Richard Groß

# SITUATIONEN MASCHINELLEN LERNENS

## Über KI-Experimente in Kunst und Wissenschaft

[transcript] Digitale Gesellschaft

Richard Groß
Situationen maschinellen Lernens

**Richard Groß** (Dr. phil.) war Promotionsstipendiat am Schaufler Lab@TU Dresden und Projektkoordinator der Arnold Gehlen-Gesamtausgabe. Er studierte Soziologie, Kunstgeschichte und Musikwissenschaft an der Technischen Universität Dresden und The New School in New York. Zu seinen Forschungsschwerpunkten zählen Sozial- und Gesellschaftstheorie sowie Technik-, Medien-, Kultur- und Zeitsoziologie.

Richard Groß

# Situationen maschinellen Lernens

Über KI-Experimente in Kunst und Wissenschaft

[transcript]

Die vorliegende Publikation ist eine überarbeitete Fassung einer an der Philosophischen Fakultät der Technischen Universität Dresden unter dem Titel »Situationen maschinellen Lernens. Über rechentechnische Unbestimmtheitstransformationen und praktische Problemkalibirierungen« angenommenen Dissertation.
Tag der Disputation: 28. Mai 2025
Erstgutachter: Prof. Dr. Dominik Schrage; Zweitgutachter: Prof. Dr. Dirk Baecker

Diese Publikation entstand am Schaufler Lab@TU Dresden, einem gemeinsamen Projekt von THE SCHAUFLER FOUNDATION und der Technischen Universität Dresden, und wurde durch ein Promotionsstipendium gefördert.
Zusätzlich wurde diese Forschungsarbeit von der Graduiertenakademie der TU Dresden, finanziert aus Mitteln der Exzellenzstrategie von Bund und Ländern, durch ein Abschlussstipendium gefördert.
Die Open Access Publikation wurde durch das Sächsische Staatsministerium für Wissenschaft, Kultur und Tourismus unterstützt.
Ein Projekt von:

SCHAUFLER LAB TU DRESDEN · Technische Universität Dresden · THE SCHAUFLER FOUNDATION

**Bibliografische Information der Deutschen Nationalbibliothek**
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über https://dnb.dnb.de/ abrufbar.



transcript Verlag | Hermannstraße 26 | D-33602 Bielefeld | live@transcript-verlag.de

Umschlaggestaltung: Kordula Röckenhaus
Druck: Elanders Waiblingen GmbH, Waiblingen
https://doi.org/10.14361/9783839458778
Print-ISBN: 978-3-8376-8047-8 | PDF-ISBN: 978-3-8394-5877-8
Buchreihen-ISSN: 2702-8852 | Buchreihen-eISSN: 2702-8860

Gedruckt auf alterungsbeständigem Papier mit chlorfrei gebleichtem Zellstoff.

*Für Rama*

# Inhalt

**Metamorphosen maschinellen Lernens 2012 – 2026** ..... 9
Einleitung ..... 10
KI/ML-Diskurse ..... 20
Ansatz und Aufbau der Untersuchung ..... 28

**Aspekte einer Soziologie künstlicher Intelligenz** ..... 35
Eine ambivalente Verfassung ..... 36
Über Algorithmen ..... 42
Zwischen Korrelation und Kausalität ..... 66
Zeitliche und räumliche Dimensionen ..... 80

**Maschinelles Lernen als Kommunikation, Praxis und Situation** ..... 91
Zugänge einer Sozialtheorie maschinellen Lernens ..... 92
Künstliche Kommunikation ..... 102
Empirische Realität und soziotechnische Praxis ..... 109
Situationen maschinellen Lernens ..... 124
Empirische ML-Situationsanalyse ..... 138

**Über Bäume, Bienen und Maschinen** ..... 155
Fallstudie 1: Visionen transformierter Bäume ..... 156
Fallstudie 2: Bienen-Maschinen-Experimente ..... 182

**Experimentelle Problemkalibrierungen** ..... 211
Doppelte Unbestimmtheit und experimentelles Kontingenztraining ..... 211
Schwache Situationalität und kalibrierte Probleme ..... 221
(Un-)Bestimmte Probleme ..... 236

**Soziotechnische Intelligenz mit stochastischen Kontingenzmaschinen** ..... 251
Situation und Intelligenz ..... 252
Rechentechnische Bestimmungen sozialer Wirklichkeit ..... 268
Stochastische Kontingenzmaschinen ..... 278
Die Realität wahrscheinlicher Fiktionen ..... 286

**Literaturverzeichnis** ..... 307

**Abbildungsverzeichnis** ..... 343

**Danksagung** ..... 345

# Metamorphosen maschinellen Lernens 2012 – 2026

Innerhalb der vergangenen sechs Jahre, in denen diese Studie Gestalt angenommen hat, hätte ich für einleitende Bemerkungen zum gesellschaftlichen Stellenwert und der wissenschaftlichen Relevanz des Themas meiner Untersuchung zu verschiedenen Zeitpunkten sehr unterschiedlich ansetzen können. Dies gilt nicht nur in dem offensichtlichen Sinne, dass die letzten sechs Jahre den Zeitraum markieren, in dem die hier präsentierten Resultate meiner Forschung erst sukzessive Form angenommen haben. Im gleichen Zeitraum hat deren Gegenstand eine solche Vielzahl an möglichen Aufhängern für eine anekdotisch vermittelte Einführung in diese soziologische Studie geliefert, dass mir irgendwann gerade die Erwähnung dieses Umstand den besten Ansatz für einen Einstieg darzustellen schien. Maschinelles Lernen (im Folgenden: ML) als seit geraumer Zeit maßgebliche Variante sogenannter Künstlicher Intelligenz (KI) hat sich als soziales Phänomen in den vergangenen Jahren nicht nur beträchtlich verändert, sondern ist dabei auch kontinuierlich Gegenstand von inner- wie außerwissenschaftlichen Kontroversen gewesen. Diese betreffen Diskussionen um den aktuellen ›Stand der Technik‹ und vor allem Beurteilungen hinsichtlich der zu erwartenden – zu erhoffenden wie zu befürchtenden – Folgen ihrer Verbreitung und Weiterentwicklung.

Die Eigenheiten des Gegenstandes dieser Studie nachvollziehbar zu machen und ebenso dessen Dynamiken als soziales Phänomen Rechnung zu tragen, ist Anliegen dieser Einleitung. Damit sollen gleich zu Beginn die Spezifika des Themas vermittelt und insbesondere für die damit einhergehenden Problematisierungsweisen der Untersuchung sensibilisiert werden. Im Folgenden werde ich daher zunächst eine Skizze der jüngeren Entwicklungen um ML anfertigen und Tendenzen des diese Entwicklungen begleitenden sozial- und geisteswissenschaftlichen Diskurses nachzeichnen. Dies wird mir dazu dienen, den Gegenstand dieser Studie sowie meine Perspektive auf diesen zu spezifizieren, bevor ich schließlich den grundlegenden Ansatz und das Vorgehen meiner Untersuchung vorstelle.

## Einleitung

›Vorläufig rückblickend‹ aus dem Jahre 2026 lässt sich eine bereits mehr als ein Jahrzehnt andauernde Phase ausmachen, in der KI beständig ökonomische, wissenschaftliche und zunehmend politische Aufmerksamkeit erfahren hat. Bereits Anfang 2020 – als die Arbeit an meinem Dissertationsvorhaben am Schaufler Lab@TU Dresden begann – waren intensive öffentliche Debatten um KI als vermeintlich die Geister scheidende »Schlüsseltechnologie des 21. Jahrhunderts« (Rothe 2020) anhaltend an der (post-)massenmedialen[1] Tagesordnung.[2] Die diskursive Prominenz von KI war (und ist) derart hoch, dass sich bisweilen der Eindruck aufdrängte, soziologische KI-Forscher:innen könnten allein mit der Aufgabe, den inmitten des Hypes sich verbreitenden Halb- bis Unwahrheiten um utopische wie dystopische KI-Szenarien mit wissenschaftlicher Auf- bzw. Abklärung beizukommen, mehr als genug Beschäftigung finden. Dies lag auch daran, dass sozial- und geisteswissenschaftliche Forschung zu ML mittlerweile eigene, den Prätentionen der an der (Weiter-)Entwicklung von ML maßgeblich beteiligten Unternehmen gegenüber durchaus kritische Perspektiven etabliert hatte. Zunehmend wurde grundlegende Kritik an ML in seiner gegenwärtigen Form vernehmbar: »If 2012 was the year in which the Deep Learning revolution began, 2019 was the year in which its sources were discovered to be vulnerable and corrupted« (Pasquinelli & Joler 2021, S. 1267). ML ist als soziales Phänomen beschreib- und problematisierbar geworden. Einsichten in seine gesellschaftliche Einbettung offenbaren etwa Abhängigkeitsverhältnisse, die Beobachter:innen gleichermaßen für wissenschaftliche Forschung interessant wie politisch problematisch erscheinen.[3]

---

1 Die vorangestellte Vorsilbe markiert die Ergänzung der ›klassischen‹ Massenmedien um internetbasierte und insbes. sog. Soziale Medien.

2 Alternative Formeln zur Unterstreichung (bzw. Bewerbung) des epochalen Charakters von KI vergleichen diese (oder semantisch nah: [*Big*] *Data*) mit mutmaßlich vitalen Energie- oder Wertressourcen, etwa Elektrizität, Gold oder Öl, vgl. Sudmann 2018b, S. 57, Höinghaus 2015. Wenig überraschend ziehen solche Vergleiche nicht selten Kritik und Widerspruch nach sich, vgl. etwa Spitz 2017.

3 »Beobachter:innen« verwende ich hier identisch mit »Beobachtern« im systemtheoretischen Sinne (mit bestimmten Unterscheidungen bezeichnende und unterscheidende Beobachtungsinstanzen wie soziale und psychische Systeme, vgl. Luhmann 1984, S. 63 sowie Luhmann 1997, S. 67f.). Die von mir verwendete geschlechtsneutrale Form könnte daher gewissermaßen als irreführend verstanden werden, erschien mir allerdings im Sinne konsistenter sprachlicher Konventionen aus pragmatischen Gründen als beste Lösung, zumal geschlechtsbezogene Unterscheidungen sich bekanntermaßen als zentral für die die Selbstbeobachtung psychischer Systeme erweisen. Zudem erlaubt mir diese Wahl, im Folgenden ›binnendifferenzierend‹ »Beobachter« von »Beobachterinnen« und wiederum »Beobachter:innen« zu unterscheiden.

»The Great AI Awakening« (Lewis-Kraus 2016) schien sich daher bereits am Anfang meiner Arbeit an der vorliegenden Studie in mancherlei Hinsicht längst vollzogen zu haben. Unter diesem Titel hatte schon Ende 2016 das New York Times Magazine eine ausführliche Reportage des Journalisten Gideon Lewis-Kraus über den *state of the art* in Sachen KI publiziert. Den Anlass dazu schuf der damals als bahnbrechend wahrgenommenen Umstellung von Such- und Übersetzungsfunktionen bei Google auf Machine Learning-Verfahren. Bei diesem Vorgang, der viele Millionen Nutzer:innen vermutlich erstmalig mit auf gegenwärtigen Formen sogenannter Künstlicher Intelligenz beruhender Software in Kontakt kommen ließ, handelt es sich um ein Kernereignis in der Geschichte des neuerlichen Aufstiegs von »konnexionistischen« bzw. »subsymbolischen« Ansätzen der KI-Forschung und -Entwicklung, die mit dem Label ML verbunden werden (vgl. Sudmann 2018b).

Diese ›KI-Schule‹ verfolgt einen Ansatz, der anders als bei früheren, rückblickend häufig »good old-fashioned AI« (*GOFAI*) genannten Strömungen nicht (mehr) auf regelbasierte Programmierung für ›künstliche‹ Agenten setzt (vgl. etwa Pasquinelli 2017). Deren Leistungsfähigkeit – etwa sogenannter Expertensysteme der 1980er und 1990er Jahre – misst sich am Komplexitätsniveau dieses Skripts von symbolisch vermittelten Anweisungen. Die Programmierung derartiger Verfahren besteht dabei – schematisch gesprochen – aus diskreten Regeln zur Transformation definierter Inputs (vgl. etwa Muhle 2018, S. 18). Demgegenüber beruht ML – in erster Annäherung – darauf, dass ein in einem Datensatz als repräsentiert angenommener Phänomenzusammenhang algorithmisch modelliert wird. Dem »ground-truthing« (vgl. Jaton 2021, Kap. 1 & 2, Jaton 2024), d.h. der Festlegung einer Datenbasis, die den interessierenden Zusammenhang abbilden soll, folgt dann eine ›Trainingsphase‹, in der ein Algorithmus eine Funktion errechnet, die in bestmöglicher Annäherung die Struktur des Datensatzes, d.h. der Verteilung einzelner Datenpunkte in einem Vektorraum von Datenbeziehungen beschreiben kann. Hier sind Input und Output in Form eines Datensatzes gegeben und eine ›Maschine lernt‹ (daher *ML*) Transformationsregeln zwischen beiden. Dieser Ansatz steckt grundsätzlich hinter allen populären als Durchbrüche gefeierten Neuerungen im Feld der KI-Entwicklung innerhalb der letzten 15 Jahre – von *Computer Vision* bis zu *Large Language Models* – und verhalf die Rede von Algorithmen und KI weit über Fachkreise heraus zu popularisieren.[4]

4 Interessanterweise werden in aller Regel nicht etwa maßgebliche Fortschritte im Bereich mathematischer oder informatischer Forschung für die rapiden Entwicklungen im Bereich ML seit 2012 verantwortlich gemacht. Wesentliche theoretische Grundlagen für die Entwicklung künstlicher neuronaler Netze der Art, wie sie in populären Verfahren zum Einsatz kommen, waren bereits in den 1980er Jahren geschaffen. Effiziente und davor zunächst überhaupt viable technische Implementierungen derselben wurden jedoch erst möglich unter der Voraussetzung von hinreichend verfügbarer Rechenleistung. Der technische ›Kniff‹, die Berechnungen der Netze von der linear rechnenden CPU (*central processing unit*) auf die parallel ope-

Weil die vorliegende Arbeit eine soziologische Auseinandersetzung mit gegenwärtigen Formen von KI zu leisten beansprucht, handelt sie von ML, was die Wahl des Titels dieser Studie erklärt. Die bis zu dieser Stelle synonyme Verwendung von *KI* und *ML* soll indes die semantische Unschärfe im diskursiven Gebrauch beider Bezeichnungen spiegeln. Ohne Rekurs auf KI lässt sich kaum einführend über ML schreiben und eine gewisse begriffliche Konfusion kann wohl als symptomatisch für die diskursiven Dynamiken um KI und ML gelten. Ich erwähne diesen Umstand auch deshalb, weil gegenwärtig schon in Fragen der Begriffswahl in einer maßgeblich von der Kommerzialisierung der in der Forschung entwickelten Produkte geprägten Lage soziologisch mitreflektiert werden muss, dass sich schon auf der Ebene der Gegenstandsbezeichnung Interessen manifestieren. Unter diesen Bedingungen ist es dann wenig überraschend, dass Forschende und anders praktisch mit KI Agierende – dies auch im Vorgriff auf die Ergebnisse meiner empirischen Fallstudie – den Begriff sogar explizit ablehnen, und eher noch mit ML in Zusammenhang gebracht werden wollen, andererseits in öffentlichen Auftritten oder Anträgen selbst wiederum darauf zurückgreifen. Im Folgenden wird zumeist von *ML*, gelegentlich von *KI*, mitunter aber auch von *KI/ML* die Rede sein. Bezieht sich erstere Bezeichnung explizit auf den Gegenstand meiner Untersuchung, verstehe ich *KI* im Kontext dieser Untersuchung als eine seit nunmehr etwa 70 Jahren etablierte Semantik zur Bezeichnung einer besonderen Art technischer Phänomene, denen ML zugerechnet wird und deren Eigenart im Folgenden erörtert werden soll. *KI/ML* hingegen verwende ich für Kontexte, deren Bezugsrahmen im angedeuteten Sinne uneindeutig ist oder aber explizit KI wie auch ML einschließt.

Für das »Great AI Awakening« als Aufstieg des ML-Paradigmas in KI-Forschung und -Entwicklung der jüngeren Zeit lässt sich pragmatisch ein Anfangspunkt im Jahre 2012 ausmachen. Die Veröffentlichung eines Fachaufsatzes von Alex Krizhevsky, Ilya Sutskever und Geoffrey Hinton (2012) lieferte den Nachweis der Leistungsfähigkeit sogenannter neuronaler Netze im Hinblick auf automatische Objekterkennung in Bildern.[5] Es führte erstmalig vor, dass eine technisch umsetzbare algorith-

---

rierende GPU (*graphics processing unit*) umzulagern, war dabei von zentraler Bedeutung und zugleich ein Grund dafür, warum ML-Anwendungen zunächst aufgrund der technischen Anforderungen an GPUs kaum auf handelsüblichen PCs zum ›Laufen‹ zu bringen waren, vgl. Sudmann 2018a, S. 22ff. sowie kritisch Whittaker 2021.

5 Seine Beiträge in der Forschung zu künstlichen neuronalen Netzen seit den 1980er Jahren brachte Geoffrey Hinton 2019 den Turing-Award und zuletzt den Nobelpreis für Physik sowie den inoffiziellen Titel des »Godfather of AI« ein, vgl. etwa Ranosa 2019, Rothman 2023. Dass dieser Titel durchaus ambivalent besetzt ist und Hintons Rolle in Vergangenheit und Gegenwart der KI-Forschung kontrovers verhandelt werden, zeigt etwa die kritische Berichterstattung der öffentlichkeitswirksamen Aufgabe seines Postens bei Google im Jahre 2023, die er damit begründete, dass er infolge dieser freier über Risiken in Verbindung mit aktuellen Entwicklungen in Sachen KI/ML sprechen könne, vgl. Marx 2023.

mische Modellierung eines hinreichend quantitativ wie qualitativ umfassenden Datensatzes von Bildern es erlaubt, das Modell nach Abschluss der sog. Trainingsphase neue Bilder automatisch ›erkennen‹ zu lassen.[6] Die Bilder im Trainingsdatensatz sind mit den Bildinhalten korrespondierenden Tags (»Hund« und über weitere 1000 Klassen) versehen und der Erfolg der Forschung bestand darin, ein Modellierungsverfahren vorlegen zu können, dass die Inhalte von bisher ungekennzeichneten Bildern zuverlässig klassifizieren, d.h. die korrekten Tags produzieren kann. Das Verfahren lieferte signifikant bessere Ergebnisse als vergleichbare Ansätze der damaligen Zeit. Noch im gleichen Jahr erschien ein weiteres Paper, das weit über wissenschaftlichen Kreise hinaus in kompakter Form der Schlagzeile für Furore sorgte, dass KI nun ›völlig autonom‹ Katzen in YouTube-Videos Katzen erkennen könne. Die Studie einer Google-Forschungsgruppe wies nach, dass ihr auf einem Verfahren des sog. *unsupervised learning* beruhendes Modell ›von sich aus‹ begonnen hatte, Bilder, in denen Katzen zu sehen sind, konsistent einer Klasse zuzuweisen, obwohl das algorithmische Modell mit nicht vorab klassifizierten Bilddaten erstellt worden war und daher Katzen gar nicht ›kennen‹ konnte (vgl. Quoc et al. 2012).

Anschließende Forschung ebnete über die nächsten Jahre hinweg den Weg zu einem Zuverlässigkeitsgrad bei visueller Objekterkennung, der dem menschlicher Proband:innen in vergleichbaren Tests entspricht. Die Prominenz von sog. *human parity* als Maßstabsgröße zur Beurteilung von Verfahren maschinellen Lernens erstreckte sich bald auf weitere Anwendungsbereiche. Bald sollte *human parity* mutmaßlich auch in den Bereichen Spracherkennung und -übersetzung erreicht worden sein.[7] Seit der Verbreitung von sog. Large Language Models (LLMs), etwa der GPT-Reihe von Open AI, ist diese Losung wieder und wieder neu spezifiziert und variiert worden.

Die bislang erwähnten Beispiele beziehen sich allesamt allein auf die Verarbeitung von Sprach- und Bilddaten und Anwendungen für Endnutzer:innen. Es sollte keineswegs übersehen werden, dass Verfahren maschinellen Lernens auch in wirtschaftliche, journalistische und anderen professionellen Praktiken Einzug gehalten haben. Die diversen Einsätze von ML in dieser Vielzahl von Kontexten sind unlängst auch zum Gegenstand soziologischen Interesses geworden. So liegen mittlerweile in vielfältiger Form Studien zur praktischen Nutzung von ML vor, z.B. im Kontext etwa von Polizeiarbeit, in der Justiz, im Versicherungswesen, in der (Präzisions-)Medizin, im Personalmanagement, am Finanzmarkt oder in der Wissenschaft.[8] Besondere publizistische Aufmerksamkeit erhielt zudem schon früh die Forschung zu »Computer Vision« im militärischen Kontext des Einsatzes

6 Konkret handelt es sich um ein sog. *Deep Convolutional Neural Network*, vgl. ebd.

7 Vgl. kritisch zu diesen Behauptungen Beaver 2022.

8 Vgl. dazu entsprechend der Reihenfolge der Themenaufzählung Text exemplarisch: Andrejevic 2018, Egbert, Kaufmann et al. 2019, Esposito & Heimstädt 2021, Završnik 2021, Cevolini

von Drohnen in Kriegsszenarien, nicht zuletzt aufgrund der von dieser und weiterer Berichterstattung ausgehenden Skandalisierung der beobachteten Praktiken (vgl. etwa Richardson 2019). Überhaupt war die sprunghafte Zunahme öffentlichen Interesses an ML zuerst bei Bilderkennungsverfahren zu beobachten – auch oder gerade, weil deren Aufstieg zu diskursiver Prominenz von brisanten Enthüllungen begleitet wurde, die problematische Aspekte der Anwendungen offenbarten.[9]

An Bestimmungen aus der Informatik angelehnt beruht ML – schematisch formuliert – auf der algorithmischen Verarbeitung von (zumeist sehr umfangreichen) Datensätzen. Im Zuge dieses auch als *Training* bezeichneten Vorgangs wird eine Modellierung des Datensatzes errechnet. Dieses Modell hat die Form einer mathematischen Funktion, die in möglichst präziser Annäherung an den Datensatz dessen Strukturen – die Beziehungen einzelner, als Input und Output definierter Datenpunkte zueinander – beschreiben soll. Die errechnete Funktion ist im beschriebenen Sinne eine Datenanalyse, die in der Folge weitere Daten(-analysen) produzieren kann (vgl. Dourish 2016, S. 7). Die Bandbreite dieser Klasse statistischer Analysen ist groß und reicht von linearer Regression bis zu neuronalen Netzen, die Milliarden von Gewichtungen umfassen – etwa die bereits erwähnten Sprachmodelle.[10] Zentral für die Einordnung als Verfahren maschinellen Lernens ist die datenbasierte Modellierung eines Datensatzes als (partiell) automatisierte Errechnung einer Funktion, die von einem bestimmten Algorithmus ›gefunden‹ wird. Dieses »function finding« (Mackenzie 2017, S. 96) impliziert, dass die Funktion inputsensibel, d.h. in Abhängigkeit von der Datenbasis variabel ist und ggf. das Modell unter Berücksichtigung neuer Daten ›weiterrechnen‹, d.h. adaptiv reagieren kann. Dies unterscheidet klassische Machine Learning-Algorithmen wie *k-means* (vgl. MacQueen 1967) oder *support vector machines* (vgl. Cortes & Vapnik 1995) von anderen algorithmischen Verfahren, weshalb Algorithmen für sich nicht prinzipiell mit ML gleichzusetzen sind. Wenngleich sich hier eine diskursive Amalgamierung beobachten lässt – so weit,

---

& Esposito 2020, Hofmann 2023, Mateescu & Nguyen 2019, Lange, Lenglet & Seyfert 2016, Both 2020.

9 Man denke etwa automatischen Klassifizierung Schwarzer Menschen als Gorillas in der Foto-App von Google im Jahre 2015. Der Mechanismus konnte offenbar auch drei Jahre später nur durch Entfernen der gesamten Klasse ›Gorilla‹ – also Label wie auch zugehörige Bilder von Gorillas unterbunden werden, vgl. Vincent 2018.

10 Die im Falle sogenannter generativen Verfahren ergänzt werden um Vorhersagefunktionen im Sinne von ›Auto-Vervollständigen‹ durch Errechnung der höchstwahrscheinlichen nächsten Elemente des Modells. So können die generative von den »diskriminierenden« Funktionen in ML unterschieden werden, wobei das Attribut allein auf den Bezug der Funktion auf die Datengrundlage und nicht etwa auf die Beurteilung von Outputs bezogen ist. Probabilistisch modellierende Bayes'sche Netze stellen ein klassisches Beispiel für generative Verfahren dar, wohingegen Support Vector Machines (vgl. Cortes & Vapnik 1995), eingesetzt als Klassifikator oder Regressor, häufig zur Erklärung von diskriminierenden Funktionen herangezogen werden, so etwa in Jebara 2004, S. 5–12.

sodass auch in den für diese Studie maßgeblichen sozial- und geisteswissenschaftlichen Diskussionen nicht immer trennscharf zwischen ML, KI und Algorithmen unterschieden wird –, halte ich terminologische Differenzierung in diesem Zusammenhang für wichtig. Folglich hat diese Studie nicht etwa Algorithmen an sich oder KI überhaupt, sondern explizit ML zum Gegenstand.[11]

Mit Blick auf die für ML typischen algorithmischen Modellierungsverfahren ist charakteristisch, dass die derart produzierten Datenanalysen sich aufgrund der zugrundeliegenden Errechnungsmethode für Beobachter:innen in ihrem Sinngehalt oft als schwer nachvollziehbar erweisen, auch wenn sie sinnförmig verfasst (Text, Bild) sind. LLMs ›verstehen‹ nicht etwa eine Frage und ›überlegen‹ sich daraufhin eine Antwort, sondern errechnen für eine Eingabe (›Prompt‹) einen wahrscheinlichen Output auf Basis stochastischer Analysen der ihm zugrundeliegenden (häufig unüberschaubar großen) Datensätze.[12]

Dieses Problem gilt in aller Regel selbst (und gerade auch) dann, wenn die Eigenschaften des Datensatzes als *ground truth* des Modells im Sinne formaler Verteilungen und Regelmäßigkeiten als bekannt wie auch im Sinne der sinnhaften Bedeutung der Daten als ›verstanden‹ vorausgesetzt werden können (vgl. Dourish 2016, S. 7). Die mitunter überraschenden, wenn nicht gar unheimlich anmutenden Outputs der Modelle (vgl. Bucher 2017) sind daher nicht ohne Weiteres in ihrer Genese ›erklärbar‹ und lassen sich hinsichtlich ihrer Implikationen meist auch im Nachhinein nicht sinnvoll erschließen (vgl. Burrell 2016).[13]

In den vorangehenden Sätzen klang bereits die wesentliche Rolle von Datensätzen an, die auch in der Hinsicht nicht zu unterschätzen ist, dass die Verfügbarkeit hinreichend quantitativ großer wie qualitativ gehaltvolle Datensätze eine der wesentlichsten Voraussetzungen von ML und damit auch von dessen rasantem Aufstieg innerhalb der letzten etwa zehn Jahre ist. Vor dem Hype um ML war der Hype um ›Big Data‹ – etwa in Form des schon erwähnten *ImageNet*. Dieser Datensatz wurde 2006 von der Informatikerin Fei-Fei Li konzeptualisiert als Verknüpfung des damals schon existierenden *WordNet*[14] mit Bildern. Innerhalb von zweieinhalb Jahren

11 Florian Jaton berichtet in seiner empirischen Studie zur sozialen Genese von Algorithmen davon, dass interessanterweise unter praktisch mit der Entwicklung von algorithmischen Verfahren Beschäftigten bisweilen oft als unklar gilt, welche Art von Algorithmen als »real machine learning« gelten können, vgl. Jaton 2021, S. 269.

12 Dessen exakte Komposition zudem nicht öffentlich bekannt ist.

13 Unter dem Label »Explainable AI« oder »XAI« wird seit längerer Zeit daran gearbeitet, diesen Umstand zu beseitigen, bislang mit nützlichen, aber überschaubaren Ergebnissen, vgl. Zednik 2021 sowie Sudmann 2020, S. 194ff.

14 Einem seit den 1980ern entwickelten lexikalischen Datensatz, der Wörter der englischen Sprache in gegenwärtig 117000 distinkten »Synsets« eingruppiert. Dabei werden die semantischen Beziehungen der Synsets zueinander erfasst und jeder Eintrag enthält zudem eine Kurzdefinition, siehe Princeton University 2010.

gelang es Lis Team, eine 3,2 Millionen (manuell in 5247 »Synsets« klassifizierte [siehe Fn. 14]) Bilder umfassende Sammlung aufzubauen (vgl. Jaton 2021, S. 272) – mittlerweile umfasst ImageNet mehr als 14 Millionen Bilder und 21841 »Synsets« (vgl. ImageNet 2021). Bemerkenswert an diesem Projekt ist nicht zuletzt, dass für die händische Annotation der Bilder Amazons Dienst »Mechanical Turk«[15] in Anspruch genommen wurde, weil das Bewältigen dieser Aufhabe allein durch studentische Hilfskräfte wohl Jahrzehnte gedauert hätte (vgl. Fei-Fei 2010). Die Inanspruchnahme des Crowdsourcing-Dienstes für *ImageNet* ist damit vermutlich eines der frühesten, wenn nicht das früheste Beispiel für sogenannte »ghost work« im Zusammenhang mit ML.[16] Mit diesem Konzept bezeichnen Mary Gray and Siddharth Suri diejenige von Menschen – zumeist (auch im Falle von AMT) aus dem Globalen Süden – verrichtete Arbeit, ohne die sogenannte Künstliche Intelligenz in ihrer gegenwärtigen Form nicht realisiert werden könnte (vgl. Gray & Suri 2019, S. IX).

Dieser Umstand berührt gleichsam unmittelbar den eben erwähnten ›Durchbruchauufsatz‹ (vgl. Krizhevsky et al. 2012). Der in Anlehnung an den Vornamen des Erstautoren Alexander Krizhevsky als *AlexNet* bekannte Ansatz nutzt nämlich den ImageNet-Datensatz (zu dem Zeitpunkt 1,2 Mio. Bilder mit 1000 Klassen) als »Ground Truth« und stellte seine Leistungsfähigkeit im Rahmen der »ImageNet Large Scale Visual Recognition Challenge« (vgl. ImageNet 2021b) im Jahre 2012 unter Beweis, die *AlexNet* mit signifikantem Abstand zum nächstplatzierten gewann (vgl. Wei 2019). Aufgrund dieser Verstrickung erscheint es nicht vermessen, die umfangreiche Annotationsarbeit für *ImageNet* als so wichtigen wie ungewürdigten Beitrag zum Erfolg von *AlexNet* zu verstehen. Es deutet sich hier bereits an, dass die soziologische Betrachtung von ML instruktive Einsichten womöglich weniger aus der Beobachtung der Nutzung von ML, sondern vor allem auch der Genese entsprechender technischer Anwendungen gewinnen kann. So präsentiert dieser erste kursorische Überblick ML als räumlich verteilten Prozess, in dessen Verlauf verschiedene Ressourcen mobilisiert werden. In einer solchen diachronen Perspektive auf die Genese von ML-Anwendungen erscheinen diesen weniger als disruptive radikale Transformationen und vielmehr als Akkumulationsvorgänge (vgl. Mackenzie 2017, S. 5).[17]

---

15 Zum Hintergrund der Crowdsourcing-Plattform vgl. Crawford 2021. S. 63–69, bes. 67f. sowie ausführlicher zur Geschichte des »Mechanical Turk«: Standage 2022.

16 Ab 2007 waren jährlich wohl ca. 20–30.000 Menschen als Arbeitskräfte an der Klassifizierung der ImageNet-Bilder beteiligt, vgl. Markoff 2012.

17 Diese Betrachtungsebene lässt indes auch den sich in der gleichen Zeit vollziehenden Aufstieg von Sozialen Medien in neuem Licht erscheinen. Versteht man diesen nämlich vor allem als Aufstieg von Data Capture-Technologien und Cloud Services (vgl. Halpern et al. 2022), die Nutzer:innendaten zur weiteren Verwendung (und Verwertung) auf privaten Servern speichern, wird leicht nachvollziehbar, dass Soziale Medien und ML alsbald in einen datenökonomischen Zusammenhang geraten würden.[17] Dieser wurde bald als Plattform- und/oder Überwachungskapitalismus (vgl. etwa Srnicek 2018, Zuboff 2018) beschrieben – und beruht

## Ein dynamischer Gegenstand

Die beschriebenen Entwicklungen seit 2012 (*AlexNet*) bzw. schon 2007 (*ImageNet*) haben schnell sozial- und geisteswissenschaftliches Interesse auf sich gezogen. Mittlerweile liegt ein vielseitiges Korpus sozial- und geisteswissenschaftlicher Forschung zu ML vor, die sich in den letzten Jahren entwickelt, differenziert und konsolidiert hat. Mittlerweile ist die Rede von »Critical AI Studies« als einem – in der Formulierung der Medien- und Literaturwissenschaftler:innen Rita Raley und Jennifer Rhee (2023) – interdisziplinär operierenden »field in formation«. ML bzw. KI scheint also, so ließe sich daraus schließen, neue Ansätze geistes- und sozialwissenschaftlicher Forschung erforderlich gemacht zu haben, die jüngeren Datums sind oder überhaupt erst noch gefunden werden müssen (»… in formation«). Eine solche Proklamation lässt sich als Versuch der Konsolidierung weit verzweigter Forschungen verstehen, der auf Umstände reagiert, die auch für die vorliegende Studie von zentraler Bedeutung sind.

Eine solche Konstatierung belegt zudem, dass nicht immer völlig klar scheint, was unter ML/KI zu verstehen sei und wie sich die Forschung – theoretisch wie methodisch – auf ihren Gegenstand beziehen sollte. Dies verweist wiederum darauf, dass »Künstliche Intelligenz« als diskursive Referenz nicht einheitlich verwendet wird. Dies verwundert wenig, da schon ein historischer Blick auf KI die Geschichte eines *moving target* offenbart (vgl. etwa Fazi 2021, S. 6). Bei diesem handelt es sich um ein kontext-, situations- und interessenabhängig variabel bestimmbares Phänomen, dessen wechselnde Verständnisse das Spannungsfeld zwischen dem technisch Möglichen – ›künstlich‹ erzeugt – und dem vermeintlich das Menschliche Auszeichnenden – häufig als Maßstab für ›Intelligenz‹ – und beständig aufs Neue ausgelotet haben.[18]

Schon die Hintergründe der ersten bekannten Verwendung von »Artificial Intelligence« im Jahre 1955 belegen, dass ein wesentliches Anliegen hinter dieser

---

im Hinblick auf Wertschöpfung nicht unwesentlich auf global distribuierter »ghost work«. Wobei gewissermaßen fraglich scheint, ob (»digitaler« [Staab 2019]) »Kapitalismus« zur Beschreibung der ›Informationsökonomie‹ überhaupt noch die richtige Bezeichnung ist. Maßgeblich dafür ist die Beurteilung der Frage, ob Daten i.S.v. Information als Kapital verstanden werden sollten, vgl. Wark 2019.

18 Sinnbildlich dafür gilt der dem Informatiker Larry Tesler zugeschriebene Satz: »AI is whatever hasn't been done yet«, so etwa populär etwa Hofstadter 1979, S. 601. Tesler jedoch hat eigenen Angaben zufolge stattdessen Folgendes von sich gegeben: »›Intelligence is whatever machines haven't done yet.‹ Many people define humanity partly by our allegedly unique intelligence. Whatever a machine – or an animal – can do must (those people say) be something other than intelligence« (Tesler [o.J.]). Was sich viel eher als Kritik eines Anthropozentrismus liest, der Menschlichkeit mit Intelligenz schlechthin liest, wurde daher vermutlich nicht ganz im Sinne des Urhebers zum geflügelten Wort.

Wortschöpfung darin bestand, einen für Finanzierungsanträge attraktiven Titel für das mit Mitteln auszustattende eigene Forschungsvorhaben präsentieren zu können (vgl. Moor 2006). Anlass dieser ersten prominienten Proklamation war bekanntlich das »Dartmouth Summer Research Project on Artificial Intelligence« und damit zunächst dessen Finanzierungantrag, ausgearbeitet und eingereicht von John McCarthy, Marvin Minsky, Nathaniel Rochester und Claude Shannon. Die im Antragstext des für den Sommer 1956 angesetzten (und durchgeführten) achtwöchigen Workshops gesetzten abstrakten Ziele lesen sich auch heute noch nachvollziehbar. Das nach wie vor prominente Schlagwort »Learning« ist ebenso enthalten – wie interessanterweise auch ein expliziter Bezug auf das Medium Sprache, das im Kontext aktueller Debatten um LLMs in aller Munde ist:

> The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. (McCarthy et al. 1956, S. 1)

Aus diesem Ursprungsszenario heraus hat sich für das Feld der KI-Forschung eine temporale Paradoxie hinsichtlich des durch ›KI‹ Bezeichneten ergeben. ›KI‹ (und gleichsam die Spezifizierung ›ML‹) bezieht sich immerzu auf aktuell die Forschung beschäftigende Probleme, die (noch) nicht technisch lösbar sind. Es handelt sich bei KI/ML daher in wesentlicher Hinsicht häufig primär im Wortsinne um eine Projektion.

Versuche, qua der Markierung eines wissenschaftlichen »field in formation« Übersicht zu schaffen, hängen von gegenwärtigen Visionen über zu Erreichendes, aber ebenso vom jeweils aktuellen gegebenen faktischen ›Stand der Technik‹ ab. Dieser entwickelt sich mit Blick auf KI – aktuell mit besonderem Fokus auf ML – ebenso dynamisch. Dies haben die vergangenen zehn Jahre Technikentwicklung wohl unbestreitbar gezeigt. ALs *moving target* erweist sich KI/ML daher mindestens in zweifacher Hinsicht.

›Durchbrüche‹ innerhalb dieses Feldes können die soziale Bedeutung und Relevanz von KI maßgeblich verändern, wenn etwa die Veröffentlichung neuer Produkte innerhalb kurzer Zeit die Defizite der vorigen vergessen macht. Sofern technische Neuerungen nicht zugleich terminologische Verschiebungen mit sich bringen, sondern unspezifisch als »KI« verhandelt werden, kann dies den Stand wissenschaftlicher Forschung schnell altern lassen. Dies gilt nicht nur für Vorhersagen über die Potenziale wie auch die Risiken von KI, die beständig Gefahr laufen, schnell widerlegt zu werden. Auch weniger an Spekulationen interessierte Forschungen stehen dem Problem gegenüber, den Anspruch auf generalisierbare Aussagen zum untersuchten Phänomen mit dessen permanenter Vorläufigkeit ver-

mitteln zu müssen. KI-Forschung ist daher als ein »field in formation« zu verstehen – ebenso wie ihr Gegenstand. Dieser ist nicht nur semantisch ein *moving target*, sondern muss auch in seiner je aktuellen technisch-materialen Gestalt immerzu als Prototyp verstanden werden.[19] Wichtig scheint mir dabei die Beobachtung, dass sich abhängig vom jeweiligen ›state of the art‹ auch der soziale Bezugsrahmen für ML verändert. Rückblickend etwa lässt sich sehen, dass im Bereich visueller Kunst die für bildgebende ML-Verfahren seit 2015 so charakteristische »visual indeterminacy« (Hertzmann 2020) vor allem Ausdruck der Möglichkeiten und Grenzen von sogenannten GAN-Architekturen war. Seit dem Einzug von Verfahren wie *Stable Diffusion* (im Jahre 2022), die nahezu fotorealistische Bilder zu generieren imstande sind, kann nicht mehr davon die Rede sein, dass visuelle »KI-Kunst« gerade durch eine charakteristische Unschärfe der Bildelemente gekennzeichnet sei.[20]

Mit Blick auf jüngere Entwicklungen könnte auch über das Feld der Kunst hinaus schnell der Eindruck entstehen, als hätte es allein seit 2020 eine Reihe von so folgenreichen Innovationen in der Entwicklung von KI-Verfahren und -Produkten gegeben, dass fraglich scheinen könnte, inwiefern die hier primär aus empirischer Feldforschung gewonnenen Beschreibungen überhaupt in irgendeiner Form generalisierbare Gehalte zur Verfügung stellen können. Vor allem im Feld der Verarbeitung und Generierung von Text- und Bildinhalten – zunehmend gilt dies auch für auditive Formen – zeigt sich die technische Realität gegenüber dem, was vor wenigen Jahren noch als innovativ angesehen wurde, als eine entschieden andere. *CLIP*, *DALL-E*, *Stable Diffusion*, *Midjourney*, *ChatGPT*, *Gemini*, Grok, *Copilot*, *DeepSeek*, *Claude* (Anthropic) – seit 2020 wurde eine ganze Reihe von Bild- und Sprachmodellen veröffentlicht, denen nachgesagt wird, ihre jeweiligen ›Vorgänger‹ in einer Vielzahl von Hinsichten übertroffen zu haben, was Leistungsfähigkeit und damit verbunden mutmaßlich auch Nützlichkeit in Anwendungsform angeht. Im Umkehrschluss heißt dies, dass vormals aktuelle Verfahren schnell in Vergessenheit geraten können.

Mit dieser Einsicht, dass bestimmte Formen von ML mitunter ›schnell altern‹, möchte ich sensibilisieren für potenzielle Differenzierungslinien in Bezug auf den Gegenstand dieser Studie. Der ausführlichen Auseinandersetzung mit dieser Problematik gewissermaßen vorauseilend soll deren Erwähnung an dieser frühen Stelle dem Erwartungsmanagement dienen: Empirischer Gegenstand dieser Studie sind spezifische Formen von ML zu einer gegebenen Zeit, als Praktiken darüber hinaus

19 Vgl. zur technischen Bedeutung des Prototyping bereits vor 40 Jahren Floyd 1984 sowie allgemeiner zum gesellschaftlichen Stellenwert des Phänomens Dickel 2019.

20 Ausführlich zu dieser Periodisierung mit der Unterscheidung GAN/Post-GAN im Hinblick auf bildgebende Verfahren im Kontext von zeitgenössischer KI-Kunst vor/seit 2019 vgl. Offert 2023.

auf spezifische Art und Weise situiert. Abstrahierende Generalisierungen zu maschinellem Lernen oder gar Künstlicher Intelligenz sind nicht das primäre Anliegen der Untersuchung und können nur unter bestimmten Voraussetzungen gewonnen werden. Damit ist zugleich freilich nicht ausgeschlossen, dass sich über verschiedene Generationen von ML hinweg konsistente Problembezüge in den mit ML verbundenen sozialen Praktiken ergeben – etwa für den beschriebenen Zeitraum seit 2012 – wie an späterer Stelle zu zeigen sein wird. Dennoch scheint mir dieser Hinweis zur Vorsicht an dieser frühen Stelle durchaus geboten, besonders mit Blick auf mediale Diskursdynamiken, innerhalb derer definitive Festlegungen zu den langfristigen Potentialen, Schwierigkeiten und Gefahren im Zusammenhang mit ML an der Tagesordnung sind.

Im Rahmen dieser kursorischen thematischen Sondierung, der in den Abschnitten zur Soziologie künstlicher Intelligenz eine systematische Erörterung des Untersuchungsgegenstandes folgen wird, soll nun nach dieser ersten Annäherung an ML mittels der Schilderung wichtiger Ereignisse und Dynamiken der letzten etwa zehn Jahre die Perspektive gewechselt werden.

## KI/ML-Diskurse

Innerhalb der letzten Jahre ist KI/ML innerhalb des sozial- und geisteswissenschaftlichen Diskurses zu einem prominenten Thema avanciert. Schon in quantitativer Hinsicht beeindruckt das rasant gestiegene Interesse, das sich unter anderem in der Zahl einschlägiger Artikel, Bücher und Schwerpunkthefte, sowie – auf institutioneller Ebene – neu ins Leben gerufener Fachzeitschriften und spezifisch denominierter Professuren, wenn nicht ganzer Forschungseinrichtungen ausdrückt.[21]

In diesem Abschnitt versuche ich einführend eine knappe Sondierung prägnanter Tendenzen dieses Diskurses in schematisierender Form, die heuristisch drei distinkte Positionen identifiziert. In erster Annäherung an die sozial- und geisteswissenschaftliche Literatur zu KI/ML werde ich nicht unmittelbar auf eine systematische soziologische Auswertung der bisherigen Forschung abzielen. Stattdessen geht es mir zunächst darum, einen Eindruck der lebhaften Debatten zu vermitteln, wie sie für den gegenwärtigen Diskurs um KI/ML dies und jenseits fachwissenschaftlicher Grenzen charakteristisch sind. Damit verfolge ich die Absicht, für die Eigenheiten eines Forschungsgegenstandes zu sensibilisieren, der nicht nur, wie beschrieben, schon immer ein *moving target* gewesen ist, sondern darüber hinaus auch gegenwärtig so anhaltend wie lautstark kontrovers verhandelt wird. Wenngleich sich

21 In der erwähnten Einführung zum Feld der »Critical AI Studies« findet sich eine Auflistung vor allem US-amerikanischer (geistes-)wissenschaftlicher Einrichtungen, vgl. Raley & Rhee 2023, S. 198f.

soziologische Forschung von solchen Dynamiken jenseits der eigenen Fachgrenzen, vor allem in Absicht auf normative Beurteilungen von KI/ML, nicht übermäßig oder gar etwa in der eigenen Theoriearbeit beeinflussen lassen sollte, so sind diese doch zumindest auf angemessene Weise als potenziell relevante Bestandteile des Forschungsgegenstandes einzuordnen. Dies gilt besonders vor dem Hintergrund der Vermutung, dass entsprechende diskursive Dynamiken auch vor empirischen Szenarien, wie ich sie im Kontext dieser Studie untersuche, kaum Halt machen werden. Und darüber hinaus ist im Sinne einer reflexiven methodologischen Herausforderung wohl allemal anzuerkennen, dass außerfachliche Diskurse zu innerfachlich verhandelten Gegenständen den soziologischen Diskurs vielmehr durchaus informieren und nicht etwa völlig unberührt lassen. Schließlich gilt, dass auch wir Soziolog:innen das, was wir über unsere Gesellschaft wissen, doch vor allem aus den (Post-)Massenmedien wissen (vgl. Luhmann 1996, S. 9).

Zur Differenzierung diskursiver Positionen zu KI/ML beziehe ich mich auf Umberto Ecos Unterscheidung von »Apokalyptikern« und »Integrierten« (Eco 1984a [1962]), die dieser im Zuge einer Studie zur Kritik an der Massenkultur der 1960er Jahre entwickelt hatte. Mit diesen Bezeichnungen charakterisiert Eco insbesondere – in Auseinandersetzung mit der Frankfurter Schule (vgl. Eco 1984b, S. 11) – Modi der kritischen Bezugnahme auf die moderne Massenkultur. Erscheint diese Apokalyptiker:innen grundlegend als in einem Zerfallsprozess befindlich, den sie als letzte echte »›Kulturmensch[en]‹« (ebd.) bezeugen, erleben Integrierte sie vielmehr als sich entwickelnd, erweiternd und auf zu begrüßende Weise wandelnd. Ecos Schema betrifft den normativen wie operativen Charakter diskursiver Bezugnahmen auf einen Gegenstand im Modus der Kritik bzw. Affirmation: Integrierte seien die Produzent:innen von Texten der Massenkultur, Apokalyptiker:innen hingegen schrieben lieber *über* sie. Im Zuge der Distanznahme qua Kritik[22] stifteten sie ihren Leser:innen dennoch Trost, da ihre Positionierung Identifikationspotenzial biete: »er läßt ihn, vor dem Hintergrund der drohenden Katastrophe, die Existenz einer Gemeinschaft von ›Übermenschen‹ erahnen, die sich über die Banalität und den ›Durchschnitt‹ zu erheben vermögen […]« (ebd.); eine »auserwählte *community*« (Eco 1984a [1962], S. 17) derer, die es besser wüssten.

> Während die Apokalyptiker gerade dadurch überleben, dass sie Theorien über den Zerfall ausbilden, versagen sich die Integrierten weitestgehend der Theoriearbeit; sie erzeugen und übermitteln ihre Botschaften in unbefangener Leichtigkeit, tagtäglich auf allen Ebenen. Die Apokalypse ist eine Besessenheit der dissenters, des Andersdenkenden; die Integration ist die konkrete Realität derjenigen, die nicht abweisen, nicht anderer Meinung sind. (ebd., S. 16)

22 Etwa mittels solcher begrifflicher Konstruktionen wie »Kulturindustrie« (vgl. ebd., S. 19).

Nun handelt es sich bei der Massenkultur, wie Eco sie sich zum Thema seiner Studie machte, um einen wesentlich anders verfassten Forschungsgegenstand als ML/KI, der zudem in der Gegenwartsgesellschaft aus einer Vielzahl von Gründen gleichsam anders zu kontextualisieren wäre. Daher ist eine Modifikation des Eco'schen Modells vorzunehmen, um zu einem aussagekräftigen analytischen Schema für die Beurteilung des KI/ML-Diskurses zu gelangen. Statt von zwei soll von drei maßgebliche Positionierungen ausgegangen werden, die entsprechend der Eigenheiten von KI/ML durch erweiterte Beurteilungskriterien bestimmt sein sollen.[23] Diese werden nicht primär hinsichtlich ihres normativen Bezugs (Affirmation/Kritik) auf ML schlechthin unterschieden, sondern in zeitlich differenzierter Form einerseits auf ML in seiner gegenwärtiger gesellschaftlichen Form und andererseits auf ihre Prognosen hinsichtlich der Zukunft von KI/ML bezogen. Diese zwei Kriterien werden darüber hinaus um das der kognitiven Beurteilung von KI/ML ergänzt. In Anlehnung an Niklas Luhmanns Unterscheidung von kognitiven und normativen Erwartungen (vgl. Luhmann 1991, S. 138f.) sowie in Anbindung an das Anliegen der vorliegen Studie soll festgehalten werden, inwiefern in den referierten Positionen ML als hinreichend verstanden (und mithin als Forschungsgegenstand gewissermaßen ›erledigt‹) oder aber als genuines wissenschaftliches Problem markiert wird und mithin als bislang eben nicht hinreichend verstanden gilt.

## Drei Positionen zu KI/ML

Ein erstes ›Lager‹ halte ich mit dem Schlagwort *Gegenwartsapokalypse* (im Folgenden kurz: *GA*) für treffend beschrieben. Diesem sollen solche Positionen zugerechnet werden, die im Modus immanenter (und häufig mit aktivistischen Absichten verbundener politisch-engagierter) Kritik auf grundlegende Probleme im Zusammenhang mit aktuellen Anwendungen von ML und deren Auswirkungen hinweisen. Apokalyptischen Charakter bekommen diese Perspektiven dabei weniger durch die Imagination einer zur willkürlichen Setzung eigener Handlungsziele fähigen, übermächtigen Form von KI/ML, die ›von sich aus‹ böswillige Anliegen entwickelt und verfolgt. Vielmehr geht es um die Rolle von KI/ML im Kontext bestehender gesellschaftlicher Gewalt- und Ungerechtigkeitsregime und drohender, wenn nicht schon manifester Klimakatastrophen. Als rechenintensive und daher ressourcenkonsumierende Technik sei von KI/ML in den Händen der Falschen eine weitere Verschlechterung der gesellschaftlichen (und planetarischen) Zustände zu

23 Wie zu sehen sein wird, weisen die folgenden Positionen bei allen maßgeblichen Unterschieden allerdings eine Gemeinsamkeit darin auf, dass sie KI/ML bzw. (Digital-)Technik nicht grundlegend kategorisch ablehnen. Derartige gewissermaßen technophobe Positionen mag es auch geben, doch spielen sie im KI/ML-Diskurs m.E. keine nennenswerte Rolle.

erwarten. Die Ergänzung des Zeitbezuges soll indes darauf hinweisen, dass GA-Positionen nicht notwendigerweise eine unaufhaltsame Verfallsgeschichte erzählen, sondern einerseits von nostalgischen Bekundungen absehen und andererseits die Zukunft insoweit als offen verstehen, dass sie nicht vollends resignieren und jedenfalls die Möglichkeit einer ›besseren‹ Form von KI/ML für denkbar halten. Wenngleich die Fortsetzung aktueller Tendenzen katastrophale Folgen mit sich brächte, gäbe es doch noch alternative Trajektorien für KI/ML. Ich rechne diesem Lager eine Reihe von Positionen zu, wie sie insbesondere im erwähnten Feld der geistes- und sozialwissenschaftlichen Feld der »Critical AI Studies« (Raley/Rhee 2023) zu finden sind. Wohl nicht gänzlich überraschend weisen die entsprechenden Positionen eine Affinität zu Kritischer Theorie im Sinne der Frankfurter Schule und feministischen, ökologischen, dekolonialen sowie/bzw. antirassistischen Ansätzen aus.[24] Als politisch engagierte Forschungsansätze stellen sie der gegenwärtigen Realität von KI/ML ein quasi-apokalyptisches Zeugnis aus und explizieren oft selbst Wege in Richtung einer besseren zukünftigen Lage, im Kontext derer sie wohl integriert gelten könnten, wenn diese Wege auch als fragil zu gelten haben und der Horizont einer nicht-apokalyptischen Zukunft zunehmend außer Reichweite zu geraten drohe (vgl. Burrell 2023).

Diesem Positionenbündel steht das Lager der *Gegenwartsintegration* (*GI*) gegenüber, innerhalb dessen die jüngeren Innovationsdynamiken um ML prinzipiell gutgeheißen werden. Dieses nicht selten technikindustrienahen Positionen zeigen sich angetan von bisherigen Entwicklungen um KI/ML und werden unter anderem durch das Führungspersonal von an diesem Prozess beteiligter Unternehmen wie Elon Musk, Eric Schmidt, Shane Legg oder Sam Altman vertreten. Ebenso wird es durch prominente (Technik-)Wissenschaftler wie etwa Yann LeCun oder die bereits erwähnten Geoffrey Hinton (bis 2023 Google, vgl. Fn. 26) und Ilya Sutskeyer (OpenAI)[25] öffentlichkeitswirksam repräsentiert. Diese Positionierung leuchtet nicht zuletzt als Ausdruck der Interessen einer Allianz von industrieller und wissenschaftlicher Forschung ein. Dabei ist – wie oben bereits angedeutet – zu beobachten, dass wirtschaftliche (Plattform-)Unternehmen zunehmend als Akteure der KI-Forschung in Erscheinung treten und aufgrund ihrer finanziellen Möglichkeiten gerade in Bereichen der Forschung erfolgreich sind, die hohe Anforderungen an Rechenleistung stellen, etwa an den erwähnten LLMs.

---

24 Vgl. etwa mit Bezug auf Aspekte von Ausbeutung, Überwachung und Klassengesellschaftlichkeit etwa Pasquinelli 2024, Noble 2018, Apprich et al. 2018, Whittaker 2021; mit Fokus auf Ökologie und Klima Bender et al. 2021, Mullaney et al. 2021; sowie bzgl. postkolonialer und rassistische Aspekte in KI/ML etwa Amaro 2022, Chun 2021, Gebru & Torres 2024.

25 Hinton und Sutskeyer sind infolge des Aufschwungs um ML-Ansätze seit Veröffentlichung ihres oben vorgestellten Aufsatzes im Jahre 2012 schließlich in die Forschungsabteilungen privatwirtschaftlicher Unternehmen gewechselt.

Tatsächlich spielen derart die gegenwärtige gesellschaftliche Realität von ML normativ affirmierende orientierte Positionen im geistes- und sozialwissenschaftlichen Diskurs um KI/ML keine relevante Rolle und kommen innerhalb der Fachdebatten in der Regel nur als Teil der Beschreibung des ML/KI-Feldes vor. Das Lager der *GI* soll an dieser Stelle dennoch genauer in den Blick genommen werden, da es gegenwärtig eine interessante Metamorphose zu durchleben scheint. Offensichtlich verfügt es über öffentlichkeitswirksame Kanäle, weil es nicht durch zuletzt das Führungspersonal von maßgeblich in der KI-Entwicklung tätigen Plattformunternehmen verkörpert wird. Aus ebendiesem Milieu heraus wird seit mehreren Jahren auf eindringliche Weise vor den Gefahren der Weiterentwicklung aktueller ML-Verfahren unter den gegenwärtigen Vorzeichen gewarnt. Diese könne zur Folge haben, dass autonome *AGI* (*Artificial General Intelligence*) oder andere Formen von ›Superintelligenz‹ (vgl. Bostrom 2014) entstünden, die sich nicht mehr kontrollieren ließen und möglicherweise gegen ›die Menschheit‹ richten könnten. Es ist deshalb mitunter die Rede von ›existenziellen‹ Risiken in der Weiterentwicklung von KI/ML die Rede, die es einzudämmen gelte. Solche Ansichten werden gegenwärtig zuhauf und großem Medienecho publiziert, etwa in Form offener Briefe. Derartiges geschah etwa im Frühjahr 2023, als im März unter anderem von Elon Musk, dem Informatiker Yoshua Bengio und dem Historiker Yuval Noah Harari zum Aussetzen der Forschung an LLMs für ein halbes Jahr aufgefordert wurde, da diese ein »profound risk to society and humanity« (Bengio et al. 2023) darstellen könnten. Zwei Monate später sahen sich unter anderem Open AI-CEO Sam Altman wie auch der ›AI Godfather‹ Geoffrey Hinton und Ilya Sutskeyer (zu dem Zeitpunkt *Chief Scientist* bei OpenAI) zu einem »Statement of AI Risk« bewegt (Hinton et al. 2023).[26] Ich bezeichne die *Gegenwartsintegrierten* deshalb als zumindest in Teilen als Lager mit einer Affinität zu Positionen, die als *zukunftsapokalyptisch* verstanden werden können.

Damit sind die Prämissen der Unterscheidung Umberto Ecos umgekehrt und eine widersprüchlich anmutende Dynamik markiert. Es sind nun die ehemals Integrierten, die apokalyptische Szenarien beschwören und dafür ihrerseits von den ›eigentlichen‹ Apokalyptiker:innen (der Gegenwart: *GA*) für ihre falsche Version der Apokalypse kritisiert werden. Im Sinne schismogenetischer Dynamiken könnten diese sich in dieser Konstellation dazu gezwungen sehen, stärker als vorher zu markieren, dass ihr Narrativ gefährlicher Tendenzen im KI/ML-Feld vielleicht doch nicht ganz so apokalyptisch war – oder zumindest unbedingt klarmachen, dass es ihnen um eine andere, gewissermaßen ›authentischere‹ Version von KI-Apokalypse

26 Als Randnotiz ist hierbei festzuhalten, dass Hinton, der vorher kaum als öffentliche Figur in Erscheinung getreten war, vor diesem Bekenntnis seinen Posten bei Google aufgab, um sich laut eigener Aussage als individuelle Person und eben nicht mit Google Affiliierter äußern zu können.

ging.[27] Ihre Kritik an der *GI*-Position richtet sich daher besonders auch auf deren Instrumentalisierung des apokalyptischen Gedankens.[28] Damit werde die Angst vor dem Einbruch der Apokalypse als Vorwand genutzt, um besser Aussichten auf zukünftige finanzielle Zuwendungen durch Investments zu gewährleisten, aber auch die eigene Position dies und jenseits von KI/ML zu stärken.[29] Auch gegenüber politischen Regulationsversuchen brächten sich ›Neo-Apokalyptiker:innen‹ in Stellung, schließlich könnten sie so für sich in Anspruch nehmen, die drohenden Gefahren im Blick zu haben und glaubwürdig anmutend behaupten, selbst mit größtem Interesse an sicherer und ›sozialverträglicher‹ (schon im Sinne von ›friedlicherer‹ oder ›katastrophale Folgen bestmöglich verhindernder‹) KI zu arbeiten (vgl. etwa Burrell 2023, Gebru & Torres 2024).

Ein drittes Lager diskursiver Positionen scheint mir schließlich mit *Epistemische Krise (EK)* treffend beschrieben. Diese Positionen zeichnet aus, dass sie die kognitive Herausforderung in der Auseinandersetzung mit KI/ML in den Vordergrund stellen. Wenngleich sie positive wie negative realweltliche Effekte durchaus nicht leugnen, entsprechen ihre kritischen Perspektiven auf KI/ML doch zumeist Beobachtungen zweiter Ordnung. So etwa, wenn sie in KI/ML das Potenzial einer weiteren narzisstischen Kränkung der Menschheit im Sinne Sigmund Freuds vermuten – wobei freilich den ›allzu humanistisch‹ Argumentierenden (vor allem

27 Das Konzept der Schismogenese geht auf Gregory Bateson zurück, der dieses im Kontext seiner Forschung zur Beschreibung von sozialen Beziehungsdynamiken »progressiver Differenzierung« (Bateson 1983a, S. 108) innerhalb von (ethnografisch beobachteten) Kollektiven entwickelte: Mitglieder einer Gruppe A verhalten sich gegenüber Mitgliedern einer Gruppe B anders als gruppenintern, und Gleiches gilt umgekehrt. Sofern das Verhalten gegenüber der anderen Gruppe von dieser als provokant wahrgenommen wird – etwa verunglimpfend oder (so bei Bateson) prahlerisch –, kann diese geneigt sein, sich zu verteidigen und den so wahrgenommenen Angriff zu erwidern, woraufhin wiederum komplementäre Reaktionen erwartet werden können usw. Es ergibt sich darin eine Dynamik wechselseitiger Verstärkung, die zu immer schärferen Angriffen, dabei aber zugleich auch zur Ausbildung und Stabilisierung der eigenen Gruppenidentität fuhrt (vgl. ebd.).

28 Diese Überlegung scheint mir parallel zu einem Gedanken Ecos hinsichtlich des Nutzens apokalyptischer Schwarzmalerei zu liegen, wenn dieser nämlich fragt: »Sind die apokalyptischen Visionen möglicherweise das raffinierteste Produkt, das sich dem Massenkonsum darbietet?« (Eco 1984a, S. 17).

29 Das Anliegen, die Menschheit vor KI zu retten, sei daher ein vorgeschobenes. Tatsächlich zeige sich im Lager der GI vor allem eine Affinität zu eugenischem Gedankengut. Das technikindustrielles Milieu derer, die an *AGI* bzw. ›Superintelligenz‹ arbeiten, zeichne sich jenseits der offiziellen Behauptungen des noblen humanistischen Kernanliegens jedoch vor allem durch rassistische, xenophobe, klassistische, ableistische und sexistische Einstellungen aus (vgl. Gebru & Torres 2024). Das Akronym *TESCREAL* fasst die maßgeblichen Ideologien zusammen, die sich in diesem Milieu anzutreffen seien, zu denen »transhumanism, Extropianism, singularitarianism, cosmism, Rationalism, Effective Altruism, and longtermism« (ebd.) zu zählen wären.

# Danksagung

Die Arbeit an dieser Studie hat mir viele inspirierende Momente bereitet, aber auch einiges abverlangt und mich in mancher Hinsicht bis an meine Grenzen geführt. Dass ich sie dennoch erfolgreich abschließen konnte, verdanke ich der Unterstützung und Hilfe aller, die mich im Zeitraum ihres Entstehens begleitet haben.

Dominik Schrage habe ich dafür zu danken, dass er mir 2019 den Weg in das damals in Entstehung befindliche Schaufler Lab@TU Dresden geebnet und als Erstbetreuer meiner Arbeit in den vergangenen Jahren neben zahllosen instruktiven Ratschlägen so viel Rückhalt geboten und Zuversicht vermittelt hat, dass es für mich und mein Projekt schlicht unmöglich wurde, vom angetretenen Pfad wieder abzukommen. Dirk Baecker danke ich für seine kritisch-konstruktive Begleitung meines Projekts, die ich als anregend wie herausfordernd und – gerade deshalb – als außerordentlich produktiv empfunden habe. Bei Susann Wagenknecht möchte ich mich herzlich für die Vermittlung neuer Perspektiven in intensiver Auseinandersetzung mit meiner Arbeit bedanken sowie vor allem für die anregende Zusammenarbeit an einem Aufsatz, die es mir in einer kritischen Phase erlaubt hat, erst wirklich zu verstehen, worum es mir in meinem Projekt überhaupt geht. Andreas Ziemann danke ich für die engagiert-richtungsweisende Betreuung meines Projekts in dessen Frühphase, als es bereits in Entwicklung, aber noch ohne Richtung war.

Den Sprecher:innen des Schaufler Lab@TU Dresden, Lutz Hagen und Kirsten Vinzenz, danke ich für das mir entgegengebrachte Vertrauen und insbesondere die vielen Freiräume, die ich als Stipendiat genießen konnte. Bei Jonas Wietelmann, Anke Woschech und Gwendolin Kremer möchte ich mich für eine angenehme und produktive Zusammenarbeit auf Augenhöhe bedanken. Ebenso möchte ich Kerstin Schankweiler und Christian Schwarke für die vielen netten Begegnungen im Zusammenhang mit dem Lab und meinem Projekt meinen Dank aussprechen. Bei meinen Leidensgenoss:innen und Kollaborateur:innen im Lab – Gina Glock, Miriam Gorr, Rita Jordan, Michael Klipphahn-Karge, Ann-Kathrin Koster, Sandra Mooshammer, Lukas Nehlsen, Rebekka Roschy, Philipp Preußger und Susanne Rentsch – möchte ich mich für eine schöne gemeinsame Zeit bedanken. The Schaufler Foundation sei an dieser Stelle sehr für die großzügige Unterstützung des Lab und die engagierte-interessierte Begleitung unserer Aktivitäten insbeson-

dere in Person von Ingo Smit gedankt. Zudem danke ich der Graduiertenakademie der TU Dresden ich für die Gewährung eines Abschlussstipendiums sowie die Finanzierung von Konferenzreisen.

Für anregende Diskussionen und gesellige Runden in einsamen Zeiten danke ich herzlich allen Mitstreiter:innen in den Promotionskolloquien der Professur für soziologische Theorien und Kultursoziologie sowie der Professur für Mikrosoziologie und technosoziale Interaktion an der TU Dresden, am Lehrstuhl für Kulturtheorie und Management der Universität Witten/Herdecke sowie an der Professur für Kultur- und Mediensoziologie der Bauhaus-Universität Weimar.

Maximilian Locher und Jonathan Harth bin ich für zahlreiche inspirierende Theoriediskussionen dankbar. Andreas Höntsch danke ich für fortwährenden spannenden Austausch und viele wichtige Anregungen, deren Bedeutung mir häufig erst mit Verspätung aufging. Bei Gereon Rahnfeld möchte ich mich für so konstruktive wie angenehme gemeinsame Auseinandersetzungen mit unseren empirischen Materialien bedanken. Außerdem danke ich Sarah Bernhardt, Jakob Claus, Niklas Egberts, Friedrich Weißbach und allen weiteren geschätzten Alltagsgefährt:innen in der Stabi dafür, dass wir unser Leid und unsere Klagen häufig so heilsam teilen konnten. Karl-Siegbert Rehberg bin ich sehr dankbar dafür, nicht nur meine Faszination für soziologisches Denken überhaupt erst geweckt, sondern durch seine ansteckende Leidenschaft für das Fach darüber hinaus dafür gesorgt zu haben, dass diese seitdem nur noch weiter gewachsen ist.

Celia verdanke ich, dass ich es überhaupt bis zu diesem Punkt geschafft habe, und so viel mehr als nur das.

Rama wird mir in seiner Neugier und warmen Zugewandtheit immer eine Inspiration bleiben. Seinem Andenken ist dieses Buch gewidmet.

*Berlin, im Mai 2026*

*Richard Groß*