen Artikel gelesen, der die Zukunft einer Journalistin im Jahr 2040 beschreibt – u. a. den täglichen Gebrauch KI-gestützter Verifizierungssysteme. Halten Sie das (CADI) für realistisch oder vielleicht sogar notwendig?</b></p> <p><b>Wieviel Arbeit, welchen Anteil des Verifizierungsprozesses können technische Hilfsmittel Journalist:innen überhaupt abnehmen, Ihrer Meinung nach?</b></p> <p>Was sehen Sie als größtes Risiko für den Journalismus im Bereich der audiovisuellen Manipulationen?</p> <p>→ Deepfakes angesprochen?</p> <p>Wie real ist Ihrer Meinung nach die Gefahr, dass Medienschaffende durch Deepfakes getäuscht werden?</p>   |

|   |  |
|---|--|
|   | <p>Sehen Sie manipulierte Inhalte selbst als größere Gefahr oder dass durch die Existenz von guten Manipulationen alles an audiovisuellem Material in Zweifel gezogen werden könnte?</p> <p>Glauben Sie, dass digitale Wasserzeichen, die die Authentizität von Material belegen, einen manuellen Verifizierungsprozess irgendwann überflüssig machen können? → Sehen Sie die Chance, dass sich so ein Mechanismus verbreitet?</p>   |
| <p><b>④ Akteure, notwendiges Handeln + Einschätzungen für heute und die Zukunft</b></p> | <p><b>Wie sehen Sie die Zusammenarbeit zwischen journalistischen Redaktionen, IT-Forensiker:innen und Recherchenetzwerken aktuell und wie könnte die Zusammenarbeit in Zukunft aussehen?</b><br/>→ Was bedeutet das für die Ausbildungsinhalte zukünftiger Journalist:innen?</p> <p><b>Wo gibt es Ihrer Meinung nach am meisten zu tun, bei den technischen Verifizierungsinstrumenten, der (staatlichen) Regulierung oder der Medienkompetenz der Menschen?</b><br/>→ Inwiefern können / müssen Social-Media-Plattformen mehr tun, um die Verifikation von audiovisuellem Material zu erleichtern? Was können sie tun?</p> <p><b>Was sind Ihrer Meinung nach die wichtigsten nationalen und internationalen Akteure im Gebiet der Forschung rund um die Verifizierung audiovisueller Inhalte?</b><br/>→ die wichtigsten Forschungsprojekte?</p> <p><b>Denken Sie, dass aktuell jede Manipulation erkannt werden kann?</b></p> <p><b>Glauben Sie, dass der Moment kommen wird, wo nicht mehr sicher festzustellen ist, ob eine Bild- oder Tonaufnahme echt oder manipuliert ist? → Wann?</b></p> <p><b>Glauben Sie, dass die Fortschritte im Bereich der audiovisuellen Manipulationen den Stellenwert und die Verwendung von Bild- und Tonmaterial in der Gesellschaft allgemein, aber auch im Journalismus verändern werden?</b></p> |

Gibt es von Ihrer Seite noch Dinge, die wir im Interview nicht ausführlich genug angesprochen haben, die Sie aber gerne noch hinzufügen möchten?

## Das Kategorienhandbuch

(=Mischung aus deduktiver und induktiver Kategorienbildung)

| Kategorie                           | Ggf. ergänzende Beschreibung der Kategorie + Die Kategorie wird codiert, wenn folgende Aspekte genannt werden...                          | Anwendungsbeispiel  |
|-------------------------------------|---|---|
| <b>1 Manipulationsmöglichkeiten</b> | Stand der Technik inkl. Aufwand zur Erstellung & Qualität auditiver / visueller Manipulationen  | „[...] Es reicht ja nicht nur ein Video zu fälschen, man muss ja irgendwie auch noch das ganze Drumherum... [...]“ (Int. Dresen, 708ff.)  |
| <b>Deepfakes</b>                    | s. o. aber ausschließlich bez. synthetischer Manipulationen inkl. deren rechtlicher Situation   | „Wenn man nachverarbeiten kann, dann kann man schon sehr, sehr gute Deepfakes erzeugen [...]“ (Int. Dresen, 219f.)  |
| <b>T2I / T2V</b>                    | s. o. aber ausschließlich bez. Text-to-Image / Text-to-Video inkl. Unterscheidung ggü. anderen synthetischen Medien                       | „Was jetzt natürlich wirklich umwälzend ist, das ist diese Verbindung zwischen diesen Sprach-Foundations-Modellen und der Bildverarbeitungskomponente [...]“ (Int. Riess, 196ff.)   |
| <b>2 Multimedia-Forensik</b>        | Systematisierung + Stand technologischer Möglichkeiten + Grundlegende Gedanken zur Verifikation audiovisueller Medien                     | „[...] Man kann diese ganzen Technologien irgendwie verschmelzen und letztendlich sprechen ja auch viele dann nur noch von Generative AI oder von synthetischen Medien. Deswegen, ich glaube... In Zukunft wird es immer schwieriger dann sogar Deepfakes alleine zu betrachten. Das wird irgendwie nach allen Seiten größer werden, schwammiger [...]“ (Int. Jähnel, 481ff.) |
| <b>System. Multimedia-Forensik</b>  | Definition / Systematisierung des Bereichs Multimedia-Forensik  | „[...] weil wir da eigentlich keine saubere Definition vielleicht haben. Ich würde es so sagen: Forensik an sich beschäftigt sich ja [...]“ (Int. Riess, 77ff.)   |
| <b>Passive Forensik</b>             | s. o. ausschl. zu Methoden der passiven Forensik  | „Wir haben immer noch, und die haben immer noch ihren wichtigen Stellenwert, traditionelle klassische Methoden, modellbasiert [...]“ (Int. Zmudzinski, 141ff.)  |
| <b>Det. v. Deepfakes</b>            | Möglichkeiten, synthetische Medien als solche zu erkennen insb. durch automatisierte Detektionssysteme (Stand der Technologie & Leistung) | „Seit 2019/2020 haben wir das Thema auf dem Schirm in unserem Referat.“ (Int. Dresen, 6f.)  |
| <b>Aktive Forensik</b>              | Sicherstellung der Integrität auditiver / visueller Inhalte präventiv durch digitale Wasserzeichen, Signaturen o. Ä.                      | /   |



|  |   |  |
|--|---|--|
| <b>Techn. Mögk. akt. F.</b>                        | Technische Umsetzungsmöglichkeiten aktiver Forensik   | „Einer der ersten großen Mechanismen aus dem Bereich der Mediensicherheit, mit dem sich die Arbeitsgruppe befasst hat und ich auch, sind tatsächlich digitale Wasserzeichen [...]“<br>(Int. Zmudzinski, 259ff.)  |
| <b>Hürden akt. F.</b>                              | Hürden / Probleme bez. des Einsatzes aktiver Forensik   | „Natürlich würde man jetzt eher sowas wie Kompression, dass das diesen Schlüssel unberührt lässt, gern zulassen. Das heißt [...]“ (Int. Dresen, 893ff.)  |
| <b>3 Praxis</b>                                    | /   | /  |
| <b>Vork. Manipulationen</b>                        | Real vorkommende Manipulationsformen: Welche sind schwierig zu detektieren? Sind Deepfakes für IT-Forensiker:innen oder Medienschaffende bereits ein reales Problem?<br><br>Aktueller Einsatz von Deepfakes + Stellenwert in der gesellschaftlichen Meinungsbildung<br><br>Studienergebnisse zu auditiven / visuellen Manipulationen im Rahmen der Covid-19-Pandemie<br><br>Auditive / visuelle Manipulationen in der akademischen Wissenschaftsliteratur | „Also ich muss dazu sagen, wir werden natürlich häufig gefragt: 'Ja, habt ihr denn viele manipulierte Bilder und man hört doch so viel davon...' Das spielt in unserem Arbeitsalltag ganz ehrlich gesagt - und ich mache das jetzt ja auch schon lange - so gut wie keine Rolle [...]“<br>(Int. Sauerbier, 218ff.) |
| <b>Praxis IT-Forensik</b>                          | Verifikationsprozesse von IT-Forensiker:innen inkl. Ablauf & genutzter Instrumente  | „Der allererste Schritt ist, in der Vorgehensweise mit dem Auftraggeber oder den Auftraggebern erstmal ganz genau zu besprechen, was überhaupt der Untersuchungsauftrag ist [...]“<br>(Int. Zmudzinski, 49ff.)   |
| <b>Praxis in Redaktionen / Recherchenetzwerken</b> | /   | /  |
| <b>Verifizierung in Re+Re</b>                      | Das Bewusstsein für die Existenz auditiver / visueller Manipulationen und der Umgang damit in Redaktionen / Recherchenetzwerken inkl. Verifikationsprozesse (Infrastruktur, Umfang, Verifikationskriterien und genutzte Hilfsmittel)  | „Es gibt natürlich ein paar Tools, die man immer nutzt, also zum Beispiel Satellitenbilder, wenn wir das jetzt als Tool bezeichnen wollen [...]“<br>(Int. Wild, 178ff.)  |
| <b>Probleme in Re+Re</b>                           | Probleme in Redaktionen / Recherchenetzwerken: Woran fehlt es? Wo liegen Hürden?  | „Aber wir finden unsere Materialien oft in den sozialen Netzwerken und die sozialen Netzwerke entfernen die Metadaten, bevor sie sie uns zur Verfügung stellen. Das heißt, damit können wir gar nicht arbeiten.“ (Int. Wild, 68ff.)  |

|  |   |  |
|--|---|--|
| <b>Zusammenarbeit (akt.)</b>           | Zusammenarbeit zwischen Redaktionen, Recherchenetzwerken und IT-Forensiker:innen  | „Also die kamen einfach zu uns dann in den Fällen, über die ich sprechen kann. Also das war tatsächlich [...]“ (Int. Zmudzinski, 494f.)  |
| <b>4 Detektierbarkeit AV-Medien</b>    | Einschätzungen zur schlussendlichen Detektierbarkeit audiovisueller Manipulationen  | /  |
| <b>Akt. Detektierbarkeit AV-Medien</b> | Einschätzungen zur schlussendl. Detektierbarkeit audiovisueller Medien zur Zeit inkl. Verifikationsmöglichkeiten, die in Redaktionen eingesetzt werden können   | „Also, ja, so in Jahrzehnten würde das wahrscheinlich // ein Problem sein. Jetzt haben wir natürlich so... Da sind wir ja noch gar nicht... Jetzt haben wir die ganzen Artefakte ja, an denen können wir es noch festmachen [...]“ (Int. Dresen, 585ff.) |
| <b>Zuk. Detektierbarkeit AV-Medien</b> | Einschätzungen zur schlussendl. Detektierbarkeit audiovisueller Medien in Zukunft inkl. Einschätzungen zu Veränderung des technologischen Abstandes zw. Manipulations- und Detektionsmöglichkeiten über die Zeit                            | „Also ich sehe jetzt keine physikalische Grenze, dass man eine perfekte Fälschung irgendwann herstellen kann.“ (Int. Dresen, 212f.)  |
| <b>Zuk. Relevanz pass. F.</b>          | Einschätzungen zur Entwicklung und zukünftigen Relevanz passiver multimedia-forensischer Methoden u.a. von manuellen / physikbasierten / automatisierten Methoden   | „Und deswegen denke ich eben auch, dass manuelle Analysen nach wie vor eine Rolle spielen. Man braucht sicherlich [...]“ (Int. Riess, 415ff.)  |
| <b>Zuk. Verbreitung akt. F.</b>        | Einschätzungen zur zukünftigen Relevanz & Verbreitung aktiver forensischer Methoden   | „Also kann man das vielleicht so in jedes Handy schon diese Funktion einführen [...]“ (Int. Wild, 261ff.)  |
| <b>Zuk. Zusammenarbeit</b>             | Einschätzungen zur zukünftigen Zusammenarbeit zw. Redaktionen und IT-Forensiker:innen: Wird eine Zusammenarbeit sinnvoll oder sogar notwendig sein?   | „Man kann natürlich super miteinander zusammenarbeiten. Also zum Beispiel [...]“ (Int. Spangenberg, 725f.)   |
| <b>Technisierung Re+Re</b>             | Einschätzungen zur zukünftigen Nutzung technischer Instrumente in Redaktionen / Recherchenetzwerken:<br>Wie technisch können / müssen Redaktionen arbeiten?<br>Wie technisch wird ein Verifizierungsprozess in Redaktionen aussehen können? | „Das heißt, der Algorithmus oder die Technik, wie auch immer, unterstützt und kann super, super tolle Hilfestellungen bieten, aber letztendlich die ultimative Entscheidung [...]“ (Int. Spangenberg, 501ff.)  |

|  |  |  |
|--|--|--|
| <b>5 Einfluss + Risiken</b>            | Einschätzungen zu Einfluss und Risiken audiovisueller Manipulationen auf/für Journalismus und Gesellschaft und deren Wahrnehmung, Einschätzungen zum Einfluss auf die Glaubwürdigkeit von Kommunikation inkl. Einschätzungen zum Einfluss von T2I / T2V auf den Stellenwert von Bildmaterial im Journalismus, zur potenziellen Gefahr, dass Medienschaffende durch Deepfakes getäuscht werden und ob Deepfakes selbst oder die ‚Liars Dividend‘ ein größeres Risiko darstellen | „Also das ist dann, was dann dadurch entsteht. Also dass auch authentisches Material, das eigentlich keine Fehler aufweist, schon angefochten werden kann [...]“ (Int. Dresen, 206ff.)                                   |
| <b>6 Handlungsfelder</b>               | Beurteilung möglicher notwendiger Handlungsfelder  | /  |
| <b>Weiterentwicklung techn. Mittel</b> | s. o. bez. der Notwendigkeit zusätzlicher technischer Hilfsmittel  | „Natürlich alles, was dabei hilft, Social Media zu durchsuchen, ist immer sehr sehr hilfreich.“ (Int. Wild, 183f.)   |
| <b>Forschung</b>                       | s. o. bez. notwendiger Forschungsinhalte   | „Wir versuchen die Forschung zu fördern. Also wir haben jetzt sozusagen ein Projekt formuliert, also was sozusagen... um Gegenmaßnahmen weiterzuentwickeln auf dem Stand der Forschung.“ (Int. Dresen, 74ff.)            |
| <b>Staatl. Stelle zur Det.</b>         | s. o. bez. staatlicher Institution zur Verifizierung politisch / gesellschaftlich relevanter Inhalte   | „Also es gibt dann schon Anfragen bei uns. Sozusagen auf dem kurzen Dienstweg, [...] Aber das ist nicht unsere Aufgabe eigentlich.“ (Int. Dresen, 143ff.)  |
| <b>Staatl. Reg.</b>                    | s. o. bez. staatlicher Regulierung   | „Klar, also die Technologie an sich würde ich jetzt nicht verbieten.“ (Int. Dresen, 1080f.)  |
| <b>S.-M.-Plattformen</b>               | s. o. bez. des Umgangs der Social-Media-Plattformen mit auditivem / visuellem Material   | „Wenn ich wüsste, dass es gute Detektoren gäbe, dann würde ich natürlich sagen, die müssen wir machen, die müssen wir durchpushen, dass das verpflichtend ist.“ (Int. Dresen, 867ff.)                                    |
| <b>Notw. Anpassung Re</b>              | s. o. bez. Anpassungen der redaktionellen Praxis, z. B. in Form der Ausbildung von Journalist:innen  | „Ich würde mir wünschen, dass es auch im Lokaljournalismus irgendwann so weit ist, dass es mehr solche Teams gibt. Weil [...]“ (Int. Wild, 158ff.)   |
| <b>Medienkompetenz</b>                 | s. o. bez. der Stärkung der Medienkompetenz von Rezipient:innen  | „Ich meine, was kann man dagegen machen? Da ist auch mit größtenteils halt die Prävention die erste Gegenmaßnahme. Und da kann man halt nur die Bevölkerung erstmal informieren, dass es das gibt.“ (Int. Dresen, 69ff.) |

## **Anhang: Erklärung zur Master-Thesis**

Ich versichere hiermit, die von mir vorgelegte Arbeit selbstständig verfasst zu haben. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Arbeiten anderer entnommen sind, habe ich als entnommen kenntlich gemacht. Sämtliche Quellen und Hilfsmittel, die ich für die Arbeit benutzt habe, sind angegeben. Die Arbeit hat mit gleichem Inhalt bzw. in wesentlichen Teilen noch keiner anderen Prüfungsbehörde vorgelegen.

Mir ist bewusst, dass sich die Hochschule vorbehält, meine Arbeit auf plagierte Inhalte hin zu überprüfen und dass das Auffinden von plagiierten Inhalten zur Nichtigkeit der Arbeit, zur Aberkennung des Abschlusses und zur Exmatrikulation führen kann.

---

Unterschrift – Sankt Augustin, 14.07.2023