

Maria Pawelec | Cora Bieß

Deepfakes

Technikfolgen und Regulierungsfragen aus ethischer
und sozialwissenschaftlicher Perspektive

Mit einer interaktiven Lehrinheit von Cora Bieß



Nomos

Kommunikations- und Medienethik

herausgegeben von

Alexander Filipović

Christian Schicha

Ingrid Stapf

Band 16

Maria Pawelec | Cora Bieß

Deepfakes

Technikfolgen und Regulierungsfragen aus ethischer
und sozialwissenschaftlicher Perspektive

Mit einer interaktiven Lehrinheit von Cora Bieß



Nomos



Onlineversion
Nomos eLibrary

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

ISBN 978-3-8487-8432-5 (Print)

ISBN 978-3-7489-2807-2 (ePDF)

Bis Band 4 erschienen bei Beltz Juventa, Weinheim.

1. Auflage 2021

© Nomos Verlagsgesellschaft, Baden-Baden 2021. Gesamtverantwortung für Druck und Herstellung bei der Nomos Verlagsgesellschaft mbH & Co. KG. Alle Rechte, auch die des Nachdrucks von Auszügen, der fotomechanischen Wiedergabe und der Übersetzung, vorbehalten. Gedruckt auf alterungsbeständigem Papier.

Geleitwort

Die Themen Vertrauen, Wahrheit und Wahrhaftigkeit begleiten seit den Anfängen der Presseberichterstattung die Debatten um Medienöffentlichkeiten und ihre Wirkungen. In diesem Zusammenhang gab es schon immer einen starken Wechselbezug zwischen der jeweiligen Medientechnik und den Inhalten der Kommunikation. Die Möglichkeiten der Medientechnik verändern die Form und die Inhalte, um die es in der gesellschaftlichen Verständigung geht.

Insbesondere der Stellenwert des *Visual Content*, also von Bildern, Videos, Piktogrammen usw. hat in der Internetkommunikation gegenüber dem geschriebenen Wort stark an Bedeutung gewonnen. Deepfakes sind ein Teil dieser visuellen Kommunikation und stellen in Frage, ob man seinen Augen noch trauen kann. Umso wichtiger ist eine kritische Auseinandersetzung mit den Gefahren und Potenzialen einer durch KI-manipulierten Bild-, Ton- und Videoproduktion.

Manipulation, Täuschung und Lüge können die demokratische Debatte nachhaltig beschädigen. In den medial vermittelten Öffentlichkeiten mit ihren vielfältigen Kommunikationsformen vom politischen Diskurs über Bildungsangebote bis zu den unterschiedlichsten Unterhaltungsangeboten ist Vertrauenswürdigkeit eine Grundbedingung, um überhaupt zur Teilhabe zu motivieren und konstruktiv zur Meinungsbildung beizutragen.

Das Thema Deepfakes bündelt die unterschiedlichen Aspekte und Erwartungen, die mit Kommunikation verbunden sind und verbindet sie mit den Kontroversen um Künstliche Intelligenz. Wie in einem Brennglas vereinigen Deepfakes zentrale Herausforderungen einer digitalen Gesellschaft. Welche Bilder und Videos sind echt? Welche Gerechtigkeitsfragen stellen sich? Welche synthetisch erzeugten Medieninhalte sind zu begrüßen? Kann und muss hier reguliert werden?

Für den Bereich der Deepfakes existieren international und national nur wenige Regulierungen. In Deutschland gibt es bislang keine spezifischen gesetzlichen Vorgaben, sondern es werden allgemeine Regelungen angewendet. Hierzu zählt u. a. das Netzwerkdurchsetzungsgesetz (NetzDG), das auch den Umgang mit durch den Einsatz von KI zustande gekommenen Manipulationen oder Desinformationen umfasst. In Hinsicht auf die Moderation der Inhalte Sozialer Medien gehen die Plattformen aber

Geleitwort

zumeist sogar strenger vor als die gesetzlichen Vorgaben und löschen Inhalte, die nicht vom NetzDG erfasst sind. Priorität haben hier die eigenen Hausregeln der Plattformbetreiber und die entsprechenden Ethikkodizes. Dies ist eine typische Konstellation für Regulierungsherausforderungen im Digitalisierungskontext im Zusammenspiel von ethischen und rechtlichen Normen, globalen Unternehmen und den individuell handelnden Nutzer:innen und Nutzern bei der Erstellung des *User Generated Content*.

Generell sind Verbote von Deepfakes, wie sie in China oder Südkorea teils durchgesetzt werden sollen, immer mit der Gefahr einer Einschränkung der Kunst-, Medien- und Meinungsäußerungsfreiheit verbunden. Darüber hinaus können sie auch die Entwicklung durchaus nützlicher synthetischer Bilder unterbinden und die Freiheit der Forschung behindern.

Das vorliegende Buch diskutiert diese und weitere kontroverse Fragen im Zusammenhang der Erstellung und Verbreitung von Deepfakes. Das Themenspektrum umfasst die Bedeutung von Deepfakes für politische Öffentlichkeiten wie auch die gezielte Anwendung von Deepfakes für die Schädigung von Frauen durch manipulierte pornografische Videos. Damit gibt es einen umfassenden Überblick zu der Frage, wie und was reguliert werden soll, und was aber auch nicht.

Die dem Band zugrundeliegende Studie wurde vom BMBF gefördert und ist eine Momentaufnahme in einem dynamischen Entwicklungs- und Governancebereich. Diese Momentaufnahme ist ethisch fundiert. Das bedeutet, dass Deepfakes in ihren unterschiedlichen Anwendungsbereichen kontextualisiert werden. Sie werden im Hinblick auf die Werte, Leitbilder und Weltanschauungen analysiert, die in ihre Entwicklung, Gestaltung und Nutzung einfließen und es wird gezeigt, dass die Möglichkeiten der Beeinflussung der Rezipient:innen und der öffentlichen Meinungsbildung zentrale Themen sind. Durch diese ethische Fundierung reicht die Momentaufnahme weit über den Moment hinaus, der sich auf einen bestimmten Stand der Technik und auf einen bestimmten Stand der Regulierungsanstrengungen bezieht. Die hier erarbeiteten ethischen Fragen zur Technikbewertung, zur Ambivalenz dieser Technik und zu Möglichkeiten und Wegen der Regulierung bleiben über den jetzigen Stand der Technik hinaus wertvoll und wichtig: Wenn Techniken wie Deepfakes gesellschaftsveränderndes Potenzial haben, dann sind die hier analysierten Bereiche (von den Kontexten der Entwicklung bis hin zu Selbst- und Opferschutz) in der hier gezeigten Differenziertheit grundlegend für eine möglichst gute und sinnvolle Techniknutzung. Die Momentaufnahme also erarbeitet die Grundlagen für die Bewertung von Deepfakes auch in

weiteren technischen und regulatorischen Stufen und zukünftigen gesellschaftlichen Konstellationen.

Darüber hinaus tut die Studie, was sie fordert: Sie schließt ab mit einer interaktiven Lehreinheit, in der deutlich wird, auf welche Weise zur Reflexion über Deepfakes angeleitet und Medienkompetenz unterstützt werden kann.

Wir wünschen der Studie viele interessierte Leser:innen.

Regina Ammicht Quinn & Jessica Heesen

Danksagung

Unser besonderer Dank gilt dem Internationalen Zentrum für Ethik in den Wissenschaften (IZEW) an der Universität Tübingen für die Ermöglichung dieser Publikation. Insbesondere danken wir Professorin Regina Ammicht Quinn und PD Dr. Jessica Heesen für ihre wertvollen Anmerkungen und ihre fachliche und persönliche Unterstützung bei der Erstellung und Veröffentlichung dieser Studie. Sie entstand im Rahmen des Forschungsprojekts TANGO (‘Technik-Innovation: sozialwissenschaftliche und ethische Analysen zu Governance‘), das 2019–2020 vom Bundesministerium für Bildung und Forschung gefördert wurde, sowie im Rahmen des Forschungsprojekts ‚digilog@bw – Digitalisierung im Dialog‘, das vom Baden-Württembergischen Ministerium für Wissenschaft, Forschung und Kunst gefördert wird.

Jovanna Horn danken wir für ihre inhaltliche Unterstützung bei der Anfertigung dieser Studie. Zudem hat Miriam Schröder einen Beitrag zur Ausarbeitung des Themenbereichs Pornografie geleistet. Darüber hinaus danken wir Dr. phil. Ingrid Stapf, Mitherausgeberin der Reihe ‚Kommunikations- und Medienethik‘, für ihre Unterstützung bei der Veröffentlichung. Nicht zuletzt bedanken wir uns bei unseren Kolleg*innen, Alexander Orlowksi und Anna Tilling.

Inhaltsverzeichnis

Hinweis	13
1 Einleitung	15
2 Theoretische und methodische Grundlagen	19
2.1 Definitionen/begriffliche Klärungen	19
2.2 Forschungsstand	24
2.3 Methodischer Zugang und theoretischer Hintergrund	28
3 Stand der Technik	37
3.1 Synthetische Bilder	38
3.2 Synthetische Videos	38
3.3 Synthetische Audiodateien	40
3.4 Entwicklungsumgebung und Akteur*innen	40
3.5 Einschätzung	41
4 Technikbewertung	43
4.1 Politik	44
4.2 Pornografie	53
4.3 Kriminalität und Sabotage	57
4.4 Strafverfolgung	59
4.5 Wirtschaft	61
4.6 (Politische) Kunst, Aktivismus und Unterhaltung	64
4.7 Bildung	66

Inhaltsverzeichnis

5	Governancemechanismen für Deepfakes	69
5.1	Entwicklungs-Infrastrukturen und sozialer Kontext	70
5.2	Infrastruktureller und sozialer Kontext der Verbreitung	76
5.3	Nationale Gesetzgebung	81
	Netzwerkdurchsetzungsgesetz	84
	Medienstaatsvertrag	89
	Pflichten von Suchmaschinenbetreibern	92
	Einschätzung	93
5.4	Supranationale Richtlinien	94
5.5	EU Joint and Coordinated Action against Disinformation	105
5.6	Plattformolicies (Medien selbstregulierung und Unternehmensethik)	114
5.7	Technische Normen und Zertifizierung	131
5.8	KI-Ethikleitlinien und Zertifizierung	133
5.9	Deepfake-Detektion und Forschungsförderung	136
	Nationale Ebene: Deutschland	137
	Supranationale Ebene: EU	138
	Internationale Forschung und Forschungsförderung	139
5.10	Authentifizierung	151
5.11	Weitere Forschungsförderung	154
5.12	Förderung professioneller Faktenprüfung	156
5.13	Außenpolitik	158
5.14	Selbst- und Opferschutz	160
5.15	Berufsethik	162
5.16	Bildung und Medienkompetenz	165
6	Fazit	168
6.1	Technikbewertung	169
6.2	Governance von Deepfakes	174
7	Handlungsempfehlungen zur Governance von Deepfakes	186
	Literaturverzeichnis	197