Bild und Recht – Studien zur Regulierung des Visuellen

Olivia Hägle

Deep Fakes vor Gericht

Nomos

Bild und Recht – Studien zur Regulierung des Visuellen

Herausgegeben von
Prof. Dr. Thomas Dreier
Dr. Dr. Grischka Petri
Prof. Dr. Wolfgang Ullrich
Prof. Dr. Matthias Weller

Band 15

Olivia Hägle

Deep Fakes vor Gericht

Eine persönlichkeits- und urheberrechtliche Untersuchung zum Umgang mit Technologie und Technologiefolgen

Nomos

Collage aus Papier. Inspiriert durch die Arbeiten von Lola Dupre. Dem Werk zugrunde liegt ein mithilfe des KI-Tools Firefly Image 5 von Adobe generiertes Portrait einer jungen Frau. Da das Modell ausschließlich mit Inhalten trainiert wurde, die durch Adobe rechtmäßig genutzt werden können, sind die mithilfe des Modells generierten Inhalte durch Adobe auch zur kommerziellen Nutzung freigegeben.

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

ISBN 978-3-7560-2053-9 (Print)
ISBN 978-3-7489-5595-5 (ePDF)

Onlineversion
Nomos eLibrary

1. Auflage 2026

Meinen Liebsten

Vorwort

Das Phänomen der Deep Fakes markiert den vorläufigen Höhepunkt der Evolution des visuellen Fakes und stellt Recht, Technologie und Gesellschaft vor neue und bekannte Herausforderungen. Die Thematik erreicht daher zunehmend auch den (rechts-)wissenschaftlichen Diskurs. Die vorliegende Arbeit beleuchtet die Problematik der Deep Fakes aus rechtswissenschaftlicher Perspektive und nimmt dabei insbesondere das Urheber- und Persönlichkeitsrecht sowie die entsprechenden Regelungen der KI-Verordnung in den Blick. Die Arbeit, die im Wintersemester 2025/2026 von der rechtswissenschaftlichen Fakultät der Albert-Ludwigs-Universität Freiburg als Dissertation angenommen wurde, entstand während meiner Tätigkeit als wissenschaftliche Mitarbeiterin am Zentrum für angewandte Rechtswissenschaft (ZAR) des Karlsruher Instituts für Technologie (KIT) und wurde Ende August 2025 eingereicht. Sie befindet sich daher im Wesentlichen auf dem Stand von August 2025. Später erschienene Literatur und Rechtsprechung (insb. die Entscheidungen des OLG Hamburg, Urt. v. 10.12.2025 – 5 U 104/24 sowie LG München I, Urt. v. 11.11.2025 – 42 O 14139/24) konnte jedoch noch teilweise vor der Drucklegung berücksichtigt werden.

Zur Entstehung dieser Arbeit haben eine Reihe von Personen beigetragen, denen mein größter Dank gilt.

Bedanken möchte ich mich zunächst herzlichst bei meinem Doktorvater Herrn Prof. Dr. Thomas Dreier M.C.J. für die hervorragende Betreuung, die meine Arbeit durch die zahlreichen Gespräche stets vorangebracht und mir zugleich alle Freiheiten gelassen hat, um meinen eigenen Weg zu gehen mit einem Thema, das zu Beginn der Arbeit an der Untersuchung noch recht unbekannt war, für das wir uns jedoch beide schon früh begeistern konnten. Mein Dank gilt darüber hinaus auch Herrn Prof. Dr. Maximilian Haedicke, LL.M. (Georgetown) für die zügige Erstellung des Zweitgutachtens.

Mein Dank gilt zudem den Herausgebern der Reihe Bild und Recht – Studien zur Regulierung des Visuellen für die Aufnahme meiner Arbeit in die Schriftenreihe sowie dem Nomos Verlag, insbesondere Herrn Marco Ganzhorn für die hervorragende Betreuung und die Verlegung meiner Arbeit. Für die finanzielle Unterstützung bei der Veröffentlichung meiner Arbeit danke ich zudem der Studienstiftung ius vivum.

Bedanken möchte ich mich darüber hinaus bei meinen Kolleginnen und Kollegen am ZAR, die meine Tätigkeit dort zu einer überaus lehrreichen, spaßigen und abwechslungsreichen Zeit machten, an die ich gerne zurückdenke. Ich danke zudem auch meinen Projektkolleginnen und -kollegen außerhalb des ZARs, insbesondere am Institut für Technikfolgenabschätzung (ITAS) des KIT, mit denen ich im Projekt

„Interdisziplinäre Zugänge zu Deepfakes“ zusammengearbeitet und diskutiert habe und durch die ich immer wieder erleben konnte, wie gewinnbringend Perspektiven außerhalb der Rechtswissenschaft auf meine eigene wissenschaftliche Forschung sein können.

Mein herzlichster Dank gilt zudem meinem persönlichen Umfeld für die unendliche Unterstützung während der Zeit meiner Promotion. Ihr hattet nicht nur immer ein offenes Ohr für meine Gedanken zu meiner wissenschaftlichen Arbeit, sondern habt mich beständig auch daran erinnert mein Leben abseits der Dissertation nicht zu vernachlässigen. Bedanken möchte ich mich dabei nicht nur bei denjenigen, die mich die gesamte Zeit meiner Dissertation über unterstützt haben, sondern auch bei denjenigen, die mich erst seit der Endphase der Dissertation begleiten und doch keinen unwesentlichen Beitrag zum erfolgreichen Abschluss meiner Promotion geleistet haben.

Mein besonderer Dank gilt zunächst Lilli Wiedemann, die mich bereits mein ganzes Leben lang begleitet und die keine Sekunde daran gezweifelt hat, dass ich auch dieses Projekt erfolgreich abschließen werde. Mein größter Dank gilt zudem meinen Eltern Heike und Franz Hägle, die mir durch ihre bedingungslose Unterstützung in jeglicher Hinsicht überhaupt erst ermöglicht haben, meinen Weg zu gehen und das Projekt der Dissertation anzugehen und insbesondere auch durch die fleißige Unterstützung beim Korrekturlesen des Manuskripts einen ganz wesentlichen Beitrag zum erfolgreichen und zügigen Abschließen der Arbeit geleistet haben. Bedanken will ich mich zuletzt bei meiner Schwester Melissa Hägle sowie bei Leon Seith, mit denen ich zahlreiche Diskussionen über die technischen und künstlerischen Implikationen von Deep Fakes und generativer KI geführt habe. Leon danke ich darüber hinaus für die Unterstützung bei der Cover-Gestaltung dieser Arbeit sowie dafür, dass er mich stets auf dem neuesten Stand darüber gehalten hat, welche Deep Fakes gerade im Umlauf sind und immer im Blick hatte, dass auch das Lachen nicht zu kurz kommt. Melissa danke ich zunächst für ihre technische Expertise und die Unterstützung beim Korrekturlesen, insbesondere der technischen Anteile meiner Arbeit. Ich danke ihr zudem dafür, dass sie nicht müde wurde mit klugem und kritischem Blick meine Erwägungen zu hinterfragen und meine Argumente in der Diskussion zu schärfen. Schließlich hatte sie nicht nur in Bezug auf meine Arbeit, sondern auch darüber hinaus immer ein offenes Ohr für mich, auch zu später Stunde, wenn während einer der zahlreichen Nachtschichten eine kurze Pause und Ablenkung von der Arbeit nötig waren. Ohne euch wäre diese Arbeit nicht möglich gewesen. Euch ist diese Arbeit daher gewidmet.

Karlsruhe, im Februar 2026 *Olivia R. Hägle*

Inhaltsübersicht

Inhaltsverzeichnis

Abbildungsverzeichnis

Abkürzungsverzeichnis

aA	Andere Ansicht
Abs.	Absatz
AcP	Archiv für die civilistische Praxis
a.F.	Alte Fassung
AfP	Archiv für Presserecht (Zeitschrift für das gesamte Medienrecht)
AI	Artificial Intelligence
APuZ	Aus Politik und Zeitgeschichte
ARSP	Archiv für Rechts- und Sozialphilosophie
Aufl.	Auflage
AusArt	AusArt: Journal for Research in Art
B.U. J. Sci. & Tech. L.	Boston University Journal of Science & Technology Law
BayObLGZ	Entscheidungen des Bayerischen Obersten Landesgerichts in Zivilsachen
Beschl.	Beschluss
BGBl.	Bundesgesetzblatt
BGHZ	Entscheidungen des Bundesgerichtshofs in Zivilsachen
BKR	Zeitschrift für Bank- und Kapitalmarktrecht
BR-Drs.	Bundesrats-Drucksache
BT-Drs.	Bundestags-Drucksache
BVerfGE	Entscheidungen des Bundesverfassungsgerichts
CGI	Computer Generated Imagery
CNN	Convolutional Neural Net(work)
CR	Computer und Recht
CRi	Computer Law Review International
Data-Act	Verordnung (EU) 2023/2854 des Europäischen Parlaments und des Rates vom 13. Dezember 2023 über harmonisierte Vorschriften für einen fairen Datenzugang und eine faire Datennutzung sowie zur Änderung der Verordnung (EU) 2017/2394 und der Richtlinie (EU) 2020/1828 (Datenverordnung), Amtsblatt der Europäischen Union 2023 Nr. L S. 1
Datenschutz-RL	Richtlinie 95/46/EG des Europäischen Parlaments und des Rates vom 24. Oktober 1995 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten und zum

	freien Datenverkehr, Amtsblatt der Europäischen Union 1995 Nr. L 281 S. 31
DÖV	Die Öffentliche Verwaltung
DSA	Verordnung (EU) 2022/2065 des Europäischen Parlaments und des Rates vom 19. Oktober 2022 über einen Binnenmarkt für digitale Dienste und zur Änderung der Richtlinie 2000/31/EG (Gesetz über digitale Dienste) - Digital Services Act (DSA), Amtsblatt der Europäischen Union 2022 Nr. L 277 S. 1
DSGVO	Verordnung (EU) 2016/679 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), Amtsblatt der Europäischen Union 2016 Nr. L 119 S. 1
DSM-RL	Richtlinie (EU) 2019/790 des Europäischen Parlaments und des Rates vom 17. April 2019 über das Urheberrecht und die verwandten Schutzrechte im digitalen Binnenmarkt und zur Änderung der Richtlinien 96/9/EG und 2001/29/EG, Amtsblatt der Europäischen Union 2019 Nr. L 130 S. 92
DuD	Datenschutz und Datensicherheit
DVBl	Deutsches Verwaltungsblatt
Erw.Gr.	Erwägungsgrund
Ent.	Entscheidung
EuCML	Journal of European Consumer and Market Law
EuR	Europarecht
EuZW	Europäische Zeitschrift für Wirtschaftsrecht
FAZ	Frankfurter Allgemeine Zeitung
FF	Forum Familienrecht
Fn.	Fußnote
FS	Festschrift
GAN	Generative Adversarial Network
Geo. L. Tech. Rev.	Georgetown Law Technology Review
Gewaltschutz-RL	Richtlinie (EU) 2024/1385 des Europäischen Parlaments und des Rates vom 14. Mai 2024 zur Bekämpfung von Gewalt gegen Frauen und häuslicher Gewalt, Amtsblatt der Europäischen Union 2024 Nr. L S. 1
GPAI	General Purpose Artificial Intelligence

GRUR	Gewerblicher Rechtsschutz und Urheberrecht
GRUR-Int.	Journal of European and International IP Law
GRUR-Prax	Gewerblicher Rechtsschutz und Urheberrecht in der Praxis
GRUR-RR	Gewerblicher Rechtsschutz und Urheberrecht Rechtsprechungs-Report
GS	Gedächtnisschrift
Harv. L. Rev.	Harvard Law Review
hM	Herrschende Meinung
IEEE	Institute of Electrical and Electronics Engineers
IIC	International Review of Intellectual Property and Competition Law
IJASCA	International Journal of Advances in Soft Computing and its Applications
InfoSoc-RL	Richtlinie 2001/29/EG des Europäischen Parlaments und des Rates vom 22. Mai 2001 zur Harmonisierung bestimmter Aspekte des Urheberrechts und der verwandten Schutzrechte in der Informationsgesellschaft – InfoSoc-RL, Amtsblatt der Europäischen Union 2001 Nr. L 167 S. 10
InTeR	Zeitschrift zum Innovations- und Technikrecht
iVm	In Verbindung mit
IWRZ	Zeitschrift für Internationales Wirtschaftsrecht
JIPITEC	Journal of Intellectual Property, Information Technology and Electronic Commerce Law
JR	Juristische Rundschau
JURA	Juristische Ausbildung
JuS	Juristische Schulung
JZ	JuristenZeitung
K&R	Kommunikation & Recht
KI	Künstliche Intelligenz
KI-VO	Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche

	Intelligenz), Amtsblatt der Europäischen Union 2024 Nr. L S. 1
KIR	Künstliche Intelligenz und Recht
KNN	Künstliches Neuronales Netz
KUG	Gesetz betreffend das Urheberrecht an Werken der bildenden Künste und der Photographie (Kunsturhebergesetz), RGBl. 1907 Nr. 3 S. 7
LMK	Fachdienst Zivilrecht – Leitsätze mit Kommentierung
LTZ	Zeitschrift für die digitale Rechtsanwendung
Marrakesch-VO	Verordnung (EU) 2017/1563 des Europäischen Parlaments und des Rates vom 13. September 2017 über den grenzüberschreitenden Austausch von Vervielfältigungsstücken bestimmter urheberrechtlich oder durch verwandte Schutzrechte geschützter Werke und sonstiger Schutzgegenstände in einem barrierefreien Format zwischen der Union und Drittländern zugunsten blinder, sehbehinderter oder anderweitig lesebehinderter Personen, Abl. L 242 S. 1
ML	Machine Learning (Maschinelles Lernen)
MDR	Monatszeitschrift für Deutsches Recht
MM	Mindermeinung
MMR	Multimedia und Recht – Zeitschrift für IT-Recht und Recht der Digitalisierung
n.F.	Neue Fassung
NJ	Neue Justiz
NJOZ	Neue Juristische Online-Zeitschrift
NJW	Neue Juristische Wochenschrift
NJW-RR	Neue Juristische Wochenschrift Rechtsprechungs-Report Zivilrecht
NStZ	Neue Zeitschrift für Strafrecht
NVwZ	Neue Zeitschrift für Verwaltungsrecht
NZV	Neue Zeitschrift für Verkehrsrecht
OdW	Ordnung der Wissenschaft
Phil. Trans. R. Soc. A.	Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences
PNAS	Proceedings of the National Academy of Sciences of the United States of America
PuK	Politik & Kultur

RBÜ	Berner Übereinkunft zum Schutz von Werken der Literatur und Kunst vom 09.09.1886, BGBl. 1973 II S. 1069
RDi	Recht Digital
Rdnr.	Randnummer
RGBl.	Reichsgesetzblatt
RuZ	Recht und Zugang
RW	Rechtswissenschaft
Schutzdauer-RL	Richtlinie 93/98/EWG des Rates vom 29. Oktober 1993 zur Harmonisierung der Schutzdauer des Urheberrechts und bestimmter verwandter Schutzrechte, Amtsblatt der Europäischen Gemeinschaften 1993 Nr. L 290 S. 9
Schutzdauer-RL	Richtlinie 2006/116/EG des Europäischen Parlaments und des Rates vom 12. Dezember 2006 über die Schutzdauer des Urheberrechts und bestimmter verwandter Schutzrechte, Amtsblatt der Europäischen Union 2006 Nr. L 372 S. 12
TDM	Text- und Data-Mining
T. Jefferson L. Rev.	Thomas Jefferson Law Review
TRIPS	Übereinkommen über handelsbezogene Aspekte der Rechte des geistigen Eigentums vom 15.04.1994, BGBl. 1994 II S. 1438
UAbs.	Unterabsatz
UCLA J.L. & Tech.	UCLA Journal of Law and Technology
UFITA	Archiv für Medienrecht und Medienwissenschaft
UGC	User Generated Content
UrhDaG	Gesetz über die urheberrechtliche Verantwortlichkeit von Diensteanbietern für das Teilen von Online-Inhalten (Urheberrechts-Diensteanbieter-Gesetz), BGBl. I 2021, Nr. 27 vom 04.06.2021, S. 1204
Urt.	Urteil
VAE	Variational Autoencoder
Va. J.L. & Tech.	Virginia Journal of Law & Technology
VerfBlog	Verfassungsblog
WCT	WIPO-Urheberrechtsvertrag (WIPO Copyright Treaty) vom 20.12.1996, BGBl. 2003 II S. 755
WPPT	WIPO-Vertrag über Darbietungen und Tonträger (WIPO Performance and Phonograms Treaty) vom 20.12.1996, BGBl. 2003 II S. 754

ZD	Zeitschrift für Datenschutz
ZEuP	Zeitschrift für Europäisches Privatrecht
ZfDR	Zeitschrift für Digitalisierung und Recht
ZfpS	Zentrum für politische Schönheit
ZfPW	Zeitschrift für die gesamte Privatrechtswissenschaft
ZGE	Zeitschrift für geistiges Eigentum
ZRP	Zeitschrift für Rechtspolitik
ZStW	Zeitschrift für die gesamte Strafrechtswissenschaft
ZUM	Zeitschrift für Urheber- und Medienrecht
ZUM-RD	Zeitschrift für Urheber- und Medienrecht - Rechtsprechungsdienst

Im Übrigen wird auf die im Literaturverzeichnis enthaltenen Abkürzungen sowie Kirchner, Abkürzungsverzeichnis der Rechtssprache, 11. Auflage Berlin/Boston, 2024 verwiesen.

Einleitung

„There is a prejudice against the spoken lie, but none against any other, and by examination and mathematical computation I find that the proportion of the spoken lie to the other varieties is as 1 to 22,894."
Quelle: Twain, My first lie and how I got out of it, in: Twain, The man that corrupted Hadleyburg: and other essays and stories, New York, 1917, S. 159-170, S. 169

§ 1 Einführung in den Problembereich: Das visuelle Fake vor Gericht

Mithilfe von Modellen generativer Künstlicher Intelligenz auf der Grundlage von Deep Learning und einer Vielzahl von Trainingsdaten ist es möglich authentisch anmutende Bild-, Video- und Audiofälschungen beliebigen Inhalts zu erstellen. Es handelt sich dabei um sogenannte *Deep Fakes*[1]. Diese Deep Fakes sind aufgrund des technologischen Fortschritts zunehmend nicht mehr von echten Inhalten zu unterscheiden und beinhalten daher ein ganz besonderes Täuschungspotenzial, das von missbräuchlichen Personen[2] gezielt manipulativ ausgenutzt und eingesetzt werden kann. Das Deep Fake macht sich dazu die besondere Überzeugungskraft visueller und audiovisueller Inhalte zunutze. Während der Wahrheitsgehalt des gesprochenen und geschriebenen Wortes vielfach angezweifelt wird, genießt das Bild, insbesondere in Gestalt von Fotografien und Videos, weithin ein ganz besonderes Vertrauen. In der Folge wird dem Deep Fake ein besonderes disruptives Potenzial zugeschrieben. Aus diesem Grund waren das Deep Fake und die Technologie der generativen Künstlichen Intelligenz auch auf der jüngsten JumiKo[3] aus verschiedener Perspektive Thema.[4]

1 Diese Schreibweise wurde hier aus dem Grund gewählt, dass auf diese Weise bereits hinreichend deutlich wird, dass man zwischen der Technologie (Deep Learning) und der Fälschung (Fake) unterscheiden muss und es entscheidend darauf ankommt, wie ein mithilfe von Deep Learning erstelltes/manipuliertes Bild zum „Fake" wird.

2 Insbesondere auch aufgrund der Gendersensibilität der Deep Fake Problematik (siehe dazu näher im weiteren Verlauf dieser Arbeit, insbesondere auch im 1. Teil unter § 2 Abschnitt V.1.) wurde hier, soweit möglich, grundsätzlich eine genderneutrale Sprache gewählt. An verschiedensten Stellen ist jedoch insbesondere auf feststehende Begriffe der Rechtssprache zu rekurrieren. Sofern diese feststehenden rechtlichen Begriffe potenziell eine Genderspezifität zu implizieren vermögen, erfolgt die Verwendung gleichwohl ausdrücklich ebenfalls mit der Intention sämtliche Geschlechter sprachlich zu inkludieren. Siehe zur geschlechtergerechten Sprache im Recht vor dem Hintergrund des Markenrechts etwa *Grabrucker*, GRUR 2023, 113.

3 *96. Konferenz der Justizministerinnen und Justizminister,* Beschlüsse, https://www.justiz.nrw.de/JM/jumiko/beschluesse/2025/Fruehjahrskonferenz_2025 [zuletzt geprüft am 01.08.2025].

4 Siehe aus persönlichkeitsrechtlicher Perspektive: *96. Konferenz der Justizministerinnen und Justizminister,* Beschluss TOP II.18: Bildbasierte sexualisierte Gewalt - Verbesserung des strafrechtlichen Schutzes, https://www.justiz.nrw.de/sites/default/files/2025-06/TOP%20II.18%20-%20Bildbasierte%20sexualisierte%20Gewalt.pdf [zuletzt geprüft am 01.08.2025]; aus urheberrechtlicher Perspektive: *96. Konferenz der Justizministerinnen und Justizminister,* Beschluss TOP I.18: KI und Urheberrecht - Urheberrechtliche Folgerungen aus der zunehmenden Verbreitung von Künstlicher Intelligenz, https://www.justiz.nrw.de/s

Gleichwohl handelt es sich bei Deep Fakes nicht um eine neue Problematik. Schon immer werden Bilder und insbesondere auch Fotografien und Videos gefälscht. Die Erscheinungsformen visueller Fakes reichen dabei von Kunstfälschungen über fotografische Fakes, die durch Einflussnahmen vor, während und nach der Aufnahme, auf analogem oder digitalem Wege entstehen können, bis hin zu vollständig synthetischen Computergrafiken. Neue Erscheinungen wie Deep Fakes sind daher also kein gänzlich neues Phänomen, sondern markieren nur den (vorläufigen) Höhepunkt der Evolution des visuellen Fakes.

Derartige visuelle Fakes geraten regelmäßig in Konflikt mit dem Recht und stellen – bedingt durch die technologischen Fortschritte – zunehmend auch die Gerichte vor neue Herausforderungen. Bereits Anfang der 2000er Jahre hatten sich Bundesgerichtshof und Bundesverfassungsgericht anhand des Falls des ehemaligen Vorstandsvorsitzenden der Deutschen Telekom AG mit dem visuellen Fake in Gestalt einer satirischen Fotomontage zu befassen.[5] Jüngst beschäftigte die deutschen Gerichte das visuelle Fake auch bereits in Gestalt von Deep Fakes, etwa im Zusammenhang mit Deep Fakes des ehemaligen Bundeskanzlers Olaf Scholz sowie des Moderators Eckart von Hirschhausen.[6] In derartigen Konstellationen, in denen das visuelle Fake den Gegenstand gerichtlicher Verfahren bildet, stellen sich regelmäßig Fragen der Kollision von persönlichkeitsrechtlichen Interessen mit kommunikativen Freiheiten. Doch auch weitergehende Kommunikationsinteressen können betroffen sein, etwa wenn visuelle Fakes wie im Fall des Olaf Scholz Deep Fakes den politischen Bereich berühren. In anderen Konstellationen, wie dem Fall des Moderators Eckart von Hirschhausen stehen auch wirtschaftliche Interessen im Raum. Der weit überwiegende Anteil der im Umlauf befindlichen Deep Fakes ist jedoch pornografischer Natur.[7] In diesem Bereich liegt auch der Ursprung der Deep Fake Technologie.[8] Diese pornografischen Fakes bieten ein immenses Schädigungspotenzial für die davon betroffenen Personen, sie können in ihren Auswirkungen jedoch auch weit über den

ites/default/files/2025-06/TOP%20I.18%20-%20KI%20und%20Urheberrecht.pdf [zuletzt geprüft am 01.08.2025].

5 Zunächst BGH, Urt. v. 30.09.2003, VI ZR 89/02 - Bewertung einer Fotomontage im Gesamtzusammenhang, NJW 2004, 596 = BGHZ 156, 206; sodann BVerfG, Beschl. v. 14.02.2005, 1 BvR 240/04 - Verwendung von Fotomontagen in satirischen Kontexten [Ron Sommer], NJW 2005, 3271; sowie im Anschluss daran wieder BGH, Urt. v. 08.11.2005, VI ZR 64/05 - Verletzung des allgemeinen Persönlichkeitsrechts bei Verwendung eines manipulierten Fotos, NJW 2006, 603.

6 LG Berlin II, Beschl. v. 13.02.2024, 15 O 579/23 - Deepfake des Zentrums für Politische Schönheit, juris; OLG Frankfurt a.M., Urt. v. 04.03.2025, 16 W 10/25 - Prüfpflichten des Hostproviders [Eckart von Hirschhausen], ECLI:DE:OLGHE:2025:0304.16W10.25.00, https://www.rv.hessenrecht.hessen.de/bshe/document/LARE250000357.

7 Nach einer Studie von 2019 handelt es sich bei 96% der im Internet im Umlauf befindlichen Deep Fakes um solche pornografischer Natur. Siehe *Ajder* et al. - The State of Deepfakes: Landscape, Threats, and Impact, 1, https://regmedia.co.uk/2019/10/08/deepfake_report.pdf [zuletzt geprüft am 15.06.2025].

8 Bereits 2017 erstmals darüber berichtend *Cole,* AI-Assisted Fake Porn Is Here and We're All Fucked, https://www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn [zuletzt geprüft am 22.06.2025].

individuellen Bereich hinausreichen, wenn derartige visuelle Fakes gezielt als Mittel im Meinungskampf eingesetzt werden, etwa um Journalistinnen einzuschüchtern.[9] Neben derartigen Anwendungsfällen, die primär unter persönlichkeitsrechtlichen Aspekten diskutiert werden, ist insbesondere in wirtschaftlich aufgeladenen Konstellationen vielfach auch das Urheberrecht von Bedeutung. Vor einiger Zeit sorgte etwa ein Audio Deep Fake der Künstler Drake und The Weeknd für Aufregung in der Musikbranche.[10] Auch in Konstellationen, in denen der Bezug zu konkreten Personen jedenfalls nicht mehr unmittelbar erkennbar erscheint[11], wie es etwa im Zusammenhang mit KI-generierten „Stockfotos"[12] der Fall ist, könnten insbesondere urheberrechtliche Interessen weiterhin beeinträchtigt sein. Auch im Zusammenhang mit weiteren Anwendungsfällen von Deep Fake-Technologie, etwa im Bereich von Kunst und Kultur, insbesondere in der Filmindustrie, die zunächst wünschenswert erscheinen, können potenzielle Konflikte mit urheber- und persönlichkeitsrechtlichen Interessen entstehen.

Aufgrund des Scheins von Authentizität können Deep Fakes jedoch nicht nur die unmittelbar betroffenen Personen in empfindlicher Weise berühren, sondern sie können aufgrund ihrer täuschenden Wirkung als spezielle Form von Mis- bzw. Desinformation durch ihre Einwirkung auf die Vorstellungen der Rezipierenden auch weitere Gefahren beinhalten. Dies gilt insbesondere in den Fällen, in denen diese – auf Grundlage von Deep Fakes getroffenen und gezielt manipulativ beeinflussten – Vorstellungen zur Grundlage für wichtige Entscheidungen werden. Da das Recht (insbesondere im gerichtlichen Verfahren) nicht nur vielfach mit Bildern konfrontiert ist, sondern sie sich bisweilen, insbesondere als Beweismittel, gezielt zunutze macht, stellen Deep Fakes bzw. das Wissen um den vermeintlichen Einsatz von Deep Fakes das Recht auch in dieser Hinsicht vor (neue) Herausforderungen.

Deep Fakes beruhen auf neuen technischen Entwicklungen und sind eine noch recht junge Erscheinung, für die es daher bislang nur wenige spezifisch darauf zugeschnittene rechtliche Regelungen gibt. Allerdings sind vergleichbare Erscheinungen visueller Fakes – und mithin regelmäßig auch die zugrunde liegenden Interessenkonflikte jedenfalls in Teilen – bekannt, sodass eine Vielzahl rechtlicher Regelungen existiert, die potenziell auch im Bereich der neuen Erscheinung Anwendung finden

9 Über ihre Erfahrungen berichtet etwa *Ayyub,* I Was The Victim Of A Deepfake Porn Plot Intended To Silence Me, 21.11.2018, https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316 [zuletzt geprüft am 15.06.2025].

10 Darüber berichtend etwa *Coscarelli* - An AI Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World, The New York Times, 19.04.2023, https://www.nytimes.com/2023/04/19/arts/music/ai-drake-the-weeknd-fake.html [zuletzt geprüft am 19.06.2025].

11 Inwieweit dies tatsächlich auch dazu führt, dass derartigen synthetischen Inhalten keine persönlichkeitsrechtliche Relevanz mehr zukommt, gilt es im 2. Teil dieser Arbeit zu untersuchen.

12 *Berger,* Stockfoto-Firma veröffentlicht 100.000 KI-Gesichter, 2019, https://www.heise.de/newsticker/meldung/Stockfoto-Firma-veroeffentlicht-100-000-KI-Gesichter-4537889.html [zuletzt geprüft am 16.06.2025].

können. Jedoch weist das Deep Fake gegenüber bekannten Erscheinungen visueller Fakes einige Spezifika auf, die besondere Herausforderungen mit sich bringen. Diese rühren zum einen aus der genutzten Technologie der generativen Künstlichen Intelligenz, die es ermöglicht ohne besonderen Aufwand automatisiert und kostengünstig (nicht nur auf den visuellen Bereich beschränkte) zunehmend authentisch anmutende Fakes herzustellen.[13] Gleichwohl ist der Geltungsanspruch, den Bilder genießen, weiterhin nahezu ungebrochen, sodass eine zunehmende Diskrepanz zwischen dem faktischen manipulativen Potenzial des Visuellen und dem visuellen Geltungsanspruch entsteht. Zum anderen kann auch ein besonderes Potenzial von Deep Fakes im Hinblick auf deren Auswirkungen und Reichweite[14] in digitalen Kommunikationsräumen[15] beobachtet werden. Dies führt einerseits zu einer Intensivierung der Folgen für unmittelbar betroffene Personen und verstärkt andererseits auch das über das unmittelbar betroffene Individuum hinausgehende manipulative Potenzial im Hinblick auf die Wirkung als Mis-/Desinformation.

§ 2 Gang, Schwerpunkt und Grenzen der Untersuchung

Sowohl Technologie als auch Technologiefolgen rufen regelmäßig ein Bedürfnis nach Regulierung hervor. Dies gilt in besonderem Maße auch für Deep Fake-Technologie und deren Folgen für Individuen und die Gesellschaft. Denn die Technologie erfordert eine Vielzahl von Trainingsbildern. Da und soweit es sich bei Deep Fakes klassischerweise um Personenbilder handelt, sind also eine Vielzahl von Personenbildnissen erforderlich, um die Modelle trainieren zu können. Damit ist aus rechtlicher Perspektive zunächst der persönlichkeitsrechtliche Schutz angesprochen. Da an den genutzten Trainingsbildern jedoch regelmäßig nicht nur den abgebildeten Personen Rechte zustehen, sondern in der fotografierenden Person auch Urheberrechte und verwandte Schutzrechte bestehen können und die Modelle zur Erstel-

13 *Patrini* et al., Commoditisation of AI, digital forgery and the end of trust: how we can fix it, 17.03.2018, https://giorgiop.github.io/posts/2018/03/17/AI-and-digital-forgery/ [zuletzt geprüft am 22.06.2025]; KI-Technologie verstärkt also einerseits die Problematik, allerdings kann die Technologie andererseits auch dazu beitragen die Gefahren einzuhegen. So wurde beispielsweise in einer empirischen Studie herausgefunden, dass KI-Sprachmodelle wie ChatGPT (damals noch auf der Grundlage des Vorgänger-Modells GPT-3) den Menschen einerseits in der Produktion von Falschinformation überlegen sind, da die synthetisch generierten Falschinformationen von Rezipierenden schlechter erkannt wurden als menschlich erzeugte Falschinformationen, andererseits jedoch auch in der Kommunikation echter authentischer Informationen den Menschen überlegen sind. Dieser Umstand könnte auch zur Bekämpfung derartiger Phänomene genutzt werden. Siehe dazu die Studie von *Spitale/Biller-Andorno/Germani*, Sci Adv 9 (2023), eadh1850.

14 In einer Studie von 2018 konnte beispielsweise beobachtet werden, dass sich falsche Informationen auf Twitter nicht nur schneller, sondern auch weiter und tiefer in dem sozialen Netzwerk verbreiteten als echte bzw. wahre Informationen. Siehe *Vosoughi/Roy/Aral*, Science 359 (2018), 1146.

15 Zu den Besonderheiten der neuen digitalen Diskursräume in sozialen Netzwerken siehe etwa *Völzmann*, MMR 2021, 619 (620).

lung qualitativ hochwertiger Inhalte insbesondere auch auf Trainingsinhalte kreativ-schöpferischer Natur angewiesen sind, ist insbesondere auch das Urheberrecht angesprochen, dem zunehmende Innovationssteuerungswirkung zukommt. Hinzu kommt nunmehr in Europa auch die spezifische Technologie-Regulierung im Wege der KI-Verordnung. Da damit zwar grundsätzlich bereits eine Reihe rechtlicher Regelungen vorhanden sind, die jedoch an verschiedenen Stellen zunehmend an ihre Grenzen (insbesondere faktischer Natur) gelangen und das Recht – wie das Beispiel der Deep Fakes eindrucksvoll zeigt – nicht losgelöst von Technologie und Gesellschaft steht, erfolgt im Rahmen dieser Arbeit eine ergänzende Betrachtung auch von Maßnahmen außerhalb des Rechts.

Die weitergehenden (mittelbaren) Folgen von Deep Fakes, die aus ihrem Täuschungspotenzial erwachsen, sind zu weitreichend, als dass sie im Rahmen dieser Arbeit umfassend diskutiert werden könnten. Mit der Problematik des irreführenden Potenzials von Deep Fakes vor Gericht soll daher hier nur ein ganz spezieller Bereich dieser Problematik herausgegriffen und unter dem Aspekt des Wahrheitsschutzes betrachtet werden.

Der erste Teil dieser Arbeit zeichnet dazu zunächst den Weg vom Original zum Deep Fake nach und führt das visuelle Fake als besondere Herausforderung für das Recht im Lichte technologischen Fortschritts ein. Dabei widmet sich die Arbeit zunächst den technologischen Hintergründen von Deep Fakes (§ 1), sowie sodann in Form der Historie des visuellen Fakes (§ 2) der Einordnung von Deep Fakes in die weitere Systematik visueller Fakes. Sodann ist auf die visuellen Besonderheiten einzugehen, die aus der besonderen Macht der Bilder erwachsen (§ 3). Der erste Teil dieser Arbeit schließt mit einem Kapitel zum Verhältnis von Recht und (KI-)Technologie (§ 4).

Der Schwerpunkt liegt sodann im zweiten und dritten Teil dieser Arbeit auf der Untersuchung des Umgangs mit den unmittelbaren Folgen von Deep Fakes, also insbesondere der Frage, ob und wie mit den vorhandenen Instrumenten aus den Bereichen Recht (2. Teil), Technik (3. Teil § 1) und Gesellschaft (3. Teil § 2) ein Schutz der Person vor Deep Fakes gelingen kann. Dazu erfolgt im zweiten Teil zunächst eine rechtliche Einordnung des Deep Fakes (§ 1), das unter bestimmten Umständen als künstlerisch-kommunikative Persönlichkeitsentfaltung potenziell seinerseits einem rechtlichen Schutz unterfallen kann. Sodann erfolgt eine Betrachtung von Deep Fakes unter dem Aspekt des Persönlichkeits- (§ 2) und Urheberschutzes (§ 3). Schließlich wird noch die spezielle Deep Fake Regulierung im Wege der KI-Verordnung betrachtet (§ 4). Abschließend folgt ein Exkurs zu den Besonderheiten im Bereich des Akustischen (§ 5). Aufgrund der interdisziplinären Fundierung dieser Arbeit werden sodann im dritten Teil das Recht ergänzende Maßnahmen aus den

Bereichen Technologie (§ 1) und Gesellschaft (§ 2) betrachtet, in deren Zusammenwirken eine angemessene Behandlung von Deep Fakes ermöglicht werden soll.

Der vierte Teil dieser Arbeit widmet sich mit dem Schutz der Wahrheit vor Deep Fakes den weit über das unmittelbar betroffene Individuum hinausreichenden Folgen von Deep Fakes. Dazu wird zunächst das Deep Fake als besondere Erscheinung von Mis- und Desinformation, mithin dessen besonderes täuschendes Potenzial betrachtet (§ 1). Sodann wird mit der Problematik des Deep Fakes vor Gericht und dem prozessualen Schutz von Wahrheit ein spezieller Aspekt der Täuschungsproblematik herausgegriffen (§ 2)

Obgleich die Deep Fake Problematik aufgrund der anhaltenden Diskussion in Politik und Gesellschaft allmählich (jedoch zunehmend intensiver[16]) auch ihren Weg in den rechtswissenschaftlichen Diskurs gefunden hat, beschränken sich die Ausführungen in der deutschsprachigen rechtswissenschaftlichen Literatur doch zumeist auf einzelne Aspekte von Deep Fakes und setzen sich insbesondere mit einzelnen der bereits angesprochenen Folgen von Deep Fakes für das unmittelbar betroffene Individuum auseinander.[17] Auch in der englischsprachigen juristischen Literatur findet bereits seit einiger Zeit eine ausführliche Diskussion verschiedenster mit der Deep Fake Problematik verbundener Fragestellungen statt.[18] Die weitergehenden Folgen von Deep Fakes, die über das einzelne Individuum hinausgehen und sich

16 Siehe etwa jüngst das Symposium des Instituts für Urheber- und Medienrecht mit dem Titel "Deep Fakes und das Recht: Medien- und urheberrechtliche Herausforderungen künstlicher Intelligenz - Chancen, Risiken & Regulierung", aus dem eine Reihe von Beiträgen hervorgingen. Siehe einleitend zur Problematik *Klass*, ZUM 2025, 485; sowie im Einzelnen *Völzmann*, ZUM 2025, 493; *Gomille*, ZUM 2025, 500; *Zurth*, ZUM 2025, 509.

17 Siehe etwa den allgemeinen Überblick zu den rechtlichen Herausforderungen im Zusammenhang mit Deep Fakes und potenziellen Lösungsansätzen bei *Lantwin*, MMR 2019, 574; *Kumkar/Rapp*, ZfDR 2022, 199; sowie speziell aus strafrechtlicher Perspektive *Lantwin,* MMR 2020, 78; *Erdogan*, MMR 2024, 379; Zu der speziellen strafrechtlichen Frage, inwieweit die Deep Fake Technologie als Mittel im Kampf gegen Kinderpornografie im Darknet eingesetzt werden darf, siehe *Wittmer/Steinebach*, MMR 2019, 650; Mit Blick auf die von Deepfakes ausgehenden Bedenken für den Einsatz bestimmter Technologien durch Sicherheitsbehörden siehe *Thiel*, ZRP 2021, 202; Zum speziellen persönlichkeitsrechtlichen Schutz vor Deepfakes durch das Kunsturheberrecht siehe *Hartmann,* Der Schutz vor Deepfakes durch das Kunsturhebergesetz, in: *Taeger,* Die Macht der Daten und der Algorithmen, Regulierung von IT, IoT und KI, 2019 S. 563 ff., 563ff.; Siehe weiterführend dazu auch *ders*., K&R 2020, 350; Ausführlich zu den Kennzeichnungspflichten nach der (künftigen) europäischen KI-Verordnung *Hinderks*, ZUM 2022, 110; sowie nunmehr auch *Block*, EuCML 2024, 184; Zu der speziellen Frage der Gefahren von Deep Fakes für Kapitalmärkte und deren Regulierung durch die KI-VO siehe *Tilson/Eichinger*, BKR 2024, 648.

18 Siehe etwa den sehr ausführlichen Überblick bei *Chesney/Citron*, California Law Review 107 (2019), 1753; sowie *Chesney/Citron*, Foreign Affairs Magazine 98 (2019), 147; *Chesney/Citron*, Maryland Law Review 78 (2019), 882; Siehe etwa auch *Blitz*, Oklahoma Law Review 71 (2018), 59; *Brown*, Va. J.L. & Tech. 23 (2020), 1; *Delfino*, Fordham Law Review 88 (2019), 887; *Farish*, Journal of Intellectual Property Law & Practice 15 (2020), 40; *Franks/Waldman*, Maryland Law Review 78 (2019), 892; *Kadri*, Maryland Law Review 78 (2019), 899; *Meskys* et al., Journal of Intellectual Property Law & Practice 15 (2020), 24; *Perot/Mostert*, Journal of Intellectual Property Law & Practice 15 (2020), 32; *Schroeder*, Syracuse Law Review 70 (2020), 1171; *Silbey/Hartzog*, Maryland Law Review 78 (2019), 960; *Meckel/Steinacker*, Morals & Machines 1 (2021), 10; *Fernandez*, UFITA 2021, 392; *Kim*, GRUR-Int. 2025, 532.

insbesondere im Zusammenhang mit der Diskussion um Mis- und Desinformation ergeben, wurden in der juristischen Literatur jedoch bislang nur ganz allmählich in den Blick genommen.[19] Anstoß für einen intensiveren dahingehenden Diskurs gab nun aber die spezifische Deep Fake Regulierung in der europäischen KI-Verordnung.[20]

Auch diese Arbeit intendiert keine umfassende Darstellung der Problematik, die aufgrund deren Vielschichtigkeit schlicht nicht möglich wäre. Zunächst erfolgt hier daher in tatsächlicher Hinsicht eine weitgehende Fokussierung auf den Bereich des visuellen Fakes, wobei bisweilen auch auf Besonderheiten im akustischen Bereich eingegangen werden soll. Während die Bereiche außerhalb des Rechts ohnehin nur im Ansatz dargestellt werden können, muss auch im Zusammenhang mit der rechtlichen Darstellung eine Schwerpunktsetzung erfolgen. Das Deep Fake wird in dieser Arbeit daher vorwiegend aus zivilrechtlicher Perspektive betrachtet, insbesondere die strafrechtliche Beurteilung bleibt anderen Forschungsarbeiten vorbehalten.[21] Auch der Bereich des Datenschutzes[22] bleibt weitgehend ausgeklammert, wobei jedoch einzelne Aspekte dessen zu betrachten sind, soweit sich Besonderheiten im Zusammenhang mit Deep Fakes ergeben. Auch die Problematik der Rechtsdurchsetzung kann nur im Ansatz dargestellt werden, da sich hier keine spezifische Gefahr von Deep Fakes verwirklicht. Zudem sollen hier auch spezifische Einsatzszenarien von Deep Fakes außen vor bleiben, die spezielle Fragestellungen hervorrufen, da das Deep Fake insoweit nur ein neues Werkzeug für bereits bekannte Probleme darstellt. Zuletzt kann auch die Desinformationsproblematik hier nicht umfassend dargestellt werden, ohnehin widmen sich der Problematik falscher und irreführender Informationen im Allgemeinen unter dem „Deckmantel" insbesondere der Fake News-Diskussion bereits einige Publikationen aus verschiedensten Blickwinkeln,[23]

19 Siehe dazu zunächst nur *Chesney/Citron*, California Law Review 107 (2019), 1753 (1776 ff.); *Mafi-Gudarzi*, ZRP 2019, 65 (65); aus interdisziplinärer Perspektive nunmehr auch *Hägle* et al., Policy & Internet 17 (2025), 1.

20 Siehe zur Deep Fake Regulierung der KI-Verordnung etwa *Becker*, CR 2024, 353; *Kumkar/Griesel*, KIR 2024, 117; *Block*, EuCML 2024, 184; *Łabuz*, Policy & Internet 16 (2024), 783; auf Grundlage der Entwurfsfassung auch bereits *Hinderks*, ZUM 2022, 110; *Kalbhenn*, ZUM 2021, 663; *Kumkar*, K&R 2023 Beilage 1 zu Heft 10, 32.

21 Siehe jüngst etwa *Blocher*, KIR 2025, 225; *Erdogan*, MMR 2024, 379; *Vassilaki*, CR 2024, 701; zuvor etwa auch bereits *Lantwin*, MMR 2020, 78.

22 Dazu aber etwa *Buchholz/Kremer*, CR 2025, 56 (58 ff.); *Martiny*, ZUM 2025, 200 (203 ff.).

23 Ausführlich widmen sich der Problematik aus einem interdisziplinären Blickwinkel beispielsweise *Steinebach* et al. - Desinformation aufdecken und bekämpfen, 2020, 1ff; Ein sehr ausführlicher Überblick zu den Risiken und Regulierungsmöglichkeiten in Bezug auf "Desinformation" findet sich auch bei *Dreyer* et al. - Desinformation: Risiken, Regulierungslücken und adäquate Gegenmaßnahmen, 1ff., https://www.medienanstalt-nrw.de/fileadmin/user_upload/NeueWebsite_0120/Themen/Desinformation/Leibnitz-Institut_LFMNRW_GutachtenDesinformation.pdf [zuletzt geprüft am 22.06.2025]; Aus strafrechtlicher Perspektive siehe etwa *Kusche,* „Fake News" – ein Fall für den Strafgesetzgeber?, in: *Beck/Kusche/Valerius,* Digitalisierung, Automatisierung, KI und Recht, Festgabe zum 10-jährigen Bestehen der Forschungsstelle RobotRecht, 2020 S. 421 ff., 421ff.; *Hoven*, ZStW 129 (2017), 718; *Schreiber* - Strafbarkeit politischer Fake News, 2021, 1ff; Siehe außerdem zahlreiche weitere einschlägige Publikationen, von denen hier beispielhaft

weshalb die allgemeine Thematik hier im Einzelnen nicht näher ausgebreitet werden soll. Mit dem prozessualen Wahrheitsschutz findet daher im Rahmen dieser Arbeit nur eine Auseinandersetzung mit einzelnen Facetten der Desinformationsproblematik anhand des Beispiels des Deep Fakes statt.

nur einige genannt werden können: *Mafi-Gudarzi*, ZRP 2019, 65; *Holznagel*, MMR 2018, 18 (18 ff.); *Lazer* et al., Science 359 (2018), 1094 (1094 ff.); *Flint* - Fake News im Wahlkampf, 1. Aufl. 2021.

1. Teil: Vom Original zum Deep Fake: Das visuelle Fake als besondere Herausforderung für das Recht im Lichte technologischen Fortschritts

Die spezielle Konstellation der Deep Fakes soll hier zum Anlass genommen werden, um zu untersuchen, wie mit Täuschungen durch bildhafte Manipulationen, insbesondere in Form der Manipulation bestehender (audio-)visueller Inhalte und durch „vollständig synthetische“[24] Generierung von Bildinhalten umgegangen wird, um einen angemessenen Interessenausgleich zu erzielen. Dabei erscheint es zunächst angezeigt, sich mit dem Begriff des *Deep Fake* auseinanderzusetzen, um die Problemstellung einzugrenzen. Bei dem Begriff *Deep Fake* handelt es sich um einen Neologismus, der sich aus den beiden Bestandteilen *Deep* und *Fake* zusammensetzt. In einem ersten Schritt (§ 1) soll sogleich der erste Begriffsbestandteil *Deep* untersucht werden. Dieser Teilbegriff geht zurück auf die Technologie, die sich hinter dem Phänomen *Deep Fake* verbirgt: Deep Learning. Im Anschluss an die Untersuchung der *Deep Fake Technologie* soll sodann in einem zweiten Schritt (§ 2) genauer beleuchtet werden, wie ein solches Deep Learning-basiertes Erzeugnis zu einem *Fake* wird.

§ 1 Technische Hintergründe: Deep Learning als Grundlage des modernen Fakes

I. KI, Künstliche Neuronale Netze und Deep Learning

Noch vor wenigen Jahren war es nur mit Hilfe von teurem Spezialequipment, aufwändiger CGI-Technik und Expertenwissen möglich, realistische und überzeugende Bild- und Videofälschungen zu erstellen. Doch ist durch neue technische Lösungen im Bereich der Bild- und Videoerstellung und -verarbeitung sowie durch die zunehmend verfügbaren Rechenkapazitäten jüngst eine rasante Entwicklung dahingehend zu beobachten, dass fortschreitend weder besondere Kenntnisse im Hinblick auf Bild- und Videobearbeitung, noch spezielle Ausstattung und ein besonderer Aufwand erforderlich sind, um überzeugende Bild- und Videofälschungen zu erzeugen, sodass zunehmend auch Laien Zugang zu solchen Techniken erlangen. Wie viele Fortschritte im gegenwärtigen Informationszeitalter ist auch diese Entwicklung im Wesentlichen auf „Künstliche Intelligenz“ („KI“) zurückzuführen.

24 Sofern zur Generierung dieser Inhalte Modelle des Maschinellen Lernens, wie insbesondere Generative Adversarial Networks und Diffusion Models, eingesetzt werden, kann aufgrund der Abhängigkeit des Outputs der Modelle von der Vielzahl an Daten, auf die das Modell zum Zwecke des Trainings angewiesen ist, gleichwohl nicht wirklich von einer synthetischen Genese ausgegangen werden. Siehe dazu insbesondere im 2. Teil dieser Arbeit unter § 3 im Zusammenhang mit der urheberrechtlichen Beurteilung dieser Modelle.

1. Überblick KI: Wie das „Künstliche" zunehmend „intelligenter" wird

Ausgehend vom allgemeinen Sprachverständnis beschreibt der Begriff *Künstliche Intelligenz*[25] eine Maschine, welche in der Lage ist, intelligentes menschliches Verhalten zu imitieren. Darauf aufbauend scheint der Begriff der *Künstlichen Intelligenz* in vielen Köpfen die Assoziation eines menschenähnlichen Roboters zu wecken, welcher in der Lage ist, sich selbstbestimmt und intelligent zu verhalten und potenziell sogar die menschliche Intelligenz zu übertreffen vermag. Von derartigen Vorstellungen, wie wir sie auch aus den bekannten Science-Fiction-Filmen[26] kennen, in denen diese Vorstellungen vielfach noch überzeichnet werden, indem wahlweise das Narrativ der gefährlichen Künstlichen Intelligenz bestimmend ist, die aufgrund ihrer Überlegenheit destruktive Tendenzen entwickelt bzw. von den Menschen zu derartigen Zwecken eingesetzt wird, oder im Gegenteil dazu die Künstliche Intelligenz als menschenähnliche, gefühlvolle und liebenswürdige Maschine erzählt wird, sind wir jedoch – trotz bemerkenswerter technologischer Fortschritte in den letzten Jahren – nach dem aktuellen Stand der Technik noch einige Schritte entfernt.[27] Die Technologien, welche

25 Da diese Systeme nicht wirklich "intelligent" sind (siehe dazu sogleich in den nachfolgenden Abschnitten), sondern Intelligenz nur simulieren, müsste der Begriff "Künstliche Intelligenz" bzw. "KI" eigentlich weiterhin in Anführungsstriche gesetzt werden. Aus Gründen der besseren Lesbarkeit wird im Folgenden jedoch darauf verzichtet und der Begriff als Eigenname verwendet. Wie hier auch *Buyx/Nida-Rümelin/Simon* - Statements: Pressekonferenz anlässlich der Veröffentlichung der Stellungnahme "Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz", 2, https://www.ethikrat.org/fileadmin/PDF-Dateien/Pressekonferenzen/pk-2023-03-20-statements.pdf [zuletzt geprüft am 22.06.2025]; vgl. daher auch *Deutscher Ethikrat* - Mensch und Maschine – Herausforderungen durch Künstliche Intelligenz, https://www.ethikrat.org/fileadmin/Publikationen/Stellungnahmen/deutsch/stellungnahme-mensch-und-maschine.pdf [zuletzt geprüft am 22.06.2025].

26 Zum Bild der "Künstlichen Intelligenz" im Film siehe *Seeßlen*, PuK 21 (2023), 24 (24).

27 Vgl. in diesem Sinne, der wohl noch (!) herrschenden Meinung etwa *Bitkom* - Künstliche Intelligenz, 31, https://www.dfki.de/fileadmin/user_upload/import/9744_171012-KI-Gipfelpapier-online.pdf [zuletzt geprüft am 22.06.2025]; *Buxmann/Schmidt* - Künstliche Intelligenz, 2019, 6; *Kaplan* - Künstliche Intelligenz, 1. Aufl. 2017, 23ff; *Volland* - Die kreative Macht der Maschinen, 1. Aufl. 2018, 14; siehe in dem Zusammenhang etwa auch den Überblick über verschiedene Prognosen in Bezug auf Artificial General Intelligence: *Heaven* - Artificial general intelligence: Are we close, and does it even make sense to try?, MIT Technology Review, 15.10.2020, https://www.technologyreview.com/2020/10/15/1010461/artificial-general-intelligence-robots-ai-agi-deepmind-google-openai/ [zuletzt geprüft am 22.06.2025]; In dem Zusammenhang kann auch eine grundsätzliche Tendenz der Wissenschaft dahingehend beobachtet werden, dass man weiteren Fortschritten im Bereich Künstlicher Intelligenz, auch mit Blick auf die Entwicklung multifunktionaler Systeme Künstlicher Intelligenz zunehmend optimistisch gegenübersteht, vgl. etwa die Studie des Projekts AI Impact von 2023 im Vergleich zu der vorangehenden aus dem Jahr 2022: *Grace* et al. - Thousands of AI Authors on the Future of AI, http://arxiv.org/pdf/2401.02843.pdf [zuletzt geprüft am 22.06.2025]; *Grace* et al. - 2022 Expert Survey on Progress in AI, https://wiki.aiimpacts.org/doku.php?id=ai_timelines:predictions_of_human-level_ai_timelines:ai_timeline_surveys:2022_expert_survey_on_progress_in_ai [zuletzt geprüft am 22.06.2025]; vgl. auch *Bengio* et al., Science 384 (2024), 842, die jedoch vor den Risiken dieser Entwicklungen warnen und regulative Schritte anmahnen; einzelne Autoren prognostizieren gar das Erreichen von Artificial General Intelligence bereits in nicht allzu ferner Zukunft. In diesem Sinne etwa auf Grundlage ökonomischer Modellierung *Macey-Dare* - How Soon is Now? Predicting the Expected Arrival Date of AGI - Artificial General Intelligence, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4496418 [zuletzt geprüft am 22.06.2025]; Für eine Darstellung verschiedener existierender Definitionen von Artificial General Intelligence, die Entwicklung verschiedener Stufen von Artificial General Intelligence und die Einordnung bestehender Systeme in diese Systematik siehe *Morris* et al. - Position:

heute ihren Siegeszug feiern, sind im Gegensatz zu einer solchen „Superintelligenz", die universell einsetzbar ist und der natürlichen Intelligenz in nichts nachsteht bzw. diese möglicherweise gar zu übertreffen vermag, vielmehr „simple" Expertensysteme in Form von Modellen, welche zur Lösung bestimmter abgrenzbarer Probleme geschult und eingesetzt werden, dabei jedoch mittlerweile ein erstaunliches Leistungsvermögen aufweisen.[28] Diese verschiedenen künstlich intelligenten Systeme werden für gewöhnlich anhand ihrer unterschiedlichen Grade von Autonomie differenziert und in die Kategorien schwache und starke Künstliche Intelligenz eingeteilt.[29] Daneben bzw. anstelle von starker Künstlicher Intelligenz ist bisweilen auch die Rede von Artificial General Intelligence.[30] Zugleich werden mit der Einteilung in starke und schwache KI vielfach auch zwei unterschiedliche Hypothesen verbunden und zwar einerseits, dass eine Maschine agieren könne als sei sie intelligent, intelligentes menschliches Verhalten also imitieren könne (schwache KI) und andererseits, dass eine solche Maschine tatsächlich über Intelligenz verfüge (starke KI).[31] Als konstituierend für starke Künstliche Intelligenz wird insbesondere die Ausbildung eines Bewusstseins, von Empathie, Kreativität oder weiteren kognitiven, genuin „menschlichen" bzw. „natürlichen" Fähigkeiten und Eigenschaften angesehen.[32] Obgleich Künstliche Intelligenz zunehmend in der Lage ist eben jene Tests zu bestehen, die entwickelt wurden, um die Ausbildung

Levels of AGI for Operationalizing Progress on the Path to AGI, http://arxiv.org/pdf/2311.02462.pdf [zuletzt geprüft am 22.06.2025].

28 Bereits 2017 "besiegte" mit AlphaZero ein Algorithmus des Maschinellen Lernens die Strategiespiele Schach, Shōgi und Go, *Silver* et al., Science 362 (2018), 1140; im Frühjahr 2023 wurde berichtet GPT-4 sei in der Lage das amerikanische Bar Exam zu bestehen *Katz* et al. - GPT-4 Passes the Bar Exam, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4389233 [zuletzt geprüft am 22.06.2025]; einschränkend jedoch *Martínez* - Re-Evaluating GPT-4's Bar Exam Performance, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4441311 [zuletzt geprüft am 22.06.2025]; ähnliche Ergebnisse wurden in verschiedenen anderen Tests erzielt. Darüber berichtend etwa *Metz/Collins* - 10 Ways GPT-4 Is Impressive but Still Flawed, The New York Times, 14.03.2023, https://www.nytimes.com/2023/03/14/technology/openai-new-gpt4.html [zuletzt geprüft am 22.06.2025]; erst jüngst gelang es zudem einem KI-Modell Mathematikaufgaben auf dem Niveau eines Goldmedaillengewinners der Internationalen Mathematik-Olympiade zu lösen, *Trinh* et al., Nature 625 (2024), 476; Das Lösen mathematischer Aufgaben wird als ein wichtiger Schritt auf dem Weg hin zu einer starken bzw. allgemeinen Künstlichen Intelligenz angesehen, da dies logisches Denken voraussetzt, vgl. etwa *Kim* - Google DeepMind's new AI system can solve complex geometry problems, MIT Technology Review, 17.01.2024, https://www.technologyreview.com/2024/01/17/1086722/google-deepmind-alphageometry/ [zuletzt geprüft am 22.06.2025].

29 Die Differenzierung in "weak AI" und "strong AI" geht zurück auf *Searle*, Behavioral and Brain Sciences 3 (1980), 417.

30 Eine dahingehende Differenzierung in "general AI" und "narrow AI" vornehmend etwa *Goertzel,* Toward a formal characterization of real-world general intelligence, in: *Baum/Hutter/Kitzelmann,* Artificial General Intelligence, March 5-8, 2010 S. 19 ff., 19ff.; siehe dazu etwa auch *Pennachin/Goertzel* - Contemporary Approaches to artificial general intelligence, in: *Goertzel/Pennachin,* Artificial General Intelligence, 2007 S. 1 ff., 1ff.; *Buxmann/Schmidt,* Künstliche Intelligenz, 6f; Zu verschiedenen Definitionen und Stufen von AGI siehe ausführlich *Morris* et al. - Position: Levels of AGI for Operationalizing Progress on the Path to AGI, http://arxiv.org/pdf/2311.02462.pdf [zuletzt geprüft am 22.06.2025].

31 *Russell* et al. - Artificial intelligence, 4. Aufl. 2022, 1032; *Kaplan,* Künstliche Intelligenz, 82f; *Käde* - Kreative Maschinen und Urheberrecht, 1. Aufl. 2021, 41.

32 Vgl. etwa *Searle*, Behavioral and Brain Sciences 3 (1980), 417; sowie *Pennachin/Goertzel*, in: *Goertzel/Pennachin*, Artificial General Intelligence, 2007, 6ff; *Buxmann/Schmidt,* Künstliche Intelligenz, 6.

von Intelligenz durch maschinelle Systeme zu beurteilen und natürliche von künstlicher Intelligenz zu unterscheiden,[33] geht man davon aus, dass selbst die führenden KI-Modelle bislang kein Bewusstsein aufweisen,[34] auf Grundlage der heutigen technischen Möglichkeiten eine „echte" Künstliche Intelligenz in dem Sinne der zuvor beschriebenen starken Künstlichen Intelligenz bzw. Artificial General Intelligence mithin noch einige Schritte entfernt ist. Gleichwohl konnten in der jüngsten Vergangenheit einige interessante Fortschritte im Hinblick auf derartige Zielvorstellungen erzielt werden,[35] und auch in die Agenden politischer Entscheidungsträger finden derartige Zukunftsprognosen zunehmend Eingang.[36]

33 Bereits 1950 etwa entwickelte Alan Turing das später nach ihm als Turing-Test benannte Experiment "The imitation game", *Turing*, Mind LIX (1950), 433; Der Turing-Test besteht in seiner weiterentwickelten Form darin, dass ein Mensch eine Unterhaltung einerseits mit einem Menschen und andererseits mit einer Maschine führt, ohne dabei Sicht- oder Hörkontakt zu den beiden zu haben und beurteilen muss, bei welchem Gegenüber es sich um den Menschen handelt und bei welchem um die Maschine. Sofern es der Maschine gelinge über ihre Eigenschaft zu täuschen, habe sie den Turing-Test bestanden, vgl. etwa *Lenzen* - Künstliche Intelligenz, 2018a, 25; Die Grenzen des Turing-Tests konnten jedoch bereits mithilfe des von Searle entwickelten - und seinerseits kritisierten - Gedankenexperiments des Chinesischen Zimmers aufgezeigt werden. In diesem Gedankenexperiment befindet sich eine Person in einem abgeschlossenen Raum mit einem Skript in chinesischer Sprache, derer sie nicht mächtig ist. Zusätzlich erhält die Person weitere Texte in chinesischer Sprache sowie ein Regelwerk für chinesische Schriftzeichen (in englischer Sprache, derer die Person mächtig ist), welches es ermöglicht den zweiten Satz Schriftzeichen mit dem ersten Satz in Beziehung zu setzen. Auf Grundlage des Regelwerks ist es der Person möglich die Schriftzeichen der beiden Texte zu vergleichen. Zusätzlich erhält die Person nun einen dritten Satz chinesischer Schriftzeichen, die Fragen, sowie eine Erklärung in englischer Sprache, die es ermöglicht, diesen dritten Satz Schriftzeichen in Beziehung zu den ersten beiden Sätzen zu setzen und die die Person dahingehend instruieren, wie sie bestimmte chinesische Symbole als Antwort auf den dritten Satz chinesischer Symbole nach draußen gibt. Nach Searle wirke es für die Außenstehenden zwar so als beherrsche die Person in dem Raum die Sprache, tatsächlich lerne diese Person jedoch kein Chinesisch, sondern verhalte sich nur wie ein Computer, der vorgegebene Rechenoperationen durchführt und so zu den Ergebnissen gelangt, *Searle*, Behavioral and Brain Sciences 3 (1980), 417; Für eine vereinfachte Darstellung siehe etwa *Lenzen* - Künstliche Intelligenz, 1. Aufl. 2020b, 66f. Weitere Tests sind etwa der alternate uses task-Test zur Feststellung von Kreativität sowie der Spiegel-Test zur Beurteilung von Bewusstsein; siehe im Zusammenhang mit der Beurteilung von Bewusstsein in Systemen Künstlicher Intelligenz jüngst etwa auch *Kosinski*, PNAS 121 (2024).

34 Vgl. etwa *Butlin* et al. - Consciousness in Artificial Intelligence: Insights from the Science of Consciousness, http://arxiv.org/pdf/2308.08708.pdf [zuletzt geprüft am 22.06.2025]; siehe dazu auch *Morris* et al. - Position: Levels of AGI for Operationalizing Progress on the Path to AGI, http://arxiv.org/pdf/2311.02462.pdf [zuletzt geprüft am 22.06.2025]; letztlich stellt sich die Frage, ob ein Nachweis von Bewusstsein überhaupt möglich ist, vgl. bereits *Nagel*, The Philosophical Review 83 (1974), 435; Gleichwohl ist es einem KI-System in Gestalt eines Large Language Models gelungen einen Entwickler bei Google davon zu überzeugen, dass es ein Bewusstsein entwickelt hätte. Darüber berichtend etwa *Tiku* - The Google engineer who thinks the company's AI has come to life, The Washington Post, 11.06.2022, https://www.washingtonpost.com/technology/2022/06/11/google-ai-lamda-blake-lemoine/ [zuletzt geprüft am 22.06.2025].

35 Erst jüngst gab es etwa Gerüchte über die Entwicklung eines neuen Modells namens Q* durch OpenAI, welches in der Lage sein soll mathematische Probleme zu lösen, mithin ein wichtiger Schritt auf dem Weg zur Entwicklung einer AGI sein könnte, *Tong/Dastin/Hu* - OpenAI researchers warned board of AI breakthrough ahead of CEO ouster, sources say, Reuters Media, 23.11.2023, https://www.reuters.com/technology/sam-altmans-ouster-openai-was-precipitated-by-letter-board-about-ai-breakthrough-2023-11-22/ [zuletzt geprüft am 22.06.2025]; *Heikkilä* - Unpacking the hype around OpenAI's rumored new Q* model, MIT Technology Review, 27.11.2023, https://www.technologyreview.com/2023/11/27/1083886/unpacking-the-hype-around-openais-rumored-new-q-model/ [zuletzt geprüft am 22.06.2025]. Siehe zum mathematischen Verständnis auf der Grundlage logischen Denkens auch bereits oben Fn. 28.

36 So etwa im Zusammenhang mit dem AI Safety Summit unter britischer Leitung, vgl. *UK Government* - The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023, https://www.

Wenn im Folgenden die Rede von Künstlicher Intelligenz ist, sollen damit jedoch nur die zuvor als schwache bzw. spezielle KI bezeichneten Expertensysteme angesprochen sein, da es derartige Systeme sind, auf welche die aktuellen Fortschritte im Zusammenhang mit Künstlicher Intelligenz zurückgehen und es auch eben jene Systeme sind, die der Generierung von Deep Fakes zugrunde liegen.

2. Eine kurze Geschichte der Künstlichen Intelligenz

Die Technologien, die unter dem Begriff *Künstliche Intelligenz* vereint werden und von denen hier die Rede ist, sind keine neuen; Forschung im Bereich der Künstlichen Intelligenz gibt es schon seit etwa 70 Jahren.[37] Als „Geburtsstunde der Künstlichen Intelligenz" wird allgemein (spätestens) das *Dartmouth Summer Research Project on Artificial Intelligence* aus dem Jahr 1956 angesehen.[38] Die Forschenden hatten es sich zum Ziel gesetzt in einem begrenzten Zeitraum und mit einer kleinen Gruppe von Forschenden verschiedene Aspekte des Lernens und weitere Merkmale menschlicher Intelligenz mithilfe einer Maschine zu simulieren, um Fortschritte in verschiedenen Bereichen erzielen zu können.[39] Nach einer anfänglichen „KI-Euphorie" in den 50er Jahren des letzten Jahrhunderts folgte ein „KI-Winter", der mit mehr und weniger großen Unterbrechungen erst zu Beginn des neuen Jahrtausends als endgültig überwunden angesehen werden kann.[40] In den letzten Jahren hat die KI-Forschung insbesondere in Gestalt des Maschinellen Lernens, speziell in Form des Deep Learnings, neuen Aufschwung erhalten und Fortschritte in einer Vielzahl verschiedener Forschungsbereiche möglich gemacht. Insbesondere die neuen gene-

gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 [zuletzt geprüft am 14.06.2025]; vgl. auch *The White House (Biden-Harris Administration)* - Executive Order 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, https://bidenwhitehouse.archives.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/ [zuletzt geprüft am 22.06.2025].

37 *Mainzer* - Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2019, 10f.

38 Siehe weiterführend zur Geschichte der KI: *Russell* et al., Artificial intelligence, 35ff; *Kaplan,* Künstliche Intelligenz, 27ff; *Buxmann/Schmidt,* Künstliche Intelligenz, 3; *Ertel* - Grundkurs Künstliche Intelligenz, 2016, 6ff; *Görz/Schmid/Wachsmuth* - § 1: Einleitung, in: *Görz/Schneeberger/Schmid,* Handbuch der Künstlichen Intelligenz, 2013 S. 1 ff., 2ff.; *Mainzer* - KI - Künstliche Intelligenz, 2003, 16ff.

39 *McCarthy* et al., A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE, 04.04.1996, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html [zuletzt geprüft am 22.06.2025].

40 Vgl. etwa *Görz/Schmid/Wachsmuth*, in: *Görz/Schneeberger/Schmid*, Handbuch der Künstlichen Intelligenz, 2013, 4ff; *Buxmann/Schmidt,* Künstliche Intelligenz, 3ff; *Mainzer,* Künstliche Intelligenz – Wann übernehmen die Maschinen?, 11ff; *Lenzen,* Künstliche Intelligenz, 24; *Russell* et al., Artificial intelligence, 35ff.

rativen KI-Modelle – etwa DALL-E[41], Midjourney[42], Stable Diffusion[43], Imagen[44] oder FLUX.1[45] von Black Forest Labs zur Generierung von Bildern aus Texteingaben und zur Sprachverarbeitung in Form sog. Chatbots – allen voran ChatGPT[46] sowie Bard bzw. nunmehr Gemini[47] – erzielen beeindruckende Ergebnisse, die auch für die allgemeine Öffentlichkeit wahrnehmbar und von unmittelbarem Nutzen für diese sind.

Dass diesen gar nicht mehr so jungen Technologien trotz allem erst in den letzten Jahren der Durchbruch gelingt, liegt insbesondere darin begründet, dass die Effizienz von Systemen Künstlicher Intelligenz wesentlich durch die Menge der verfügbaren Daten bedingt ist, welche im Vorfeld in die Modelle eingespeist werden müssen, um diese im Hinblick auf ein bestimmtes Ziel trainieren zu können. Die Verfügbarkeit großer Datenmengen wurde insbesondere durch die weite Verbreitung des Internets und auch durch die breite Nutzung bestimmter Systeme, etwa Suchmaschinen und sozialer Netzwerke ermöglicht, sind es doch im Wesentlichen die Unternehmen, die hinter eben jenen Plattformen stehen, die nun auch über die erforderlichen Datenmassen verfügen, um die besonders leistungsstarken KI-Modelle trainieren zu können.[48] Wenn derart große Mengen von Daten (*BigData*)[49] verarbeitet werden sollen, müssen jedoch auch eine Vielzahl von Rechenoperationen durchgeführt werden, was wiederum eine Steigerung der benötigten Rechenkapazitäten bedingt. Wesentlich für den aktuell zu beobachtenden Erfolg von KI-Systemen und insbesondere von Algorithmen des Maschinellen Lernens sind also zum einen die gigantische Menge der mittlerweile zur Verfügung stehenden Daten, und zum anderen die Tatsache, dass heute – anders als noch zu Zeiten der Anfänge des Forschungszweigs der Künstlichen Intelligenz – ausreichend leistungsfähige Prozessoren zur Verfügung stehen, um die benötigten Datenmengen in einem angemessenen Zeitraum zu verarbeiten.[50]

41 DALL·E 2, https://openai.com/dall-e-2/ [zuletzt geprüft am 22.06.2025].
42 Midjourney, https://www.midjourney.com/home/ [zuletzt geprüft am 14.06.2025].
43 Stable Diffusion 2.0 Release, 24.11.2022, https://stability.ai/blog/stable-diffusion-v2-release [zuletzt geprüft am 22.06.2025].
44 Imagen: Text-to-Image Diffusion Models, https://imagen.research.google/ [zuletzt geprüft am 22.06.2025].
45 Black Forest Labs - FLUX.1, 01.08.2024, https://blackforestlabs.ai/announcing-black-forest-labs/ [zuletzt geprüft am 22.06.2025].
46 Introducing ChatGPT, 30.11.2022, https://openai.com/blog/chatgpt [zuletzt geprüft am 22.06.2025].
47 *Hsiao,* Bard heißt jetzt Gemini, 08.02.2024, https://blog.google/intl/de-de/unternehmen/technologie/bard-gemini-advanced-app/ [zuletzt geprüft am 22.06.2025].
48 Siehe zum Ansatz der Human-Aided AI, die sich feedback loops zunutze macht, sowie zu den zugrunde liegenden Machtverhältnissen, insbesondere aus ethischer Sicht *Mühlhoff* - Die Macht der Daten, 2023, 6ff.
49 Weiterführend zur Bedeutung von BigData für die Fortschritte im Bereich der KI-Forschung: *Russell* et al., Artificial intelligence, 44.
50 *Buxmann/Schmidt,* Künstliche Intelligenz, 7f; *Grossenbacher,* Verblüffende Videofälschungen - Von Magie nicht mehr zu unterscheiden, 17.08.2018, https://www.srf.ch/news/panorama/verblueffende-videofaelschungen-von-magie-nicht-mehr-zu-unterscheiden [zuletzt geprüft am 22.06.2025]; *Bitkom* - Künstliche

3. Begriffsbestimmung Künstliche Intelligenz: Versuch der Definition des Undefinierbaren

Der Begriff der *Künstlichen Intelligenz* bzw. *Artificial Intelligence* geht zurück auf das Thema der Konferenz am Dartmouth College im Sommer 1956, in deren Förderungsantrag *Artificial Intelligence* in etwa als die Simulation von Eigenschaften von Intelligenz wie unter anderem von verschiedenen Aspekten des Lernens eingeführt wird.[51] Dieser Definitionsversuch wie weitere ähnlich gelagerte Ansätze bleiben jedoch zirkulär, da sie jeweils auf den Begriff der *intelligence* bzw. *Intelligenz* rekurrieren und daher die Definitionsproblematik nur verschieben.[52] Gerade eine klare Umgrenzung des Teilbegriffs der *Intelligenz*[53] ist jedoch, wenn nicht gar unmöglich, so doch jedenfalls problematisch, sodass eine präzise Definition von *Künstlicher Intelligenz* bereits aus diesem Grund kaum erreichbar scheint.[54] Überdies ist eine eng umgrenzte Begriffsbestimmung von *Künstlicher Intelligenz*, die den Anspruch auf Allgemeingültigkeit erhebt, auch nicht zielführend, denn es handelt sich um ein sehr weites und sich ständig weiterentwickelndes Forschungsgebiet, das aus einer Vielzahl verschiedener Perspektiven betrachtet werden kann.[55] Gleichwohl erscheint es in einigen Bereichen angezeigt, eine Begriffsbestimmung von *Künstlicher Intelligenz* für bestimmte Zwecke zu entwickeln; dies gilt insbesondere auch im Bereich des Rechts. Im Folgenden kann und soll daher keine umfassende Darstellung sämtli-

Intelligenz, 26f., https://www.dfki.de/fileadmin/user_upload/import/9744_171012-KI-Gipfelpapier-online.pdf [zuletzt geprüft am 22.06.2025].

51 *McCarthy* et al., A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE, 04.04.1996, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html [zuletzt geprüft am 22.06.2025].

52 Ausführlich dazu *Herberger*, NJW 2018, 2825 (2826); *McCarthy* - What is Artificial Intelligence?, 1ff., http://www-formal.stanford.edu/jmc/whatisai.pdf [zuletzt geprüft am 22.06.2025]; *Linke* - Urheberrechtlicher Schutz von „KI" als Computerprogramme – Squeezing today´s innovations into yesterday´s system?, in: *Hetmank/Rechenberg,* Kommunikation, Kreation und Innovation - Recht im Umbruch?, 2019 S. 29 ff., 31ff. [zuletzt geprüft am 25.08.2020].

53 Weiterführend zum englischen Begriff "intelligence": *Legg/Hutter*, Minds & Machines 17 (2007), 391 (391 ff.); für eine ausführliche Darstellung siehe *dies.* - A Collection of Definitions of Intelligence, in: *Goertzel/Wang,* Advances in artificial general intelligence: concepts, architectures and algorithms, 2007 S. 17 ff., 18ff.; *Pennachin/Goertzel*, in: *Goertzel/Pennachin*, Artificial General Intelligence, 2007, 6ff; Zum Verhältnis der englischen "intelligence" zur deutschen "Intelligenz" siehe *Herberger*, NJW 2018, 2825 (2826 f.); Die zahlreichen Stufen bzw. Ausprägungen von Intelligenz ausführlich darstellend sowie "künstliche" Intelligenz in diesen Kontext einordnend etwa *Otte* - Intelligenz und Bewusstsein, APuZ, 13.10.2023, https://www.bpb.de/shop/zeitschriften/apuz/kuenstliche-intelligenz-2023/541495/intelligenz-und-bewusstsein/ [zuletzt geprüft am 22.06.2025]. Zu den verschiedenen Tests zur Beurteilung von "Künstlicher Intelligenz" siehe bereits oben Fn. 33.

54 Vgl. zu den Schwierigkeiten bei der Suche nach einer klaren Definition des Begriffs der "Künstlichen Intelligenz": *Mainzer,* Künstliche Intelligenz – Wann übernehmen die Maschinen?, 2f; *Buxmann/Schmidt,* Künstliche Intelligenz, 6f; *Görz/Schmid/Wachsmuth*, in: *Görz/Schneeberger/Schmid*, Handbuch der Künstlichen Intelligenz, 2013, 2ff.

55 Siehe etwa zur Uneinheitlichkeit der Begriffsdefinition aufgrund der Weite des Forschungsfeldes: *Bitkom* - Künstliche Intelligenz, 28ff., https://www.dfki.de/fileadmin/user_upload/import/9744_171012-KI-Gipfelpapier-online.pdf [zuletzt geprüft am 22.06.2025]; vgl. auch *Buxmann/Schmidt,* Künstliche Intelligenz, 6.

cher Ansätze[56] zur Definition von *Künstlicher Intelligenz* erfolgen. Vielmehr soll im Anschluss an die Darstellung insbesondere von Ansätzen, die sich mit Künstlicher Intelligenz bzw. bestimmten Formen von KI als Regulierungsobjekt beschäftigen, für die Zwecke dieser Arbeit eine Arbeitsdefinition von *Künstlicher Intelligenz* gefunden werden, die eine Einordnung der hier angesprochenen KI-basierten Technologien in den Gesamtzusammenhang Künstliche Intelligenz ermöglicht.

Die verschiedenen Definitionsansätze lassen sich grundsätzlich in zwei verschiedene Kategorien einteilen: Primär normativ angelegte Ansätze versuchen Künstliche Intelligenz über deren Zielsetzung, welche in der Imitation intelligenten menschlichen Verhaltens besteht, zu definieren. Da eine Abgrenzung rein über die Zielsetzung von Systemen Künstlicher Intelligenz sich jedoch aus den oben genannten Gründen als schwierig erweist, werden derartige Ansätze vielfach zumindest über deskriptive Elemente ergänzt bzw. gar durch solche ersetzt.[57]

In Europa wird die Diskussion über Künstliche Intelligenz als Gegenstand von Regulierung unter dem Begriff des *KI-Systems* geführt. Da auch die vorliegende Untersuchung an Künstliche Intelligenz als Regulierungsobjekt anknüpft, sollen die europäischen Erwägungen zum KI-System[58] auch der vorliegenden Untersuchung zugrunde gelegt werden. Soweit ersichtlich wurde der Begriff des *KI-Systems* erstmals 2018 durch die Europäische Kommission in ihrer KI-Strategie[59] gebraucht. Die Europäische Kommission definiert Künstliche Intelligenz danach rein normativ als Systeme, die sich „intelligent" verhalten, ihre Umgebung analysieren und in bestimmtem Maße autonom handeln, um bestimmte Ziele zu erreichen.[60] Da diese Definition – wie grundsätzlich auch andere rein normative Ansätze – unter dem Rekurs auf den Begriff *intelligent* leidet, der über die Merkmale der Umgebungsanalyse und der Autonomie nur bedingt konkretisiert wird und dessen Bestimmbarkeit im Übrigen vorausgesetzt wird, nimmt die *High Level Expert Group on Artificial Intelligence*[61] eine Ergänzung dieser rein normativen Definition um deskriptive Elemente dahingehend vor, dass es sich bei Künstlicher Intelligenz um eine Disziplin handelt, welche bestimmte Technologien wie Machine Learning, Machine Reasoning und

56 Für einen Überblick über zahlreiche verschiedene Ansätze siehe *Legg/Hutter*, Minds & Machines 17 (2007), 391 (424 ff.); *dies.*, in: *Goertzel/Wang*, Advances in artificial general intelligence: concepts, architectures and algorithms, 2007, 21f.

57 Zur Beschreibung von "Künstlicher Intelligenz" über einen deskriptiven Ansatz: *Linke*, in: *Hetmank/Rechenberg*, Kommunikation, Kreation und Innovation - Recht im Umbruch?, 2019, 32ff; für die Zwecke ihrer Arbeit ebenfalls einen primär deskriptiven Ansatz bevorzugend *Käde,* Kreative Maschinen und Urheberrecht, 40f.

58 Siehe ausführlich zur Entwicklung des Begriffs des KI-Systems in Europa: *Borges*, CR 2023, 706 (707 ff.); sowie *Molavi Vasse'i*, KIR 2025, 190; zur Begriffsabgrenzung siehe ausführlich *Siemerling*, CR 2024, 554.

59 *Europäische Kommission* - Künstliche Intelligenz für Europa, COM(2018) 237 final, 1.

60 *Europäische Kommission* - Künstliche Intelligenz für Europa, COM(2018) 237 final, 1.

61 Die High-Level Expert Group on Artificial Intelligence wurde im Juni 2018 von der Europäischen Kommission als unabhängige Expertengruppe eingerichtet.

Robotik unter diesem Oberbegriff zusammenfasst.[62] Einen ähnlichen Ansatz der Kombination von normativen und deskriptiven Elementen verfolgt auch die Bundesregierung in ihrer KI-Strategie.[63]

Eine Ergänzung ihrer zunächst rein normativen Definition um deskriptive Elemente nimmt denn auch die Europäische Kommission in ihrem Vorschlag für ein Gesetz über Künstliche Intelligenz (KI-VO)[64] vor. Mit diesem Ansatz verfolgt sie das Ziel der Formulierung einer klaren und damit rechtssicheren und gleichzeitig hinreichend flexiblen (mit Blick auf die Anpassungsfähigkeit an neue technische Entwicklungen) Begriffsbestimmung für das *KI-System*.[65] Nach Art. 3 Nr. 1 KI-VO-E handele es sich bei einem „System der künstlichen Intelligenz" (KI-System) um *„eine Software, die mit einer oder mehreren der in Anhang I aufgeführten Techniken und Konzepte entwickelt worden ist und im Hinblick auf eine Reihe von Zielen, die vom Menschen festgelegt werden, Ergebnisse wie Inhalte, Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das Umfeld beeinflussen, mit dem sie interagieren"*. Weiterhin setzt die Kommission voraus, dass die Systeme dabei in unterschiedlichem Umfang autonom arbeiten.[66] Anhang I benennt als die in Art. 3 Nr. 1 KI-VO-E angesprochenen Techniken und Konzepte das Machine Learning, insbesondere in Form des Deep Learning (lit. a)), Logik- und wissensgestützte Konzepte (lit. b)) sowie statistische Ansätze (lit. c)). Diese Definition wird in der Literatur jedoch vielfach als zu weitgehend kritisiert, da sie – insbesondere aufgrund der Bezugnahme statistischer Ansätze – nahezu jegliche Form von Software erfasse,[67]

62 *High-Level Expert Group on Artificial Intelligence* - A Definition of AI: Main Capabilities and Disciplines, 7, https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf [zuletzt geprüft am 22.06.2025]: *"Artificial intelligence (AI) refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems)."*.

63 *Bundesregierung* - Strategie Künstliche Intelligenz der Bundesregierung, 4f., https://www.bundesregierung.de/breg-de/service/publikationen/strategie-kuenstliche-intelligenz-der-bundesregierung-2018-1551264 [zuletzt geprüft am 22.06.2025].

64 *Europäische Kommission* - Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für Künstliche Intelligenz (Gesetz über Künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union - KI-Verordnung (KI-VO), COM(2021) 206 final.

65 Vgl. Erw.Gr. 6 S. 1, 4 KI-VO-E.

66 Erw.Gr. 6 S. 3 KI-VO-E.

67 Siehe etwa *Bomhard/Merkle*, RDi 2021, 276 (277); *Engelmann/Brunotte/Lütkens*, RDi 2021, 317 (318); *Steege*, MMR 2022, 926; *Hacker/Berz*, ZRP 2023, 226 (227); den deskriptiven Ansatz grundsätzlich als positiv hervorhebend, jedoch ebenfalls eine weitere Konkretisierung fordernd *Roos/Weitz*, MMR 2021, 844 (845).

mithin die KI-Regulierung zur allgemeinen Software-Regulierung wird.[68] Durch den deskriptiven Ansatz können derartige Begriffsbestimmungen möglicherweise auch den dynamischen technischen Entwicklungen nicht gerecht werden. Außerdem fokussieren sich deskriptive Ansätze in der Regel zu sehr auf den Bereich des Maschinellen Lernens. Dies verleitet gar zu einer Gleichsetzung der Künstlichen Intelligenz mit dem Maschinellen Lernen, wie sie aktuell vielfach im öffentlichen Diskurs vorgenommen wird. Tatsächlich handelt es sich beim Maschinellen Lernen jedoch nur um einen Ausschnitt der Künstlichen Intelligenz.[69] Die Gleichsetzung der Begriffe *Künstliche Intelligenz* und *Maschinelles Lernen* ist daher zwar verkürzt, die synonyme Verwendung wird den praktischen Verhältnissen jedoch vielfach gerecht,[70] denn ein Großteil der Fortschritte im Bereich der Künstlichen Intelligenz kann auf das Maschinelle Lernen zurückgeführt werden. Um die im Hinblick auf deskriptive Ansätze angeführten Kritikpunkte zu vermeiden, findet in den Positionen von Parlament und Rat daher zunehmend eine Rückbesinnung auf eher normative Ansätze statt, die robuster im Hinblick auf die technische Entwicklung sind, gleichzeitig leiden sie jedoch auch unter dem Umstand, dass sie dazu in gewissem Maße immer auch wage bleiben (müssen).

Einen kombinierten Ansatz normativer und deskriptiver Elemente verfolgt daher der Rat in seiner allgemeinen Ausrichtung zu Künstlicher Intelligenz und führt dabei einerseits das Merkmal der Autonomie ein und begrenzt Künstliche Intelligenz andererseits auf die Verfahren des Maschinellen Lernens sowie logik- und wissensgestützte Konzepte, vgl. Art. 3 Nr. 1 EC-Mandat zum KI-VO-E.[71] Nach dem Vorschlag des Europäischen Parlaments, der sich stark an der aktualisierten Definition der OECD[72]

68 In diesem Sinne vor dem Hintergrund auch noch der Vorschläge von Rat und Parlament *Hacker/Berz*, ZRP 2023, 226 (227).

69 Vgl. *Alpaydin* - Maschinelles Lernen, 3. Aufl. 2022, 3; *Görz/Schneeberger/Schmid* - Handbuch der Künstlichen Intelligenz, 5. Aufl. 2013, 14.

70 *Käde,* Kreative Maschinen und Urheberrecht, 40; vgl. auch *Buxmann/Schmidt,* Künstliche Intelligenz, 7; vgl. auch *Staehelin*, GRUR 2022, 1569 (1569).

71 *Rat der Europäischen Union* - Allgemeine Ausrichtung zum Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz), 14954/22, 71.

72 *OECD* - Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 7, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [zuletzt geprüft am 22.06.2025]: „*An AI system is a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.*"; Nach der inoffiziellen deutschen Übersetzung: „*Ein KI-System ist ein maschinenbasiertes System, das für bestimmte von Menschen definierte Ziele Voraussagen machen, Empfehlungen abgeben oder Entscheidungen treffen kann, die das reale oder virtuelle Umfeld beeinflussen. KI-Systeme können mit einem unterschiedlichen Grad an Autonomie ausgestattet sein*"; *OECD* - Inoffizielle Übersetzung: Empfehlung des Rats zu künstlicher Intelligenz, OECD/LEGAL/0449, 5, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 [zuletzt geprüft am 22.06.2025]. Im Vergleich zu der englischen Originalfassung fehlt in der inoffiziellen deutschen Übersetzung der Verweis auf die Anpassungsfähigkeit des KI-Systems nach seiner Einführung. Da sich dieses Merkmal sowohl in der englischen als auch in der französischen Originalfas-

orientiert[73], handelt es sich bei einem KI-System um „*ein maschinengestütztes System, das so konzipiert ist, dass es mit unterschiedlichem Grad an Autonomie operieren kann und das für explizite oder implizite Ziele Ergebnisse wie Vorhersagen, Empfehlungen oder Entscheidungen hervorbringen kann, die das physische oder virtuelle Umfeld beeinflussen*".[74] Als entscheidendes Merkmal wurde mithin ein variierender Autonomiegrad eingeführt, welcher dann vorliegen solle, wenn das KI-System zumindest bis zu einem gewissen Grad unabhängig von menschlicher Kontrolle agiert und in der Lage ist, ohne menschliches Eingreifen zu arbeiten.[75] Allerdings wird bezweifelt, ob sich die Ziele des Parlaments auf diesem Wege erreichen lassen.[76]

Im Trilog[77] wurde schließlich eine Einigung in Bezug auf die Definition des *KI-Systems* dahingehend erzielt, dass es sich bei einem KI-System um ein maschinengestütztes System handelt, das für einen in unterschiedlichem Grade autonomen Betrieb ausgelegt ist und das nach seiner Betriebsaufnahme anpassungsfähig sein kann und das aus den erhaltenen Eingaben für explizite oder implizite Ziele ableitet, wie Ausgaben wie etwa Vorhersagen, Inhalte, Empfehlungen oder Entscheidungen erstellt werden, die physische oder virtuelle Umgebungen beeinflussen können. Diese Definition[78] findet sich nunmehr auch in Art. 3 Nr. 1 KI-VO[79]. Nach der Intention des Europäischen Gesetzgebers soll diese Definition des *KI-Systems* insbesondere eine Abgrenzung von Systemen der Künstlichen Intelligenz gegenüber einfachen bzw. gewöhnlichen Software-Systemen und Ansätzen ermöglichen, die auf durch den Menschen vorgegebenen Regeln beruhen und diese nur automatisiert ausfüh-

sung findet, scheint es sich dabei um ein redaktionelles Versehen in der inoffiziellen deutschen Übersetzung zu handeln.

73 Damit verfolgt das Europäische Parlament das Ziel den Begriff des KI-Systems klar zu definieren und eng mit KI-bezogenen Tätigkeiten internationaler Organisationen abzustimmen, um eine rechtssichere und harmonisierte Regulierung sicherzustellen, die hohe Akzeptanz genießt und gleichzeitig die hinreichende Flexibilität aufweist, um den technologischen Entwicklungen gerecht werden zu können, vgl. Erw.Gr. 6 S. 1 der Abänderung EP zu KI-VO-E.

74 *Europäisches Parlament* - Abänderungen des Europäischen Parlaments vom 14. Juni 2023 zu dem Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz), P9_TA(2023)0236 (Abänderung EP zu KI-VO-E), 131, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_DE.pdf [zuletzt geprüft am 22.06.2025]. In der englischen Fassung: „*artificial intelligence system' (AI system) means a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual environments*".

75 Vgl. Erw.Gr. 6 S. 3 Abänderung EP zu KI-VO-E.

76 *Becker/Feuerstack*, MMR 2024, 22 (23).

77 Zur juristischen Bewertung des Trilogergebnisses siehe *Bomhard/Siglmüller*, RDi 2024, 45.

78 Siehe zum Begriff des KI-Systems der KI-VO ausführlich *Wendehorst* et al., MMR-Beilage 2024, 605.

79 Verordnung (EU) 2024/1689 des Europäischen Parlaments und des Rates vom 13. Juni 2024 zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz und zur Änderung der Verordnungen (EG) Nr. 300/2008, (EU) Nr. 167/2013, (EU) Nr. 168/2013, (EU) 2018/858, (EU) 2018/1139 und (EU) 2019/2144 sowie der Richtlinien 2014/90/EU, (EU) 2016/797 und (EU) 2020/1828 (Verordnung über künstliche Intelligenz) - KI-Verordnung (KI-VO), in der Fassung der Bekanntmachung vom 12.07.2024, Amtsblatt der Europäischen Union 2024 Nr. L S. 1 ff.

ren.[80] Stattdessen werden bei Systemen Künstlicher Intelligenz die Ergebnisse aus den Eingaben abgeleitet,[81] etwa indem Verfahren des Maschinellen Lernens oder logik- und wissensbasierte Verfahren eingesetzt werden.[82] Entscheidende Merkmale sind einerseits die Fähigkeit der Systeme Schlüsse zu ziehen, die insbesondere eine Lernfähigkeit[83] und Anpassungsfähigkeit[84] ermöglicht, sowie dass die Systeme bis zu einem gewissen Grad unabhängig (autonom)[85] von menschlicher Beteiligung agieren. Der Europäische Gesetzgeber verweist für die Abgrenzung von Künstlicher Intelligenz gegenüber gewöhnlicher Software also auf die Funktionsweise der Systeme: Erstere leiten ihre Ergebnisse unmittelbar aus den Eingaben ab, wohingegen letztere auf durch den Menschen vorgegebene Regeln angewiesen sind. Weiterhin werden auf dieser Grundlage die Lernfähigkeit, Anpassungsfähigkeit sowie die Autonomie der KI-Systeme hervorgehoben.

Die praktische Handhabbarkeit dieser Definition wird sich bewähren müssen, sie vermeidet jedoch einerseits die zu starke Fokussierung auf bestimmte Formen Künstlicher Intelligenz, indem sie bestimmte Technologien nur als Anwendungsfälle benennt und keinen rein deskriptiven Ansatz verfolgt und andererseits den vielen normativen Ansätzen anhaftenden Rekurszirkel auf den Intelligenzbegriff, indem sie auf bestimmte Merkmale abstellt, die kennzeichnend für Systeme maschineller Intelligenz sind. Aufgrund dieser Vorteile eines solchen Ansatzes, soll die Begriffsbestimmung des *KI-Systems* in Art. 3 Nr. 1 KI-VO auch der vorliegenden Untersuchung zugrunde gelegt werden. Wenn also im Folgenden von *Künstlicher Intelligenz* die Rede ist, sollen damit Expertensysteme angesprochen sein, die die Lösung bestimmter vorgegebener und abgrenzbarer Probleme zum Ziel haben und diese Ziele – jedenfalls in Teilen – autonom verfolgen, indem sie sich zur Analyse ihrer Umwelt verschiedener Technologien wie insbesondere des Maschinellen Lernens bedienen und aus gegebenen Daten Ergebnisse ableiten und sich durch die Fähigkeiten der Lern- und Anpassungsfähigkeit auszeichnen.

4. Maschinelles Lernen

Wesentliche Fortschritte im Bereich der KI-Forschung können in den letzten Jahren auf die Technologie des Maschinellen Lernens[86] als spezielle Form Künstlicher Intel-

80 Erw.Gr. 12 S. 2 KI-VO.
81 Erw.Gr. 12 S. 3, 4 KI-VO.
82 Erw.Gr. 12 S. 5 KI-VO.
83 Erw.Gr. 12 S. 6 KI-VO.
84 Erw.Gr. 12 S. 12 KI-VO.
85 Erw.Gr. 12 S. 11 KI-VO.
86 Im Zusammenhang mit Maschinellem Lernen wird häufig auch der Begriff des Data Mining gebraucht. Hierbei handelt es sich um einen speziellen Anwendungsfall des Maschinellen Lernens, bei dem aus großen Datenmengen (Big Data), welche häufig in Form von Datenbanken vorliegen, im Wege des Maschinellen Lernens Wissen extrahiert wird, *Alpaydin,* Maschinelles Lernen, 2f. Siehe zu der Abgrenzung

ligenz zurückgeführt werden; Auch die hier zu diskutierenden Deep Fake Modelle sind eben jenem Bereich Künstlicher Intelligenz zuzuordnen. Im Gegensatz zur Entwicklung eines klassischen Algorithmus[87] wird die in einem Modell des Maschinellen Lernens – insbesondere des Deep Learnings – enthaltene Arbeitsanweisung im Grundsatz nicht von der Person vorgegeben, welche das Modell entwickelt, sondern dieses „lernt" selbstständig aus gegebenen Daten eine Lösung für ein vorgegebenes Problem zu extrahieren.[88] Das Lernen in einem Machine Learning Modell erfolgt anhand der Trainingsdaten, indem darin Muster erkannt werden und anschließend eine Generalisierung vorgenommen wird, um das Gelernte auch auf neue Fälle übertragen zu können.[89] Im Anschluss an den Lernvorgang ist das Machine Learning Modell (bestenfalls) in der Lage eine hinreichende Approximation zur Lösung des vorgegebenen Problems zu finden. Das Maschinelle Lernen, besonders in Form des Deep Learnings, hat sich insbesondere in denjenigen Bereichen als hilfreich erwiesen, in denen zwar keine ausreichende Wissensgrundlage beim Entwickelnden gegeben ist,[90] doch genügend Daten vorhanden sind, um einen Algorithmus darauf trainieren zu können.[91] Wenngleich der Mensch hierbei als Vorbild und Grundlage für die Maschine diente und sich die Modellierung des Maschinellen Lernens weiterhin grundsätzlich an den biologischen Prozessen orientiert, so bestehen doch Unterschiede zwischen dem menschlichen Lernprozess und dem Maschinellen Lernen, die dazu führen, dass die Maschine dem Menschen zwar in bestimmten Teilbereichen in der Lösung abgegrenzter Probleme ebenbürtig bzw. gar überlegen ist, sie jedoch bislang nicht das umfassende Potenzial des menschlichen Lernens zu erreichen vermag.[92]

---

der verschiedenen Begriffe (Künstliche Intelligenz, Machine Learning und (Text- und) Data-Mining) aus der Perspektive des Urheberrechts im Zusammenhang mit den urheberrechtlichen Schrankenregelungen zugunsten des Text- und Data-Minings im 2. Teil dieser Arbeit § 2 2.c.bb.(III)(2).

87 Unter einem Algorithmus versteht man eine Folge von Arbeitsanweisungen an den Computer, um aus einer Eingabe eine Ausgabe zu erzielen, *ders.,* Maschinelles Lernen, 2.

88 *Hartmann* - KI & Recht kompakt, 2020, 9f; *Alpaydin,* Maschinelles Lernen, 2; *Patel* - Praxisbuch unsupervised learning, 1. Aufl. 2020, 4f; Allerdings setzt das Algorithmendesign gleichwohl ein gutes Verständnis der Daten und der darin enthaltenen Strukturen voraus. Beim klassischen Maschinellen Lernen kommt daher dem Feature Engineering eine herausragende Bedeutung zu, vgl. *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436); *Mohri/Rostamizadeh/Talwalkar* - Foundations of machine learning, 2012, 4. Siehe in Abgrenzung dazu aber insbesondere das "Deep Learning" als spezielle Form des Maschinellen Lernens. Grundlegend LeCun et al. a.a.O. sowie näher dazu auch sogleich unter 6.

89 *Marsland* - Machine learning, 2015, 4.

90 Vgl. etwa auch *Russell* et al., Artificial intelligence, 669.

91 Vgl. *Alpaydin,* Maschinelles Lernen, 2.

92 Vgl. dazu sowie zu der Frage, wie Systeme des Maschinellen Lernens entwickelt werden können, die noch näher an das menschliche Vorbild heranreichen: *Lake* et al., Behavioral and Brain Sciences 40 (2017), e253; sowie *Tenenbaum* et al., Science 331 (2011), 1279; vgl. auch *Lenzen* - Der elektronische Spiegel, 2023, 159ff; zu dem Versuch der Nachbildung des Lernens von Sprache durch Kleinkinder mithilfe des Maschinellen Lernens siehe *Vong* et al., Science 383 (2024), 504; zum Vergleich der Leistungsfähigkeit des Menschen mit derjenigen des maschinellen Lernens in Bezug auf die Mustererkennung in bestimmten Konstellationen, in denen nur wenige Trainingsdaten zur Verfügung stehen, siehe etwa *Kühl* et al., Cognitive Systems Research 76 (2022), 78. Siehe zum Vergleich des Maschinellen Lernens mit dem menschlichen

Es lassen sich drei verschiedenen Arten des maschinellen Lernens unterscheiden: supervised learning, unsupervised learning und reinforcement learning.[93] Im klassischen Fall des supervised learning lernt ein Modell anhand der Rückmeldung einer Supervisionsinstanz darüber, ob das Modell die richtige Entscheidung getroffen hat oder nicht.[94] Dies setzt voraus, dass zu einem gegebenen Satz von Trainingsdaten (Input) der gewünschte Output (z.B. Label im Sinne einer Klassifikation) bereits bekannt ist.[95] Während der Trainingsphase wird der von dem Machine Learning Modell erzielte Output mit dem gewünschten Output verglichen und das Modell entsprechend angepasst.[96] Ziel des Trainings ist die Minimierung der Fehlerfunktion.[97] Der Impuls für die Anpassung kommt hier also von außen. Die Güte des Modells, insbesondere auch inwieweit das Modell gelernt hat von den Trainingsdaten zu abstrahieren, wird im Anschluss an die Trainingsphase anhand von unbekannten Testdaten überprüft.[98] Von überwachten Lernprozessen zu unterscheiden ist zunächst das unsupervised learning. Beim unsupervised learning ist keine Supervisionsinstanz beteiligt und es ist nicht erforderlich, dass der gewünschte Output bereits bekannt ist.[99] Das Ziel des Machine Learning Modells besteht vielmehr darin, allein anhand des gegebenen Inputs zu lernen und darin Muster zu erkennen.[100] Dieses Ziel lässt sich am besten erreichen, wenn man große Datenmengen zur Verfügung hat und daraus bestimmte Erkenntnisse ziehen und damit Wissen extrahieren kann.[101] Die dritte Form des Lernens ist das sogenannte reinforcement learning (bestärkendes Lernen). Anders als beim supervised learning erhält das Modell hier für erzeugte Outputs keine Korrektur bzw. Anleitung, wie das erwünschte Ziel erreicht werden soll, sondern lediglich eine Bewertung in dem Sinne, dass das Modell Anreize für gute Entscheidungen und „Bestrafungen" für schlechte Entscheidungen erhält

Lernen aus rechtlicher Perspektive auch noch näher im 2. Teil dieser Arbeit, insbesondere für die Zwecke der urheberrechtlichen Beurteilung generativer Künstlicher Intelligenz unter § 3 Abschnitt III.2.c.bb.(IV)(2)(c).

93 *Russell* et al., Artificial intelligence, 671; vgl. auch *Murphy* - Probabilistic machine learning : an introduction, 2022, 1ff; *Marsland,* Machine learning, 6.

94 *Russell* et al., Artificial intelligence, 671.

95 *Alpaydin,* Maschinelles Lernen, 11; *Marsland,* Machine learning, 6; *Buxmann/Schmidt,* Künstliche Intelligenz, 9f.

96 Vgl. *Marsland,* Machine learning, 1ff; *Mohri/Rostamizadeh/Talwalkar,* Foundations of machine learning, 4f.

97 *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436); *Alpaydin,* Maschinelles Lernen, 10.

98 Vgl. *Mohri/Rostamizadeh/Talwalkar,* Foundations of machine learning, 5; *Buxmann/Schmidt,* Künstliche Intelligenz, 10.

99 Vgl. *Alpaydin,* Maschinelles Lernen, 11ff; *Murphy,* Probabilistic machine learning: an introduction, 14.

100 *Alpaydin,* Maschinelles Lernen, 11f; *Buxmann/Schmidt,* Künstliche Intelligenz, 10; *Marsland,* Machine learning, 6; *Murphy,* Probabilistic machine learning : an introduction, 14.

101 Wenn mithilfe von statistischen Methoden oder im Wege des Maschinellen Lernens Wissen aus einer Vielzahl von Daten gewonnen wird, spricht man auch von Data Mining. Die zu lösenden Fragestellungen und die einsetzbaren Mittel beim Maschinellen Lernen und beim Data Mining entsprechen sich im Wesentlichen, weshalb die Ausführungen zum Maschinellen Lernen im Wesentlichen auch für den Bereich des Data Mining gelten. *Ertel,* Grundkurs Künstliche Intelligenz, 196.

und sich selbstständig korrigieren muss.[102] Ziel des Modells ist die Maximierung der Belohnung auf lange Sicht.[103] Derartige Lernmethoden werden regelmäßig dann eingesetzt, wenn es an hinreichenden Trainingsdaten fehlt.[104] Letztlich gehen die Modelle, die sich Verfahren des reinforcement learnings zunutze machen, nach dem Trial-and-Error-Prinzip vor, um ihr langfristiges Ziel zu erreichen.[105]

Mithilfe des Machine Learnings lassen sich eine Vielzahl verschiedener Probleme lösen; Machine Learning Modelle bieten daher zahlreiche Anwendungsmöglichkeiten.[106] Die Aufgaben der Modelle reichen von der Assoziation, der Klassifikation und der Regression über die Kompression bis hin zur Generierung neuer Inhalte. Je nach Zielsetzung unterscheiden sich auch die eingesetzten Lernmethoden, die jeweils unterschiedliche Stärken und Schwächen aufweisen: Bei Regressions- und Klassifikationsproblemen kommen etwa vielfach Verfahren des supervised learnings zum Einsatz;[107] besteht die Aufgabenstellung in einer Dimensionalitätsreduktion (Kompression)[108], Merkmalsextraktion[109] oder Clusteranalyse[110] werden hingegen regelmäßig Verfahren des unsupervised learnings eingesetzt. Zur Lösung komplexer Aufgaben werden für gewöhnlich verschiedene Lernverfahren kombiniert.[111] Neben den Künstlichen Neuronalen Netzen (KNN) als besonders wichtiger Klasse von Modellen des Maschinellen Lernens, die auch den zur Erstellung von Deep Fakes genutzten Modellen zugrunde liegen, können etwa auch Decision Trees, Support Vector Machines, Bayesian Networks sowie eine Vielzahl weiterer Modellarchitekturen eingesetzt werden.[112] Dabei macht sich das Machine Learning auch Methoden der Statistik[113] zunutze: Indem Erfahrungsdaten aus der Vergangenheit ausgewertet und Muster darin erkannt werden, können die Machine Learning Modelle Ableitungen für die Zukunft bilden.[114]

---

102 *Alpaydin,* Maschinelles Lernen, 582; *Murphy,* Probabilistic machine learning: an introduction, 18; *Marsland,* Machine learning, 6.

103 *Patel,* Praxisbuch unsupervised learning, 26; *Mohri/Rostamizadeh/Talwalkar,* Foundations of machine learning, 8; *Buxmann/Schmidt,* Künstliche Intelligenz, 11; *Alpaydin,* Maschinelles Lernen, 582.

104 *Russell* et al., Artificial intelligence, 840; *Ertel,* Grundkurs Künstliche Intelligenz, 313.

105 *Ertel,* Grundkurs Künstliche Intelligenz, 313.

106 Vgl. etwa *Alpaydin,* Maschinelles Lernen, 4ff.

107 *Patel,* Praxisbuch unsupervised learning, 11; *Alpaydin,* Maschinelles Lernen, 10; vgl. auch *Murphy,* Probabilistic machine learning : an introduction, 1ff; *Marsland,* Machine learning, 6ff.

108 Vgl. etwa *Murphy,* Probabilistic machine learning : an introduction, 15f; *Buxmann/Schmidt,* Künstliche Intelligenz, 10.

109 *Alpaydin,* Maschinelles Lernen, 126, 131.

110 Vgl. etwa *Murphy,* Probabilistic machine learning : an introduction, 14f; *Ertel,* Grundkurs Künstliche Intelligenz, 245; *Alpaydin,* Maschinelles Lernen, 175.

111 Vgl. etwa im Zusammenhang mit dem Deep Learning: *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436 ff.); *Ertel,* Grundkurs Künstliche Intelligenz, 299ff.

112 Vgl. *Ertel,* Grundkurs Künstliche Intelligenz, 191ff.

113 Siehe auch *Alpaydin,* Maschinelles Lernen, 4.

114 Zum Verhältnis von Machine Learning und Statistik siehe etwa *Bzdok/Altman/Krzywinski*, Nat Methods 15 (2018), 233 (233 f.).

5. Künstliche Neuronale Netze

Die Anfänge der Disziplin der Künstlichen Neuronalen Netze reichen gar noch weiter zurück als die Forschung zu Künstlicher Intelligenz als solcher: genauer auf die Entwicklung des *McCulloch-Pitts-Neurons*[115] im Jahr 1943.[116] Bei der Entwicklung der Künstlichen Neuronalen Netze (KNN) hat man sich wiederum den Menschen zum Vorbild genommen und versucht die Prozesse der Informationsverarbeitung, welche in den natürlichen neuronalen Netzen des menschlichen Gehirns ablaufen, mithilfe eines mathematischen Modells nachzubilden.[117] Ein tieferes Verständnis für die Funktionsweise der Modellarchitektur der Künstlichen Neuronalen Netze erfordert also zunächst einen kurzen Blick in die Kognitions- bzw. Neurowissenschaften, um ein grundlegendes Verständnis für die Strukturen zu erhalten, die der Informationsverarbeitung im menschlichen Organismus zugrunde liegen.

a. Das menschliche Vorbild: Informationsverarbeitung über Neuronen und Synapsen

Die Reiz- und Informationsverarbeitung und in der Folge auch Lernvorgänge laufen im menschlichen Gehirn über Neuronen und Synapsen ab, welche in ihrem Zusammenwirken ein hochkomplexes neuronales Netz ergeben.[118] Ein einzelnes Neuron besteht aus einem Zellkörper, an welchen sich die Dendriten mit den Synapsen anschließen. Der Zellkörper dient als Speicher für die an einem Neuron ankommenden elektrischen Spannungen, er fungiert also als eine Art Kondensator.[119] Über das vom Zellkörper abgehende Axon gelangt man zu den Synapsen, den eigentlichen Informationsübertragungsorganen. Über diese Synapsen, die Verbindungsstellen, können Spannungen übertragen werden. Die ankommenden Spannungsimpulse werden an der Synapse gesammelt, bis genügend Spannung eingegangen ist, sodass ein bestimmter Schwellenwert überschritten wird. Wird diese Schwelle erreicht, so wird die Spannung über das Axon und die Verknüpfungsstellen der Synapsen weitergeleitet.[120] Für das spätere Lernen in einem Neuronalen Netz sind insbesondere diese Synapsen von Bedeutung, denn diese können ihre Leitfähigkeit verändern und in der Folge mehr bzw. weniger Spannungsimpulse übertragen.[121]

115 *McCulloch/Pitts*, Bulletin of Mathematical Biophysics 5 (1943), 115 (115 ff.).
116 *Russell* et al., Artificial intelligence, 35; *Ertel,* Grundkurs Künstliche Intelligenz, 9.
117 *Alpaydin,* Maschinelles Lernen, 285f; *Ertel,* Grundkurs Künstliche Intelligenz, 265f.
118 Siehe weiterführend zur Informationsverarbeitung im menschlichen Gehirn: *Anderson* - Kognitive Psychologie, 7. Aufl. 2013, 10ff.
119 *Ertel,* Grundkurs Künstliche Intelligenz, 266.
120 *Ertel,* Grundkurs Künstliche Intelligenz, 266; ausführlich *Wendt* - Allgemeine Psychologie - Wahrnehmung, 1. Aufl. 2014, 59ff.
121 *Ertel,* Grundkurs Künstliche Intelligenz, 268.

b. Mathematische Modellierung der Informationsverarbeitung in Form eines Künstlichen Neuronalen Netzes

Diese Abläufe hat man sich bei der Modellierung der Künstlichen Neuronalen Netze zunutze gemacht und softwaretechnisch im Wege einer Folge von mathematischen Rechenoperationen implementiert.

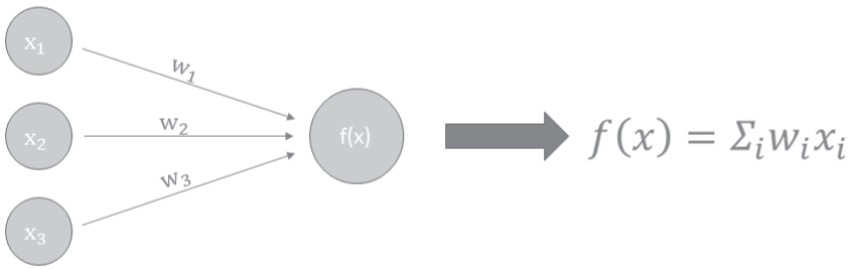

Abbildung 1: Künstliches Neuron mit den Eingabewerten x_1, x_2 und x_3 und der Ausgabe $f(x)$ sowie den synaptischen Gewichten w_1, w_2 und w_3

Das Perzeptron[122] als die grundlegende Verarbeitungseinheit eines Künstlichen Neuronalen Netzwerks besteht im Wesentlichen aus zwei Einheiten: der Eingabe- und der Ausgabeschicht.[123] Unter Bezugnahme des menschlichen Vorbildes treten bei der Modellierung des Künstlichen Neurons an die Stelle der Dendriten die Inputwerte x_i und die Outputwerte $f(x)$ an die Stelle des Axons.[124] Die anpassungsfähigen Synapsen werden über Gewichte in Form von Multiplikationsfaktoren w_i repräsentiert. Der Output wird schließlich im einfachsten Fall, in dem die Aktivierungsfunktion des Output-Neurons in der Identität besteht (in diesem Fall gilt also $f(x) = x$), durch die Summe der gewichteten Inputwerte repräsentiert.[125]

$$f(x) = \sum_i w_i x_i$$

Um dieses einfache Modell eines Neurons als binären Klassifizierer[126] einsetzen zu können, muss es sich bei der Aktivierungsfunktion des Output-Neurons um eine Schwellwertfunktion (Heavisidesche Stufenfunktion) handeln, sodass man als

122 Dieses grundlegende Modell geht auf die Arbeit von Rosenblatt zur Mustererkennung zurück: *Rosenblatt*, Psychological Review 65 (1958), 386 (386 ff.).
123 *Alpaydin*, Maschinelles Lernen, 289f.
124 *Mainzer*, Künstliche Intelligenz – Wann übernehmen die Maschinen?, 104.
125 *Alpaydin*, Maschinelles Lernen, 289f; *Ertel*, Grundkurs Künstliche Intelligenz, 269.
126 Mithilfe eines solchen binären Klassifizierers lassen sich im Grunde beliebig komplexe Klassifikationsprobleme lösen. So etwa das bekannte Beispiel, in dem Fotos danach unterschieden werden sollen, ob sie Hunde oder Gebäck abbilden. Siehe das Beispiel bei *Zack*, chihuahua or muffin, 11.02.2021, https://twitter.com/teenybiscuit/status/707727863571582978 [zuletzt geprüft am 11.02.2021].

Output jeweils die Ausgabe einer der beiden möglichen Klassen erhält: Die Ausgabe nimmt dann im Falle der binären Codierung entweder den Wert 0 oder den Wert 1 an.[127]

c. Lernen in einem Künstlichen Neuronalen Netz

Um eine vorgegebene Aufgabenstellung lösen zu können, muss ein Modell lernen. Durch die Modellierung des Künstlichen Neuronalen Netzes anhand des menschlichen Vorbildes erfolgt auch das Lernen in diesem Künstlichen Neuronalen Netz in Anlehnung an das menschliche Lernen. In denjenigen Bereichen, in denen es nicht genügt eine Vielzahl von Fällen einfach auswendig zu lernen[128], lernt der Mensch, indem er im Anschluss an das Trainieren einer Vielzahl von Beispielen eine Generalisierung vornimmt und damit eine abstrakte Lernregel findet.[129] Anpassungsfähig und damit lernfähig sind beim menschlichen Vorbild nur die Synapsen, welche ihre Leitfähigkeit verändern können.[130] Je öfter und intensiver eine Synapse genutzt wird, desto besser wird ihre Leitfähigkeit und desto mehr Gewicht erlangt sie; Umgekehrt können Synapsen verkümmern, die nicht genutzt werden.[131] Entsprechendes gilt im Grundsatz auch für die technische Implementierung: Das Lernen in einem Künstlichen Neuronalen Netz erfolgt über eine sukzessive Anpassung der Modellgewichte in Reaktion auf das Training anhand einer Vielzahl von Trainingsdaten.[132]

Um bei dem einfachen Beispiel eines linearen Klassifizierers zu bleiben: Soll ein Künstliches Neuronales Netz in der Lage sein, Äpfel von Birnen zu unterscheiden, so muss es anhand einer Vielzahl von Apfel- und Birnen-Daten trainiert werden. Um das binäre Klassifikationsproblem zu lösen, ist eine Funktion zu finden, die die Klassifikation der linear separablen Klassen der Äpfel und Birnen korrekt abbildet. Dazu ist in einem n-dimensionalen Merkmalsraum eine $n-1$-dimensionale Hyperebene zur Trennung der Klassen gesucht.[133]

127 *Ertel,* Grundkurs Künstliche Intelligenz, 269; *Alpaydin,* Maschinelles Lernen, 290f; *Russell* et al., Artificial intelligence, 700f.

128 Das Auswendiglernen von Daten ist technisch über das bloße Speichern von Informationen abbildbar. Allerdings bezwecken Modelle der Künstlichen Intelligenz und des Maschinellen Lernens gerade nicht die reine Wiedergabe bekannter Daten, sondern eine Abstraktion über die konkreten Trainingsdaten hinaus. Vgl. dazu auch *Ertel,* Grundkurs Künstliche Intelligenz, 192.

129 *ders.,* Grundkurs Künstliche Intelligenz, 192.

130 *Mainzer,* Künstliche Intelligenz – Wann übernehmen die Maschinen?, 105.

131 *Ertel,* Grundkurs Künstliche Intelligenz, 268; ausführlich *Mallot/Hübner* - § 11: Neuronale Netze, in: *Görz/Schneeberger/Schmid,* Handbuch der Künstlichen Intelligenz, 2013 S. 357 ff., 363f.

132 *Alpaydin,* Maschinelles Lernen, 290; siehe ausführlich zu der Modellierung Künstlicher Neuronaler Netze im Einzelnen *Mallot/Hübner*, in: *Görz/Schneeberger/Schmid*, Handbuch der Künstlichen Intelligenz, 2013, 365ff.

133 Vgl. *Ertel,* Grundkurs Künstliche Intelligenz, 199f.

Bei einem „klassischen" Künstlichen Neuronalen Netz kommt in diesem Zusammenhang dem Feature Engineering eine entscheidende Bedeutung zu,[134] also insbesondere der Identifizierung relevanter Merkmale in den Daten. In dem Beispiel der Klassifikation von Äpfeln und Birnen könnten dem Modell etwa die Merkmale Farbe, Form und Oberflächenstruktur vorgegeben werden. Diese Merkmale werden als Input-Neuronen modelliert. Entsprechend den identifizierten Merkmalen erhalten die zu klassifizierenden Eingabedaten einen Merkmalsvektor. Die Modellgewichte werden initial zufällig gewählt und im Laufe des Lernprozesses sukzessive angepasst. Die Summe der gewichteten Inputwerte wird an die Outputschicht weitergeleitet und das Output-Neuron nimmt nach Anwendung der Aktivierungsfunktion φ den Wert 0 oder 1 an, nimmt mithin eine Klassifizierung als Apfel oder als Birne vor.

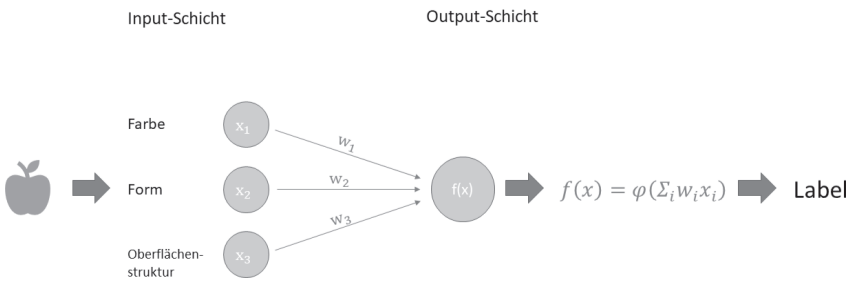

Abbildung 2: Beispiel eines einfachen künstlichen neuronalen Netzes mit drei Eingabeneuronen und einem binär codierten Ausgabeneuron zur Klassifizierung von Äpfeln und Birnen

Wird zum Training des Modells ein überwachtes Lernverfahren genutzt, erfolgt die Gewichtsanpassung, mithin der Lernvorgang, über einen Vergleich des auf Grundlage der initialen Gewichte erzielten Outputs (hier in Form des berechneten Labels: Apfel oder Birne) mit dem gewünschten Output (tatsächlich zutreffendes Label: Apfel oder Birne). Dies geschieht in einer Vielzahl von Trainingsrunden anhand sämtlicher Daten des Trainingsdatensets, bis das Netz eine hinreichende Approximation gefunden hat. Um dieses Ziel zu erreichen muss der mittlere Fehler minimiert werden.[135] Dazu wird vielfach auf den Backpropagation-Algorithmus[136] zurückgegriffen.[137] Dabei wird auf Grundlage des Vergleichs von anhand eines

134 Siehe dazu bereits zuvor Fn. 88 sowie sogleich unter 6.

135 Vgl. *Ertel,* Grundkurs Künstliche Intelligenz, 284f.

136 Grundlegend *Rumelhart/McClelland/Williams* - Learning Internal Representations by Error Propagation, in: *Rumelhart/McClelland James L./Feldman,* Parallel Distributed Processing, 1987 S. 318 ff.

137 Goodfellow et al. stellen klar, dass sich der Backpropagation-Algorithmus auf das Verfahren zur Berechnung des Gradienten beschränkt, das Lernen als solches hingegen erfolge auf Grundlage eines weiteren Algorithmus, etwa mithilfe des in der Praxis weit verbreiteten stochastic gradient descent-Verfahrens. Siehe *Goodfellow/Bengio/Courville* - Deep Learning, 1. Aufl. 2018, 225.

Trainingsbeispiels erzieltem Output mit dem gewünschten Output die Gewichtsanpassung mithilfe der generalisierten Delta-Regel berechnet.[138] Im Anschluss an den Trainingsvorgang wird das Künstliche Neuronale Netz anhand der unabhängigen Testdaten daraufhin überprüft, ob eine hinreichende Generalisierung stattgefunden hat, um einem reinen Auswendiglernen der Trainingsdaten vorzubeugen.

6. Deep Learning

Die bislang erörterten Algorithmen und Lernmechanismen sind zwar grundsätzlich in der Lage eine Vielzahl von Problemen – auch hochkomplexer Art – zu lösen, allerdings leiden sie an gleich mehreren Schwächen. Zum einen erfordern die klassischen Modelle des Maschinellen Lernens ein gewisses Verständnis für die Strukturen, welche in den Eingabedaten enthalten sind; gerade bei hochkomplexen Problemen wie der Bilderkennung ist dies jedoch mit einem immensen Aufwand seitens derjenigen Personen verbunden, die die Algorithmen designen.[139] Zudem steigt mit der Komplexität der zu lösenden Problematik der Trainingsaufwand des Modells, es ist also erforderlich, die Daten zu komprimieren.[140] Um diese Kompression zu erzielen und die relevanten Merkmale zur Lösung komplexer Aufgabenstellungen zu extrahieren, macht man sich das *Deep Learning (tiefes Lernen)* zunutze. Der Vorteil des Deep Learning gegenüber dem gewöhnlichen Maschinellen Lernen besteht gerade darin, dass die Merkmale nicht von derjenigen Person vorgegeben werden müssen, die das Modell designt, mithin kein aufwändiges Feature Engineering, das ein tieferes Verständnis für die in den Daten enthaltenen Strukturen voraussetzt, erforderlich ist.

Ähnlich dem menschlichen Vorbild wird beim Deep Learning in einer Vielzahl von Neuronen-Schichten ein immer tiefer gehendes Verständnis für eine bestimmte Problematik erlernt.[141] Dabei werden die zur Erfassung der Problematik zu lernenden Merkmale in den verschiedenen Schichten zunehmend abstrakter.[142] Dies geschieht, indem über eine Komprimierung der Eingangsdaten zunächst eine Merkmalsextraktion stattfindet und im Anschluss durch einen Vergleich mit dem erwünschten Output, die optimalen Gewichtsanpassungen berechnet werden.[143] Im ersten Schritt der Merkmalsextraktion werden in den einzelnen Schichten die verschiedenen Merkmale gelernt. Dabei steigt das Abstraktionslevel, je tiefer die Schichten in das neuronale Netz führen. So werden im Zusammenhang mit der Bilderkennung beispielsweise

138 Vgl. *Rumelhart/McClelland/Williams*, in: *Rumelhart/McClelland James L./Feldman*, Parallel Distributed Processing, 1987, 322ff; vgl. auch *Ertel,* Grundkurs Künstliche Intelligenz, 284ff.
139 *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436).
140 *Ertel,* Grundkurs Künstliche Intelligenz, 300.
141 Siehe auch *ders.,* Grundkurs Künstliche Intelligenz, 301.
142 *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436); *Alpaydin,* Maschinelles Lernen, 328.
143 Siehe *Alpaydin,* Maschinelles Lernen, 327ff.

in der Merkmalsschicht, die der Eingabeschicht am nächsten ist, typischerweise Kanten und Linien erkannt, in einer weiteren Schicht werden diese dann bereits zu bestimmten Mustern zusammengesetzt, erst in einer späteren Merkmalsschicht werden diese Muster dann verbunden und es können Objekte erkannt werden.[144] Diese Merkmalsextraktion ist mit relativ geringem menschlichen Aufwand möglich, indem der Algorithmus die einzelnen Merkmalsschichten mithilfe eines Universallernverfahrens selbstständig aus den vorhandenen Daten lernt.[145] Zur Datenkompression und Merkmalsextraktion kann etwa das Autoencoderverfahren eingesetzt werden.[146] Nach Abschluss der Merkmalsextraktion schließt sich ein klassisches überwachtes Netz an.[147] Als Eingabe dient hier die komprimierte Repräsentation der Ausgangsdaten, die durch die Merkmalsextraktion gewonnen wurde. Dieses überwachte Netzwerk wird nun trainiert, indem ein Vergleich der gegebenen Daten mit der gewünschten Ausgabe stattfindet. Im Falle der Gesichtserkennung mithilfe eines Deep Learning Algorithmus etwa wird der Algorithmus in diesem überwachten Netz dahingehend trainiert, dass er das passende Label für die Eingabedaten ausgibt, das Gesicht also richtig klassifiziert. Das geschieht wiederum, indem das vom Algorithmus berechnete Label mit dem vorgegebenen angestrebten Label verglichen wird und die Gewichte angepasst werden, indem von diesem gewünschten Output aus zurückberechnet wird, wie die Gewichte verändert werden müssten, um möglichst nahe an das angestrebte Ziel zu gelangen.[148] Im Anschluss an den Lernvorgang wird das Deep Learning Netzwerk anhand unabhängiger Test-Daten daraufhin überprüft, ob es eine hinreichende Generalisierung vorgenommen hat und so in der Lage ist eine Vielzahl unbekannter Fälle richtig zu lösen.

144 *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436); *Ertel,* Grundkurs Künstliche Intelligenz, 301.

145 *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436 f.). In der Praxis wird dazu vielfach auf das Verfahren des stochastic gradient descent zurückgegriffen. Siehe a.a.O. S. 437.

146 Vgl. *Ertel,* Grundkurs Künstliche Intelligenz, 301ff; Zur Anwendung des Autoencoder-Verfahrens zur Erstellung von Deep Fakes siehe sogleich unter II.

147 *ders.,* Grundkurs Künstliche Intelligenz, 303.

148 Siehe weiterführend dazu: *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436 ff.).

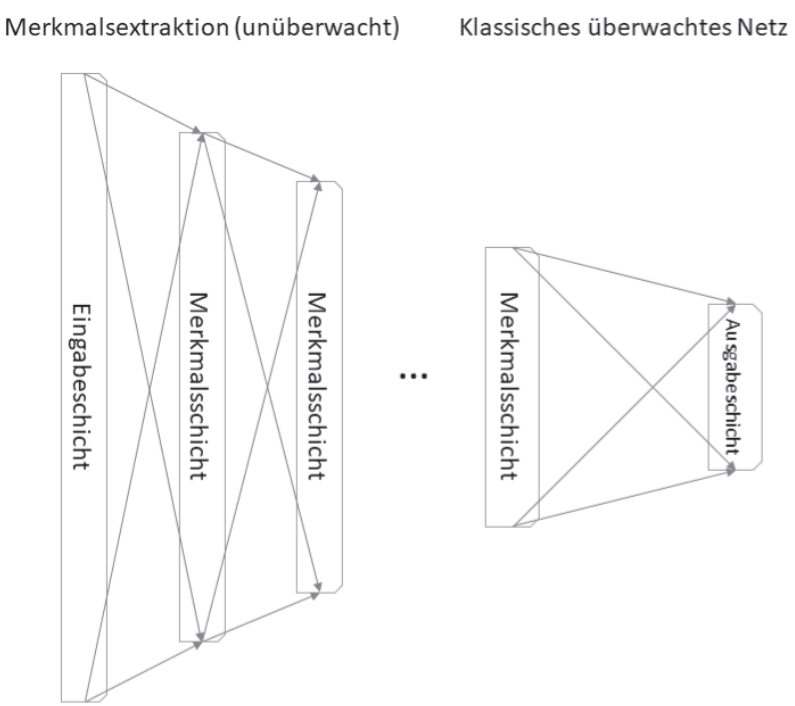

Abbildung 3: Deep Learning Netzwerk bestehend aus einem unüberwachten Netz zur Merkmalsextraktion sowie einem daran anschließenden überwachten Netz

Deep Learning Verfahren können in einer Vielzahl von Bereichen zur Anwendung gelangen und sind den gewöhnlichen Methoden dabei aufgrund der immanenten Vorteile deutlich überlegen.[149]

7. Convolutional Neural Nets (CNNs)

Im Zusammenhang mit der Verarbeitung von Bildern ist zudem auf eine spezielle Netzwerkarchitektur eines Deep Feedforward Nets hinzuweisen: Convolutional

149 *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (436) nennen hier etwa die Bereiche der Bild- und Spracherkennung; Allerdings weist das Deep Learning auch Nachteile auf. Siehe dazu etwa *Alpaydin,* Maschinelles Lernen, 328; sowie *Ertel,* Grundkurs Künstliche Intelligenz, 304.

Neural Networks (CNNs)[150].[151] Der besondere Erfolg von CNNs im visuellen Bereich liegt unter anderem in der Schwäche der gewöhnlichen neuronalen Netze im Umgang mit Bildern begründet, da bei der Verarbeitung in einem gewöhnlichen Modell sämtliche Informationen über räumliche Abhängigkeiten einzelner Pixel verloren gehen,[152] sodass ein solches Modell nicht in der Lage ist Objekte in Bildern positionsunabhängig zu erkennen.[153]

Digitale Bilder bestehen je nach ihrer Auflösung aus einer Vielzahl von Pixeln. Diese Pixel werden durch Pixelwerte beschrieben. Wenn man zunächst vom einfachsten Fall eines Binärbildes ausgeht, welches aus 64x64 Pixeln besteht, dann würde eine Anzahl von 64x64 = 4.096 Eingabeneuronen benötigt, um dieses Bild in das Neuronale Netz einspeisen zu können. Jedes dieser Neuronen stünde stellvertretend für ein Pixel und würde im Falle des obigen Beispiels eines Schwarz-Weiß-Bildes abhängig vom Informationsgehalt an der jeweiligen Stelle im Bild einen Wert zwischen 0 und 1 annehmen. Diese Methode funktioniert zwar grundsätzlich, in der Praxis hat sich im Bereich der Bildverarbeitung jedoch die Netzwerkarchitektur der Convolutional Neural Networks durchgesetzt.

Anders als gewöhnliche neuronale Netze sind Convolutional Neural Nets in der Lage die Input-Daten in Form einer Matrix zu verarbeiten.[154] Das zu verarbeitende

150 Grundlegend zu der Netzwerkarchitektur, die nunmehr als Convolutional Neural Net bekannt ist: *LeCun* - Generalization and Network Design Strategies, http://yann.lecun.com/exdb/publis/pdf/lecun-89.pdf [zuletzt geprüft am 13.06.2024]; *LeCun* et al., Neural Computation 1 (1989), 541; *LeCun* et al., Handwritten Digit Recognition with a Back-Propagation Network, in: *Touretzky,* Advances in Neural Information Processing Systems, Collected papers of the 1989 IEEE Conference on Neural Information Processing Systems - Natural and Synthetic, held November 27 - 30, 1989, in Denver, Colorado, 1989 S. 396 ff., https://proceedings.neurips.cc/paper_files/paper/1989/file/53c3bce66e43be4f209556518c2fcb54-Paper.pdf [zuletzt geprüft am 22.06.2025]; *LeCun* et al., Proceedings of the IEEE 86 (1998), 2278.

151 Siehe ausführlich zu Convolutional Neural Nets: *O'Shea/Nash* - An Introduction to Convolutional Neural Networks, 1ff., https://arxiv.org/pdf/1511.08458.pdf [zuletzt geprüft am 22.06.2025]; *Albawi/Mohammed/Al-Zawi,* Understanding of a convolutional neural network, in: *IEEE,* Proceedings of 2017 International Conference on Engineering & Technology (ICET'2017), Akdeniz University, Antalya, Turkey, 21-23 August 2017, 2017 S. 1 ff., 1ff. [zuletzt geprüft am 15.06.2025]; *Goodfellow/Bengio/Courville,* Deep Learning, 369ff; *Saha,* A Comprehensive Guide to Convolutional Neural Networks, 15.12.2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 [zuletzt geprüft am 14.12.2020]; Weiterführend zum Einsatz von Convolutional Neural Nets zur Merkmalsreduktion auch: *Ertel,* Grundkurs Künstliche Intelligenz, 303.

152 *Saha,* A Comprehensive Guide to Convolutional Neural Networks, 15.12.2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 [zuletzt geprüft am 14.12.2020].

153 *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021].

154 *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021]; Neben der Bildverarbeitung eignen sich CNNs daher besonders gut auch zur Verarbeitung von Audio und Video, vgl. *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (439); Vgl. für einen solchen Ansatz etwa *Thies* et al. - Neural Voice Puppetry: Audio-driven Facial Reenactment, https://arxiv.org/abs/1912.05566 [zuletzt geprüft am 22.06.2025], die aus Audio-Input Video-Material generieren.

Bild muss also nicht in einer langen Reihe als Vektor dargestellt werden, sondern kann als Matrix bestehend aus den Pixelwerten verarbeitet werden und behält somit die Informationen über die räumlichen Zusammenhänge im Bild.[155] Die Architektur der Convolutional Neural Nets birgt aus diesem Grund entscheidende Vorteile in der Bildverarbeitung. Denn benachbarte Datenpunkte weisen häufig hohe Korrelationen auf und lassen sich zu Motiven zusammensetzen, die auf der Grundlage von CNNs leichter und losgelöst von deren Bildposition zu erkennen sind.[156] Hinzu kommt, dass Convolutional Neural Nets leichter zu trainieren sind und eine bessere Generalisierung aufweisen als „gewöhnliche" Deep Neural Nets, bei denen angrenzende Schichten vollständig verbunden sind.[157] In der Regel handelt es sich bei den zur Erstellung von Deep Fakes eingesetzten Modellen also um solche, die sich die Netzwerkarchitektur eines Convolutional Neural Nets zunutze machen.

Inspiriert durch die Vorstellungen der visuellen Neurowissenschaften[158] ist ein Convolutional Neural Net im Grundsatz aus verschiedenen Schichten unterschiedlicher Funktionalität aufgebaut:[159] den Convolutional Layers, den Pooling Layers und den Fully-connected Layers.[160] Grundlegende Verarbeitungseinheit eines solchen CNNs sind die Convolutional Layers.[161] In diesen Convolutional Layers (aus dem Englischen: convolution = Faltung) findet die Faltung statt, mit dem Ziel die wesentlichen Merkmale des Bildes herauszufiltern.[162] Um dieses Verständnis für das Bild zu erhalten, wird eine Faltungsmatrix (Kernel-Filter) benötigt, welche über die Pixelwerte des Eingabebildes gelegt wird.[163] Die Werte der Faltungsmatrix und die des Eingabebildes werden jeweils multipliziert und im Anschluss addiert, sodann wird die Matrix verschoben und dasselbe passiert für den neuen Bereich bis das Bild

155 *O'Shea/Nash* - An Introduction to Convolutional Neural Networks, 3f., https://arxiv.org/pdf/1511.08458.pdf [zuletzt geprüft am 22.06.2025]; *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021].

156 *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (439).

157 *dies.*, Nature 521 (2015), 436 (439).

158 Zu den neurowissenschaftlichen Grundlagen siehe *Hubel/Wiesel*, The Journal of Physiology 160 (1962), 106; *Felleman/van Essen*, Cerebral Cortex 1 (1991), 1; *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (439); *Goodfellow/Bengio/Courville,* Deep Learning, 405ff.

159 *LeCun/Bengio/Hinton*, Nature 521 (2015), 436 (439).

160 Vgl. *LeCun* et al., Proceedings of the IEEE 86 (1998), 2278 (2283 ff.); *O'Shea/Nash* - An Introduction to Convolutional Neural Networks, 4, https://arxiv.org/pdf/1511.08458.pdf [zuletzt geprüft am 22.06.2025].

161 *O'Shea/Nash* - An Introduction to Convolutional Neural Networks, 5, https://arxiv.org/pdf/1511.08458.pdf [zuletzt geprüft am 22.06.2025].

162 *Saha,* A Comprehensive Guide to Convolutional Neural Networks, 15.12.2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 [zuletzt geprüft am 14.12.2020].

163 *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021].

vollständig erfasst und die darin enthaltenen Merkmale komprimiert wurden.[164] Das Ergebnis der Convolution ist eine Matrix von der Größe der Faltungsmatrix.[165]

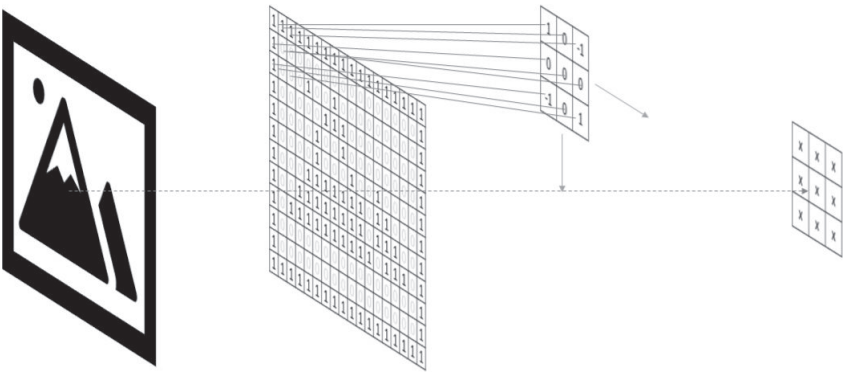

Abbildung 4: Ablauf der Convolution in einem Convolutional Neural Net

In der Netzwerkarchitektur eines CNN werden die anpassungsfähigen Gewichte durch diese Faltungsmatrix beschrieben. Für jedes Pixel des Ausgangsbildes wird ein Input-Neuron modelliert und jedes dieser Input-Neuronen wird über die Gewichte der Faltungsmatrix mit den Neuronen der nachfolgenden Schicht verbunden.[166] Wenn man die Matrix nun verschiebt, erhält man eine Vielzahl von geteilten Gewichten.[167] Im Regelfall besteht ein CNN aus einer Vielzahl solcher Convolutional Layers, in denen schrittweise ein tiefergehendes Verständnis für die Input-Daten erlangt wird, indem das Netz ortsunabhängige Strukturen erkennt – zunächst werden nur rudimentäre Elemente wie Kanten sowie Farben erkannt, in nachfolgenden Schichten werden zunehmend komplexere Informationen extrahiert.[168] Noch effizienter arbeitet ein Convolutional Neural Net, wenn man zwischen die Convolutional

164 *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021]; Siehe ausführlich und weiterführend zu den Convolutional Layers: *O'Shea/Nash* - An Introduction to Convolutional Neural Networks, 5ff., https://arxiv.org/pdf/1511.08458.pdf [zuletzt geprüft am 22.06.2025].

165 *Saha,* A Comprehensive Guide to Convolutional Neural Networks, 15.12.2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 [zuletzt geprüft am 14.12.2020]; *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021].

166 *O'Shea/Nash* - An Introduction to Convolutional Neural Networks, 1ff., https://arxiv.org/pdf/1511.08458.pdf [zuletzt geprüft am 22.06.2025].

167 *Rikiya Yamashita* et al., Insights into Imaging 2018, 611 (614).

168 *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021].

Layers sogenannte Pooling Layers schaltet. In diesen Schichten werden Informationen zusammengefasst, indem zum Beispiel beim Max-Pooling nur der höchste Wert einer Matrix übernommen wird, um so die Dimensionalität der Merkmale zu reduzieren und damit die benötigte Rechenleistung weiter zu verringern.[169] An die Convolutional- und Pooling-Layers schließt sich das Modell eines klassischen neuronalen Netzes an (Fully-connected Layers).[170] Auch das Modell eines Convolutional Neural Networks muss zunächst trainiert werden, dazu wird auch hier das Netz mit Trainingsdaten gespeist und im Anschluss werden durch Vergleich mit dem erwünschten Output die Gewichte mithilfe von Backpropagation angepasst.[171]

Die Vorteile des CNN gegenüber gewöhnlichen neuronalen Netzen liegen darin, dass die Informationen über relative Positionen (also die Beziehungen zwischen benachbarten Pixeln), welche häufig von Bedeutung sind, bei einem klassischen Netz komplett verloren gehen, wohingegen einem Convolutional Neural Network das Lernen dieser Informationen durch die matrizenbasierte Verarbeitung der Eingangsdaten bereits immanent ist.[172] In der Folge können Objekte in einem Bild positionsunabhängig erkannt werden, was einen entscheidenden Vorteil gegenüber herkömmlichen Netzen bietet.[173]

II. Deep Fake-Modelle

Im Falle von Deep Fakes wird nun eine Kombination verschiedener Deep Learning Architekturen mit der Zielsetzung der Generierung eines Produkts (Deep Fake) eingesetzt. Typischerweise besteht der Output eines solchen Deep Fake-Modells in einem Bild. Dieselbe Technologie ermöglicht jedoch neben der Ausgabe einfacher Bilder auch – durch Übertragung der Funktionsweise auf andere Medieninhalte – die Erzeugung von bewegten Bildern sowie von Tonspuren für Audioinhalte. Klassischerweise werden dabei Medieninhalte verarbeitet, die einen besonderen Persönlichkeitsbezug aufweisen (Personenbildnisse oder Stimmaufnahmen), sodass dem

169 *Saha,* A Comprehensive Guide to Convolutional Neural Networks, 15.12.2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 [zuletzt geprüft am 14.12.2020]; *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021].

170 *O'Shea/Nash* - An Introduction to Convolutional Neural Networks, 4ff., https://arxiv.org/pdf/1511.08458.pdf [zuletzt geprüft am 22.06.2025].

171 *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021].

172 *Saha,* A Comprehensive Guide to Convolutional Neural Networks, 15.12.2018, https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53 [zuletzt geprüft am 14.12.2020].

173 vgl. *Goodfellow/Bengio/Courville,* Deep Learning, 379f; *Becker,* Convolutional Neural Networks - Aufbau, Funktion und Anwendungsgebiete, 2019, https://jaai.de/convolutional-neural-networks-cnn-aufbau-funktion-und-anwendungsgebiete-1691/ [zuletzt geprüft am 18.02.2021].

Training der Modelle und deren Ausgaben eine besondere persönlichkeitsrechtliche Relevanz zukommt. Daneben ist es jedoch grundsätzlich auch möglich mithilfe derartiger Deep Fake-Modelle Medieninhalte ohne Persönlichkeitsbezug zu erzeugen (z.B. weil dort Sachen abgebildet werden oder nicht-menschliche Geräusche erzeugt werden). Gleichwohl kann auch die Nutzung von Inhalten ohne Persönlichkeitsbezug jedenfalls Fragestellungen urheberrechtlicher Art aufwerfen, die sich jedoch grundsätzlich nicht von denjenigen im Kontext von Inhalten mit Persönlichkeitsbezug unterscheiden, sodass auf diese Phänomene ohne persönlichkeitsrechtliche Relevanz im Folgenden nur am Rande eingegangen werden soll.

Da es eine Vielzahl verschiedener Technologien[174] gibt, die zur Herstellung von Deep Fakes genutzt werden (können), deren umfassende Darstellung nicht möglich und für die Zwecke dieser Arbeit auch nicht zielführend ist, soll im Folgenden eine Beschränkung auf drei grundlegende Technologien zur Erstellung von Deep Fakes vorgenommen und anhand von Beispielen erläutert werden: diese sind Autoencoder, Generative Adversarial Networks und Diffusion Models. Als Anwendungsfälle seien hier zum einen das bekannte satirische Video[175] genannt, auf dem man augenscheinlich den früheren US-amerikanischen Präsidenten Barack Obama sieht, wie er sich in abfälliger Weise über seinen Amtsnachfolger äußert. Zum anderen geht es um Technologien, mit deren Hilfe auch vollständig neue Bilder erzeugt werden können,[176] also Abbildungen von Personen, für die es kein direktes Vorbild gibt und Bilder von Personen, die es so überhaupt nicht gibt. Diese Technologien ermöglichen es, mithilfe einer hinreichenden Anzahl an Trainingsbildern – für einige Modelle genügt gar ein einziges Bild der Zielperson –[177] und einigen Stunden Trainings mit einem hinreichenden Prozessor überzeugende Deep Fakes herzustellen,[178] in denen Mimik und Gestik einer Person auf eine andere Person übertragen werden oder eine Person vollständig synthetisch generiert wird.

174 Für einen Überblick über verschiedene konkrete Deep Fake-Modelle siehe *Mirsky/Lee*, ACM Computing Surveys 54 (2022), 1.

175 *BuzzFeedVideo* - You Won't Believe What Obama Says In This Video!, https://www.youtube.com/watch?v=cQ54GDm1eL0 [zuletzt geprüft am 22.06.2025].

176 Siehe dazu etwa: *Berger,* Stockfoto-Firma veröffentlicht 100.000 KI-Gesichter, 2019, https://www.heise.de/newsticker/meldung/Stockfoto-Firma-veroeffentlicht-100-000-KI-Gesichter-4537889.html [zuletzt geprüft am 16.06.2025]; siehe etwa auch This Person Does Not Exist, https://thispersondoesnotexist.com/ [zuletzt geprüft am 14.06.2025].

177 Siehe zum Beispiel die Überlegungen von *Siarohin, Aliaksandr* et al. - First Order Motion Model for Image Animation, in: *Wallach*, et al., Advances in Neural Information Processing Systems 32 (NIPS 2019), http://papers.nips.cc/paper/8935-first-order-motion-model-for-image-animation.pdf [zuletzt geprüft am 22.06.2025].

178 Siehe beispielsweise diesen Versuch der Erstellung eines eigenen Deep Fakes durch einen "KI-Laien": *Schreiner,* KI: Deepfake selbst erstellen - so geht es, so lange dauert es, 14.02.2021, https://the-decoder.de/ki-deepfake-selbst-erstellen-so-geht-es-so-lange-dauert-es/ [zuletzt geprüft am 22.06.2025].

1. Autoencoder

Eine Möglichkeit zur Erstellung von Deep Fakes besteht darin, Mimik und Gestik bzw. gar weitergehende Bewegungen und Handlungen einer Quell-Person auf eine Ziel-Person zu übertragen. Eine solche Technologie liegt auch dem bereits angesprochenen Deep Fake-Video des ehemaligen US-amerikanischen Präsidenten Barack Obama zugrunde, in dessen Verlauf offengelegt wird, dass das Video mithilfe von Videomaterial des Schauspielers und Comedians Jordan Peele erstellt wurde.[179] Die Technologie, die hinter der Erstellung derartiger Videos steht und in der die Wurzeln des Phänomens *Deep Fake* liegen, nennt sich Autoencoder.[180] Ziel eines solchen Deep Fake-Autoencoders ist eine möglichst authentisch anmutende Übertragung von Mimik und Gestik einer Quell- auf eine Zielperson, welche es ermöglicht einer Person Aussagen und Taten in (audio-)visuell unterstützter Weise zuzuschreiben, die sie so nie gesagt oder getan hat.

a. Die Netzwerkarchitektur des Autoencoders

aa. Zielsetzung des Encoder-Decoder-Modells

Bei dem Modell des Autoencoders handelt es sich um eine spezielle Architektur Künstlicher Neuronaler Netze bestehend aus zwei Netzwerken *(Encoder und Decoder)*, deren Zielsetzung in der komprimierten Darstellung und anschließenden Rekonstruktion eines bestimmten Eingabe-Datensatzes besteht.[181] Dazu bedienen sich derartige Modelle insbesondere unüberwachter Lernverfahren.[182] Um eine Kompression zu „erzwingen", kann etwa die Anzahl der Einheiten in den verdeckten Merkmalsschichten im Vergleich zur Inputschicht verringert werden,[183] sodass eine

179 *BuzzFeedVideo,* You Won't Believe What Obama Says In This Video!

180 Die ersten Deep Fakes, die 2017 bei Reddit auftauchten, können auf die Technologie der Autoencoder zurückgeführt werden. Siehe *Nguyen* et al., Computer Vision and Image Understanding 223 (2022), 103525 (2).

181 *Gao* et al. - Extract Features Using Stacked Denoised Autoencoder, in: *Huang/Han/Gromiha,* Intelligent Computing in Bioinformatics, 2014 S. 10 ff., 10; *Goodfellow/Bengio/Courville,* Deep Learning, 563; *Ertel,* Grundkurs Künstliche Intelligenz, 302.

182 *Vincent* et al., Extracting and composing robust features with denoising autoencoders, in: *McCallum/Roweis,* ICML 2008, Proceedings of the twenty-fifth International Conference on Machine Learning, 2008 S. 1096 ff., 1096; *Vincent* et al., Journal of Machine Learning Research 11 (2010), 3371 (3372). Siehe zum unsupervised learning bereits zuvor unter I.4.

183 *Ertel,* Grundkurs Künstliche Intelligenz, 302; *Zucconi,* An Introduction to Neural Networks and Autoencoders, https://www.alanzucconi.com/2018/03/14/an-introduction-to-autoencoders/ [zuletzt geprüft am 22.06.2025]; Allerdings hat sich gezeigt, dass eine derartige Kompression über eine reduzierte Anzahl verdeckter Neuronen im Vergleich zur Eingabeschicht keine optimalen Ergebnisse liefert, sodass in der Praxis andere Verfahren, wie etwa das sog. Sparse-Coding bzw. Denoising, eingesetzt werden. Siehe Ertel a.a.O. S. 302; Für einen derartigen Ansatz des Denoising siehe etwa *Vincent* et al., in: ICML 2008 S. 1096 ff.; sowie *Vincent* et al., Journal of Machine Learning Research 11 (2010), 3371; und *Gao* et al., in: *Huang/Han/Gromiha,* Intelligent Computing in Bioinformatics, 2014, 10ff. Im weiteren Verlauf dieser Arbeit wird einheitlich die Terminologie "Dimensionalitätsreduktion" genutzt, wenngleich der

Dimensionalitätsreduktion stattfinden muss.[184] Diese Modellierung intendiert, dass in den Merkmalsschichten die wichtigsten Merkmale extrahiert werden und so nach und nach ein Verständnis für die Daten gewonnen wird *(Encoding)*. Aufgrund der Vielzahl von verdeckten Schichten, die nötig sind, um die wichtigsten Merkmale etwa eines Bildes zu extrahieren, handelt es sich regelmäßig um Deep Learning Modelle. An die Merkmalsextraktion in den verdeckten Schichten schließt sich sodann der Decodierungsprozess *(Decoding)* an. Dessen Ziel besteht darin aus den komprimierten Daten das Original zu rekonstruieren.[185] Ziel des Trainings eines solchen Autoencoders ist die Minimierung des mittleren Rekonstruktionsfehlers *(Autoencoder-Problem)*.[186]

Abbildung 5: Vereinfachte Darstellung eines Autoencoder-Modells

In dem zuvor beschriebenen Deep Learning Netzwerk (Abbildung 3) entspricht der Encoder dem unüberwachten Netz zur Merkmalsextraktion, das schrittweise ein tiefer gehendes Verständnis für die Daten erlangt, und der Decoder dem anschließenden überwachten Netz zur Rekonstruktion des Input, in dem durch Vergleich des erzielten Outputs (Rekonstruktion) mit dem erwünschten Output (Kopie des Inputs) die optimalen Gewichte berechnet werden (vgl. Abbildung 6).

Encoding-Prozess nicht zwangsläufig mit einer Reduktion der Neuronen der verdeckten Schichten im Vergleich zur Input-Schicht einhergeht.

184 *Alpaydin,* Maschinelles Lernen, 315f; *Zucconi,* An Introduction to Neural Networks and Autoencoders, https://www.alanzucconi.com/2018/03/14/an-introduction-to-autoencoders/ [zuletzt geprüft am 22.06.2025].

185 *Ertel,* Grundkurs Künstliche Intelligenz, 303.

186 *Pierre Baldi*, Proceedings of ICML Workshop on Unsupervised and Transfer Learning 2012, 37 (39).

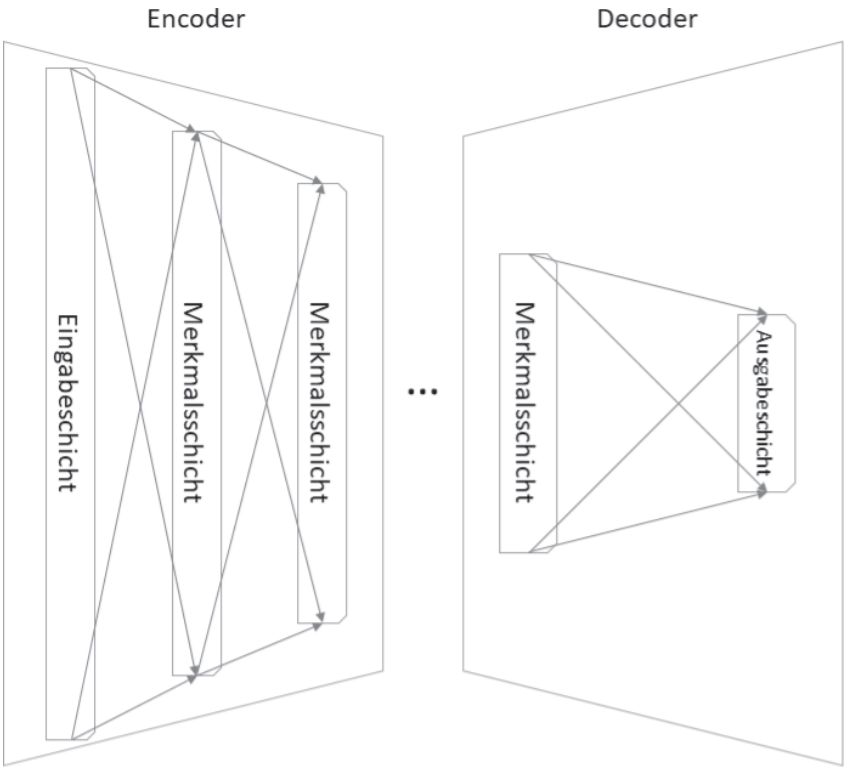

Abbildung 6: Ergänzung der Abbildung 3: Deep Learning Autoencoder Netzwerk

Obgleich die Zielsetzung von Autoencodern konzeptuell schlicht in der Abbildung einer Eingabe auf eine Ausgabe besteht, bietet dieses einfache Modell, insbesondere durch die Eigenschaft der Dimensionalitätsreduktion[187], immenses Potenzial für den Bereich des Maschinellen Lernens, so gelangen Autoencoder insbesondere auch im Bereich des Deep Learnings zur Anwendung.[188] Auf dieser Grundlage bieten Autoencoder-Modelle eine Vielzahl potenzieller Anwendungsmöglichkeiten, etwa

187 vgl. *Goodfellow/Bengio/Courville,* Deep Learning, 564.

188 *Pierre Baldi*, Proceedings of ICML Workshop on Unsupervised and Transfer Learning 2012, 37 (37). Vgl. auch Abbildung 6.

im Bereich der Klassifizierung, aber auch im generativen Bereich;[189] sie können insbesondere auch zur Generierung von Deep Fakes genutzt werden.[190]

bb. Modellierung des Autoencoders

In technischer Hinsicht handelt es sich bei dem Modell eines Autoencoders um zwei aufeinander folgende Funktionen g (Encoder) und h (Decoder). Bei einem gegebenen n-dimensionalen Eingaberaum R (Input) lernt der Encoder eine k-dimensionale komprimierte Repräsentation K des Inputs im latenten Raum darzustellen. Der Decoder kehrt diese Dimensionsreduktion um und lernt aus dem komprimierten Merkmalsraum K den Input zu rekonstruieren. Der Decoder erzeugt also eine Rekonstruktion $\ddot{R}$. Dabei gilt $k \ll n$.

Sowohl bei dem Encoder-Netz als auch bei dem Decoder-Netz handelt es sich in der Grundkonstruktion um vielschichtige Perzeptronen. Mathematisch werden diese grundsätzlich jeweils beschrieben durch eine Aktivierungsfunktion f, welche auf die Summe der gewichteten Inputwerte x_i angewendet wird und so einen Output generiert.

Es gilt also wieder:

$$f(x) = \sum_i w_i x_i$$

Aufgrund der beschriebenen Schwächen der „einfachen" Netzwerkarchitekturen im Zusammenhang mit der Verarbeitung von multidimensionalem Input wie Bild und Audio werden im Zusammenhang mit Deep Fake Autoencodern vielfach Convolutional Neural Nets eingesetzt.[191]

(I) Encoder

Im ersten Schritt findet in den verdeckten Schichten des Encoders zunächst die Kompression bzw. Merkmalsextraktion statt, wobei das Abstraktionslevel steigt, je tiefer die Schichten im Netz liegen. Zur Merkmalsextraktion in den einzelnen Merkmalsschichten kann wiederum das Autoencoder-Verfahren genutzt werden, indem in der jeweiligen Merkmalsschicht zunächst durch unüberwachtes Lernen

189 Im generativen Bereich kommen etwa Variational Autoencoder (VAE) zum Einsatz, vgl. *Kingma/Welling* - An Introduction to Variational Autoencoders, 2ff., https://arxiv.org/pdf/1906.02691 [zuletzt geprüft am 22.06.2025]; grundlegend bereits *Kingma/Welling* - Auto-Encoding Variational Bayes, https://hdl.handle.net/11245/1.434281 [zuletzt geprüft am 22.06.2025]; vgl. etwa auch *Gregor* et al., Proceedings of Machine Learning Research 2015, 1462.

190 *Nguyen* et al., Computer Vision and Image Understanding 223 (2022), 103525 (1 ff.).

191 Vgl. *Mirsky/Lee*, ACM Computing Surveys 54 (2022), 1 (6).

eine Dimensionalitätsreduktion erzielt und dann im Anschluss über den Decoder im Wege des überwachten Lernens der Input der Eingabeschicht aus deren latenter Repräsentation in Form der komprimierten Merkmale rekonstruiert wird.[192] Zur Merkmalsextraktion wird etwa das Verfahren des Denoising[193] eingesetzt, indem das Modell mithilfe von zufällig verrauschten Bildern darauf trainiert wird, das Rauschen zu entfernen und so die wesentlichen Merkmale aus den Input-Daten zu extrahieren.[194] Durch das Hinzufügen von Rauschen wird verhindert, dass das Modell (wie bei einem einfachen Autoencoder-Modell[195]) die Input-Daten einfach repliziert; stattdessen lernt das Modell eines Denoising Autoencoders zum Zwecke des Denoising die wesentlichen Merkmale zu identifizieren.[196] Um den Encoder im Hinblick auf das Denoising zu trainieren, werden wiederum der mittlere Rekonstruktionsfehler berechnet und die Gewichte entsprechend angepasst, um den Fehler zu minimieren.[197] Die Berechnung des Gradienten erfolgt mithilfe von Backpropagation. Der Rekonstruktionsfehler kann etwa über den mittleren quadratischen Fehler oder mithilfe der Kreuzentropie bestimmt werden.[198] Das Lernen als solches, also die Berechnung der optimalen Gewichtsanpassung, erfolgt sodann mithilfe des stochastischen Gradientenabstiegsverfahrens (Stochastic Gradient Descent).[199] Wie die Merkmalsextraktion als solche im Wege des unüberwachten Lernens im Encoder funktioniert, ist im Einzelnen noch weitgehend ungeklärt.[200] Es gibt jedoch mittlerweile einige Studien, welche sich mit der Erinnerung lernfähiger Modelle an Trainingsdaten beschäftigen.[201]

192 *Ertel,* Grundkurs Künstliche Intelligenz, 301f. Zur Merkmalsextraktion in den Schichten eines CNN siehe bereits zuvor unter I.7.

193 Siehe weiterführend zur Merkmalsextraktion im Wege des Denoising im Rahmen der Behandlung der Diffusionsmodelle im Abschnitt II.3.

194 *Vincent* et al., Journal of Machine Learning Research 11 (2010), 3371 (3378 f.); siehe bereits *Vincent* et al., in: ICML 2008 S. 1096 ff., 1098; *Ertel,* Grundkurs Künstliche Intelligenz, 302.

195 Siehe zu den Fällen der unter- und übervollständigen Autoencoder *Goodfellow/Bengio/Courville,* Deep Learning, 565.

196 *Vincent* et al., in: ICML 2008 S. 1096 ff., 1097f.; *Vincent* et al., Journal of Machine Learning Research 11 (2010), 3371 (3378 f.).

197 *Vincent* et al., in: ICML 2008 S. 1096 ff., 1098; *Vincent* et al., Journal of Machine Learning Research 11 (2010), 3371 (3379).

198 *Vincent* et al., Journal of Machine Learning Research 11 (2010), 3371 (3379), siehe weiterführend zu diesen Verfahren der Fehlerminimierung a.a.O. S. 3376f.

199 *dies.*, Journal of Machine Learning Research 11 (2010), 3371.

200 *Ertel,* Grundkurs Künstliche Intelligenz, 302.

201 Siehe ausführlich dazu im 2. Teil dieser Arbeit unter § 3 im Abschnitt III.3.