Für ein produktives Zweifeln

Eine ethnografische Re-Konfiguration von Risiko-KI-Assemblagen am Beispiel des Robo-Advisory

Martina Klausner und Matthias Kloft

Künstliche Intelligenz im öffentlichen Diskurs und in anthropologischer Forschung

Öffentliche Diskurse um den Einsatz von Künstlicher Intelligenz (KI) bzw. automatisierter Datenverarbeitung und Entscheidungsfindung rangieren wie so oft bei technischen Innovationen zwischen einem Versprechen technischer Lösungen für vielfältige gesellschaftliche Probleme und KI als wirtschaftlichem Erfolgsfaktor einerseits und Warnungen vor problematischen fundamentalen Transformationen, die solche Innovationen mit sich bringen könnten, andererseits. Problematisiert werden Auswirkungen auf einzelne Sektoren, wie beispielsweise durch den Einsatz von Chatbots wie ChatGPT zur Textproduktion im Bildungssektor oder die Bedrohung von Arbeitsplätzen in verschiedenen Bereichen.[1] Aber auch allgemeinere Bedrohungen wie die Manipulation demokratischer Prozesse oder die zunehmende Schwierigkeit zwischen Täuschung und Wahrheit zu unterscheiden, werden angemahnt bis hin zu einer Warnung vor der Entmachtung oder gar Auslöschung der Menschheit durch intelligente Maschinen.[2]

Welche Rolle kann und soll eine anthropologische Forschung in solchen öffentlichen Debatten einnehmen? Vor allem: Welche Art der Forschung braucht es, um jenseits der polemischen Diskussionen differenzierte Impulse und

1 Vgl. OpenAI (2023): »GPTs are GPTs«. https://openai.com/research/gpts-are-gpts (letzter Aufruf: 25.4.2024).

2 Vgl. Egan (2023); PauseAI Global (o.J.): PauseAI. https://pauseai.info/ (letzter Aufruf: 25.4.2024).

Problematisierungen (vgl. Rabinow 2003) in öffentliche Diskurse einzubringen? Neben bislang eher vereinzelten Auseinandersetzungen mit technischen Innovationen und deren gesellschaftlichen Auswirkungen in der Empirischen Kulturwissenschaft (vgl. Klausner/Niewöhner/Seitz 2023; Amelang 2022; Koch 2005; Mathar 2010), gibt es in der angloamerikanischen Kulturanthropologie und dem interdisziplinären Feld der Science and Technology Studies bereits seit mehreren Jahrzehnten Auseinandersetzungen mit diesen Fragen, insbesondere auch im Bereich Künstlicher Intelligenz (vgl. Suchman 1987; Collins 1992; Forsythe 2001, Mackenzie 2017), die wiederum eng verknüpft sind mit verschiedenen Formen kritisch-engagierter oder interventionistischer Forschung.

In ihrer Diskussion feministischer Auseinandersetzungen mit digitalen Technologien bzw. mit den »Wissenschaften des Künstlichen« hebt Lucy Suchman die unterschiedlichen Konfigurationen von Mensch-Objekt-Verhältnissen hervor, die sich durch kritisch-engagierte Forschungen aufzeigen, aber vor allem auch (neu) ziehen lassen (vgl. Suchman 2008). Das Konzept der Konfiguration versteht Suchman (2012) als analytisches Werkzeug, um erstens die Komposition und Eingrenzung dessen, was überhaupt als ein Objekt verstanden wird (von den Entwickler*innen und Anwender*innen aber ebenso von den Forscher*innen), nachvollziehbar zu halten; und zweitens, um die kulturellen Imaginationen und Bedeutungszuschreibungen technischer Innovationen und deren konkrete Materialisierung gleichzeitig in den Blick zu nehmen. Wie Suchman ausführt, thematisieren cyberfeministische Ansätze (vgl. Haraway 1985; Hawthorne/Klein 1999; Hayles 2006) seit den frühen 1980ern die Verschiebungen der Grenzen des Menschlichen und Nicht-Menschlichen durch technowissenschaftliche Entwicklungen und problematisieren dabei insbesondere vergeschlechtlichte sowie andere soziale Einschreibungen in Technologien und deren Reifizierung. Gesucht wurde und wird in dieser kritischen Auseinandersetzung nicht ein distanziertes Außen, sondern der Dialog mit Akteur*innen des Feldes, der bis hin zur Zusammen-/Mitarbeit an experimentellen Alternativen von »sociomaterial assemblages« (Suchman 2008: 150) reichen kann. Mit Verweis auf Donna Haraway versteht Suchman das Anliegen einer gegenwärtigen feministischen Technoscience als »recognition of the material consequences of the figural and the figural grounds of the material« (ebd.: 153). Beide, Haraway und Suchman, positionieren sich dezidiert *in* den Technowissenschaften und setzen sich sowohl ernsthaft mit den Figuren, wie beispielsweise wirkmächtigen Metaphern menschlicher Handlungsfähigkeit oder maschineller Intelligenz,

auseinander – »the figural grounds of the material« (ebd.) – als auch mit den materiellen Konsequenzen solcher Figurationen. Wir greifen in unserem Artikel dieses Anliegen, sich mit den Figurationen *und* den Materialitäten zu beschäftigen, auf und fragen konkret nach den Konsequenzen, die sich durch die gegenwärtige dominierende Erzählung von KI als Risiko ergeben.[3] Welche möglichen Risiken bzw. problematischen Entwicklungen werden KI zugeschrieben und damit sichtbar gemacht, während andere ausgeblendet werden?

Riskante KI ist nicht nur wiederkehrendes Motiv in medialen Auseinandersetzungen mit neuen Entwicklungen, sondern zugleich die zentrale Denkfigur in Regulierungsansätzen, die KI-Systeme einerseits als Innovationsmotor für wirtschaftliche Entwicklungen sehen, aber gleichzeitig deren Risiko für grundlegende demokratische Werte und Grundrechte hervorheben. Ein aktuelles Beispiel, das auch der Aufhänger für unseren Beitrag ist, ist der EU AI Act (AIA), der als eines der ersten grundlegenden KI-Gesetze weltweit derzeit einige Aufmerksamkeit erhält. Vereinfacht gesagt wird das Risiko in Anwendungen von KI vor allem in der Opazität solcher Systeme gesehen, die kaum noch menschlicher Kontrolle unterstehen. Sicherheit und Vertrauen sollen entsprechend durch Maßnahmen, die für mehr Transparenz sorgen, erzeugt werden. Wir nehmen die anstehende Ratifizierung des AIA und die entsprechenden Diskurse als Aufhänger, um das Verhältnis von KI und Risiko zu beleuchten. Um nicht nur Figurationen, sondern auch Materialitäten als wirkmächtig in soziomaterieller Praxis zu diskutieren, konzentrieren wir uns auf eine Fallstudie, die im Banken- und Finanzsektor angesiedelt ist. Interessant an dieser Fallstudie ist insbesondere, dass wir den Risiko-Ansatz im AIA mit anderen Risikovorstellungen aus dem Bereich des Finanzmarkthandels konfrontieren und erweitern können.

3 Unsere Forschung zu Regulierung von KI war Teil einer Projektgruppe des hessischen *Zentrums für verantwortungsbewusste Digitalisierung* (https://zevedi.de/themen/noki/ [letzter Aufruf 25.4.2024]), das maßgeblich aus einer normwissenschaftlichen Perspektive, also insbesondere aus der Perspektive von Recht und Ethik, auf Digitalisierung blickt, und an der Schnittstelle zwischen Regulierung und Politik rechtliche wie ethische Impulse liefert. Unsere Rolle als Anthropolog*innen in der Projektgruppe war bewusst offengehalten und ermöglichte es uns eigene Problematisierungen einzubringen, die über den engeren rechtlichen und ethischen Rahmen hinausreichen und dafür weiterführende gesellschaftliche und epistemische Fragen in die Diskussion einbringen konnten, wie wir in diesem Beitrag weiter ausführen werden.

Da wir sowohl Regulierungsansätze als auch die konkreten soziomateriellen Praktiken rund um KI analytisch in den Blick nehmen, verstehen wir KI in diesem Beitrag als Teil einer Assemblage, eines Gefüges, das menschliche wie nicht-menschliche Akteur*innen umfasst. Ansätze einer »assemblage ethnography«, wie sie sich in den letzten Jahrzehnten in den Anthropologien als konzeptionelle Antwort auf verschiedene Herausforderungen etabliert haben (vgl. Wahlberg 2022; Welz 2021; Farías/Bender 2011; Ong/Collier 2005), betonen Emergenz und Prozessualität als zentral für ein *assemblage thinking*, um Phänomene auf unterschiedlichen Maßstabsebenen in den Blick zu nehmen und zugleich die lokalen Effekte adressieren zu können. Für die Analyse von KI-Systemen ist dieser Zugriff gerade deshalb hilfreich, weil damit die Distribuiertheit und Prozesshaftigkeit von KI hervorgehoben und die Produktlogik, die für den AIA zentral ist, irritiert werden kann. Versteht man KI nicht als Produkt und damit als Entität, die zeitlich und räumlich verortet werden kann, sondern als kontinuierlichen Prozess, der Daten, Software, Infrastrukturen ebenso umfasst wie Regulierung und technowissenschaftliche Expertise, lassen sich unterschiedliche Konfigurationen des Figuralen und Materiellen diskutieren.

Bevor wir auf die Fallstudie eingehen und diese im Hinblick auf ihre unterschiedlichen KI-Risiko-Beziehungen als Teil einer Assemblage diskutieren, stellen wir den rechtlichen Rahmen, insbesondere den AIA, etwas ausführlicher vor. Daran anschließend folgt der konzeptionelle Rahmen, in dem wir den Mehrwert eines *assemblage thinking* für eine ethnografische Analyse von KI-Anwendungen herausarbeiten. Aus dieser Zusammenschau ergibt sich ein ethnografisches Programm für eine anthropologische Auseinandersetzung mit KI, das als Grundlage für Impulse in öffentlichen Diskursen um KI auch jenseits der vorgestellten Fallstudie produktiv gemacht werden kann.

Von *Risky AI* zu *Trustworthy AI*: KI in regulatorischer und anthropologischer Perspektive

Seit Februar 2020 verhandeln die EU-Mitgliedsstaaten über eine neue KI-Verordnung, die, laut EU, »the world's first comprehensive AI law«[4] darstellt. Die-

4 European Parliament (2023): »EU AI Act: First Regulation on Artificial Intelligence«. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (letzter Aufruf: 3.5.2024).

se EU KI-Verordnung oder auch AIA[5] soll dem Umstand Rechnung tragen, dass KI-Systeme einerseits große gesellschaftliche und wirtschaftliche Potentiale bergen, aber gleichzeitig auch Bedenken hervorrufen, wie sich diese neuen Technologien auf Sicherheitsfragen und Grundrechte, wie beispielsweise Diskriminierungsfreiheit, auswirken. KI wird in solchen politischen und regulatorischen Diskursen als disruptive Technologie gerahmt, die Ängste erzeugt und von einer riskanten Technologie zu einer vertrauenswürdigen Technologie entwickelt werden muss (European Commission 2020). KI wird im AIA als Produkt adressiert und sektorübergreifend in den Blick genommen, d.h. nicht als Teil spezifischer Fach- und Rechtsbereiche. Der AIA muss entsprechend mit bereits bestehenden Gesetzen in den spezifischen Sektoren, wie Finanzmarkt oder Medizin, harmonisiert werden.

Die Grundlage des AIA ist ein Risiko-basierter Ansatz, nach dem KI-Systeme in vier Grundkategorien eingeordnet werden: *Unacceptable Risk AI* (unzulässige KI-Systeme, die verboten werden), *High-Risk AI* (KI-Systeme mit hohem Risiko, die insbesondere aufgrund der potentiellen Anwendung in sensiblen Bereichen zahlreichen Zertifizierungs- und Monitoringverfahren unterliegen), *Limited Risk AI* (KI-Systeme mit begrenztem Risiko, für die gewisse minimale Transparenzverpflichtungen gelten) und *Minimal Risk AI* (KI-Systeme mit geringem Risiko, für die die Maßnahmen des AIA nicht angewandt werden müssen und die vermutlich die meisten Anwendungen umfasst). Im Fokus des AIA stehen KI-Systeme mit hohem Risiko und die entsprechenden Ausführungen, wann diese Kategorisierung zutrifft und wie mit diesen Fällen umgegangen werden muss. Die Abstufung in unterschiedliche Risikoklassen mit daran geknüpften Prüf- bzw. Verbotsmechanismen zielt vor allem darauf ab, aus einer *risky AI* eine *trustworthy AI* zu schaffen.

Aus einer regulatorischen Perspektive auf potenziell riskante KI-Systeme wird vor allem gefragt, wer an welcher Stelle wie Verantwortung trägt und im Zweifelsfall für einen Schaden haftbar gemacht werden kann. Als zentrales Problem gilt, dass KI-Systeme potenziell opak sind und verschiedene Maßnahmen entwickelt werden müssen, um diese Opazität handzuhaben. Ein exemplarischer Auszug aus der Entwurfsfassung des AIA:

5 European Commission (2021): »Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS«. In: EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206 (letzter Aufruf: 25.4.2024).

> »To address the opacity that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems. Users should be able to interpret the system output and use it appropriately. High-risk AI systems should therefore be accompanied by relevant documentation and instructions of use and include concise and clear information, including in relation to possible risks to fundamental rights and discrimination, where appropriate« (AIA o.J: 47).

Offenlegungs- und Dokumentationspflichten, Zertifizierungsprozesse, die Registrierung von Anwendungen auf einer zentralen Plattform und dergleichen sollen Transparenz schaffen, dadurch Diskriminierung und andere Risiken verhindern und Vertrauen herstellen. KI wird als Produkt adressiert, dessen Herstellungsprozess und Funktionen zumindest im *high-risk*-Bereich nachvollziehbar sein müssen, um im Zweifelsfall den Punkt, an dem ein potenzielles Risiko entsteht, zu erfassen und den Fehler zu beheben bzw. die Anwendung zu verbieten. Zentrale Begriffe, die im Kontext von KI-Systemen und Risiko-Einstufungen auftauchen, sind Opazität, als eine Form der Undurchdringbarkeit algorithmischer Prozesse, und als Gegenstück Transparenz und entsprechende Maßnahmen. Mit den Erzählungen von Opazität maschinell lernender Algorithmen hat sich Jenna Burrell (2016) ausführlich auseinandergesetzt und drei unterschiedliche Formen von Opazität identifiziert: Erstens Opazität als interessensgeleiteter und absichtsvoller Schutz von proprietären algorithmischen Anwendungen durch Unternehmen und Institutionen; zweitens Opazität als Effekt hochspezialisierter Skills und Expertise, wie dem Schreiben von Code, die Laien in der Regel nicht zur Verfügung stehen; und drittens Opazität als eine Art Mismatch von hochkomplexen mathematischen Optimierungsprozessen einerseits und den Erwartungen an menschliche Kontrolle und Interpretationsmöglichkeiten andererseits. In den verschiedenen Regulierungsansätzen und politischen Debatten wird Opazität allerdings in der Regel nicht differenziert und am ehesten auf die zweite von Burrell identifizierte Form abgezielt: Die fehlende Expertise und entsprechend fehlende Interpretationsmöglichkeit des Outputs von KI-Systemen durch *die Bürger*innen/Verbraucher*innen* lässt KI als Black Box erscheinen, die durch verschiedene Maßnahmen geöffnet bzw. transparent gemacht werden soll. Durch Information, Dokumentation und Zertifizierung von solchen Systemen soll Transparenz in Bezug auf Input und Output und die damit verbundenen Risiken hergestellt werden. In den Science and Technology Studies ist »black

boxing« als Teil von Technologieentwicklung bzw. das Öffnen der Black Box als eine leitende analytische Metapher seit langem etabliert, um unterschiedliche Komponenten und deren komplexe Assoziationen als Teil von Technologie zu adressieren und zu analysieren (vgl. Latour 2002). Mittlerweile ist diese Metapher des Öffnens der Black Box in öffentlichen Diskursen und Policy-Prozessen um KI angekommen und wird dort mehr oder weniger synonym für die Forderungen nach Erklärbarkeit, im Sinne einer »explainable AI«, verwendet (Brożek et al. 2023).

Für unsere Diskussion und Fallstudie sehen wir in diesem Black Box-Transparenz-Nexus allerdings verschiedene Probleme: Erstens reproduzieren die Tropen der Black Box und Opazität, wie wir sie in regulatorischen Aushandlungen, aber auch kritischen Auseinandersetzungen mit KI vorfinden, die dominante Erzählung von KI-Systemen als im Grunde objektive, neutrale, technisch-mathematische Verfahren. Verunreinigungen dieser Objektivität, wie beispielsweise ein Bias, können, wenn man den entsprechenden fehlerhaften Schritt im Prozess bzw. in der Black Box erfasst hat, identifiziert und beseitigt werden. Damit einhergehend und zweitens liegt dieser Black Box-Vorstellung ein technisches Verständnis zugrunde, das gerade für neuere Verfahren generativen Maschinellen Lernens, die auf Deep Learning basieren, wenig Sinn macht. Solche Verfahren werden nicht in einem nachvollziehbaren, sequentiellen Prozess entwickelt, wie es noch bei klassischen, beispielsweise statistischen Algorithmen der Fall ist. Vielmehr handelt es sich um automatisierte, vielschichtige (*deep*) Informationsverarbeitungsprozesse, die aufgrund ihrer Komplexität und Mehrdimensionalität gar nicht transparent gemacht werden können. Der entscheidende Punkt ist, dass es sich gar nicht um eine Black Box im eigentlichen Sinne handelt. Louise Amoore (2020) hat dieses Black Box-Denken in Bezug auf KI grundlegend kritisiert. Dominante kritische Perspektiven auf KI-Systeme, so Amoore, würden einseitig auf mögliche Bias-Probleme und fehlerhafte Berechnungen fokussieren und als Lösung mehr Transparenz und *accountability* fordern (vgl. Amoore 2020: 5). Amoore argumentiert hingegen, dass zum einen diese Kritik die weitreichenden Veränderungen in Bezug auf neue emergente Wahrheitsregimes durch den Einsatz von KI-Systemen vernachlässigt bzw. reproduziert. Zum anderen ignorieren diese Ansätze, dass gerade neue Generationen von generativen KI-Anwendungen nicht mehr als Black Box funktionieren. Sie schlägt daher für Deep Learning-Algorithmen das Bild der russischen Matroschka Puppe vor: Jedes Mal, wenn man eine Schicht Puppe geöffnet hat, findet man darin nur die nächste Schicht Puppe – es gibt keine nachvollziehbaren Schritte und

eindeutigen Komponenten mehr, die man differenzieren und deren Wert-Einschreibungen man nachvollziehen könne, sondern nur »multiple combinatorial possibilities and connections« (ebd.: 13). Amoore schlägt vor, dass wir sehr viel grundlegender den Erzählungen von KI als einer objektiven Mathematik widersprechen und anstelle von Eindeutigkeit und Objektivität die Qualität des Zweifelns, die sie im Kern von Deep Learning-Algorithmen sieht, stark machen müssen. Vereinfacht gesagt, sind gerade Momente des Nicht-Wissens, Zweifelns und immer wieder Anpassens das, was Amoore als zentral für diese Verfahren maschinellen Lernens sieht. Im Umkehrschluss plädiert sie deshalb für nichts weniger als eine zweifelnde Wissenschaft, die sich nicht einfach für eine optimierte, beispielsweise diskriminierungsfreie, KI einsetzt, sondern jeglicher Erzählung von Eindeutigkeit Narrative von Mehrdeutigkeit und Zweifeln entgegensetzt.

Insgesamt diskutiert Amoore Algorithmen bzw. KI-Systeme als spezifische Arrangements von Aussagen (»arrangements of propositions«, Amoore 2020: 12) und grundlegend relational, um die iterativen und experimentellen Kapazitäten von KI hervorzuheben. Diese Rahmung von KI als Arrangements hat, so Amoore, entscheidende Konsequenzen für die Analyse und für ethische Problematisierungen: »The arrangement of propositions means that an apparently optimal output emerges from the differential weighting of alternative pathways through the layers of an algorithm. In this way, the output of the algorithm is never simply either true or false but is more precisely an effect of the partial relations among entities« (ebd.: 13f.). Auch wenn Amoore selbst den Begriff der Assemblage nicht heranzieht, sehen wir in dieser konzeptionellen Rahmung von Algorithmen als Arrangement von Aussagen ein *assemblage thinking* in Bezug auf KI-Systeme aufscheinen, dass wir im Folgenden weiter ausarbeiten. Wie wir in unserer Einleitung hervorgehoben haben, ermöglicht ein relationales, *assemblic* Verständnis von KI-Systemen, die Beziehungen zwischen KI und Risiko zu vervielfältigen und Problematisierungen jenseits der Produkt- und Black Box-Logik anzubieten. Während Amoores Arrangement-Begriff erlaubt in die konkrete soziomaterielle Konfiguration von KI hineinzuzoomen, ermöglicht uns der Assemblage-Ansatz gewissermaßen hinauszuzoomen.

KI-als-Assemblage

In einem Überblick zu den Ansätzen und Entwicklungen einer *assemblage ethnography* hebt Ayo Wahlberg (2022) hervor, dass trotz unterschiedlicher Genealogien *assemblage thinking* vor allem als anti-essentialistische Alternative zu den typischen anthropologischen Analyseeinheiten wie Gesellschaft, Ethnos oder Kultur verstanden wurde und eine Aufmerksamkeit für soziale Phänomene als emergent, verteilt an unterschiedlichen Orten, auf unterschiedlichen Maßstabsebenen und in Praktiken entgegensetzte (vgl. auch Welz 2021; Farías/Bender 2011; Ong/Collier 2005). Wahlberg sieht die verschiedenen Ansätze als methodologische Antwort »to a globalising world where technocratic organisation and technoscientific expertise continue to impact on and shape the lives of people everywhere in the world« (Wahlberg 2022: 139). Anstelle wie in der Anthropologie lange Zeit üblich »cultural formations in the world system« (Marucs 1995 zitiert in Wahlberg 2022: 129) zu untersuchen, lag bereits der Fokus früher *assemblage ethnographies* auf den Konfigurationen, die lokale Lebenswelten und soziale Phänomene auf Makro-Ebene strukturieren. Verstärkt wurden diese Neuausrichtungen seit der Jahrtausendwende durch eine zunehmende anthropologische Hinwendung auf Phänomene und Prozesse von globaler Reichweite, wie Gisela Welz in ihrer Zusammenschau und Diskussion des Assemblage-Begriffs aufzeigt (vgl. Welz 2021). Statt empirisch abgrenzbare Einheiten rücken Phänomene in den anthropologischen Blick, die als Gefüge menschliche wie nicht-menschliche Akteur*innen umfassen und in einem konstanten Werden und in vielfältigen Relationen unterschiedlicher Reichweiten zueinander verstanden werden.

In ähnlicher Weise kommen KI-Systeme, wenn wir sie als Assemblage denken, grundlegend anders in den Blick als in der Produkt-Logik des AIA oder als Objekt kritischer Interventionen und öffentlicher Diskurse. Wir plädieren dafür, KI-Systeme, generatives Maschinelles Lernen oder Algorithmen als Assemblage in den Blick zu nehmen, die aus vielfältigen Gefügen von Gefügen und partiellen Verbindungen zwischen unterschiedlichen Entitäten, gewissermaßen Arrangements von Arrangements, besteht, die oftmals asynchron und nicht einfach lokalisierbar sind. Hinzukommt, dass frühe *assemblage thinking*-Ansätze in einer *Anthropology of Policy* ebenfalls unterschiedliche Maßstabsebenen von Politik und Regulierung betont haben und damit Regulierung selbst als Teil von Praxisgefügen betrachtet werden kann – und nicht unabhängig des zu regulierenden Objekts. Für unsere Untersuchung hat diese hier angebotene Lesart von KI-Systemen als Assemblage mehrfache Konsequenzen:

Erstens geht es in dieser Lesart nicht mehr um eine klassische Mensch-Technik-Interaktion oder Beziehung (der Mensch entwickelt bzw. trainiert das Produkt und kann daher als verantwortlich, haftbar gedacht werden; oder der Mensch ist von den automatisierten Entscheidungen, die von diesem Produkt ausgehen, unmittelbar betroffen), sondern um eine Vervielfältigung relevanter Akteur*innen. Software, Daten, Infrastrukturen ebenso wie Entwickler*innen, Regulierer*innen und Anwender*innen sind Teil des jeweiligen Praxisgefüges. Zudem sind diese, das wäre der zweite zentrale Unterschied, nicht mehr einfach lokalisierbar, sondern können als auf unterschiedlichen Maßstabsebenen und zeitlich verteilt gedacht werden. Gerade da sich KI-Systeme laufend weiterentwickeln und als grundlegende mathematische Verfahren in immer neuen Kontexten angewandt werden, ist der Prozess zeitlich nie abgeschlossen und bleibt emergent. Drittens lässt sich Regulierung damit nicht außerhalb von diesen Praxisgefügen denken, sondern als produktiver und konfigurierender Bestandteil von *KI-als-Assemblage*. Und viertens argumentieren wir mit Verweis auf Amoore, dass mit generativen KI-Systemen und deren immer weiteren Verbreitung neue Formen der Wissensproduktion und damit neue Regimes der Verifizierung und Wahrheitsproduktion entstehen. Diese weitreichende Transformation bleibt in regulatorischen Ansätzen ausgeblendet, fordert aber gerade eine kritische Analyse und Problematisierung (vgl. Rabinow 2003).

Fallstudie: Robo-Advisory

Übertragen auf unser empirisches Beispiel, den Einsatz von KI im Finanzmarkt, zweifeln wir das spezifische Verhältnis von KI und Risiko an, wie es in politischen Diskursen und Gesetzgebungsansätzen wie der KI-Verordnung vertreten wird. Damit meinen wir nicht, dass von KI-Systemen keine Risiken ausgehen. Vielmehr zielen wir darauf ab, das Verhältnis von Risiko und KI zu vervielfältigen und nach den konkreten und vielschichtigen Risikoverhältnissen auf unterschiedlichen Maßstabsebenen zu fragen und danach, welche Rolle KI dabei spielen kann. Das Beispiel, das wir für unsere Fallstudie gewählt haben, ist auf den ersten Blick relativ simpel. Wir haben uns verschiedene Beispiele des Robo-Advisory angeschaut. Zwei dieser Systeme werden von spezialisierten Investment- und FinTech-Unternehmen angeboten und ein weiteres von einer großen deutschen Bank. Vereinfacht gesagt geht es beim Robo-Advisory um eine Automatisierung der klassischen Vermögensbera-

tung. Anstelle einer Bankmitarbeiterin, die eine Kundin in der Anlage von Wertpapieren berät, kommen verschiedene teil- bis vollautomatisierte Prozesse zum Einsatz. Während in gängigen Analysen Robo-Advisory als Produkt gefasst wird, in dem vor allem das Verhältnis zwischen KI und Anleger*in problematisiert wird, schlagen wir vor, im ersten Schritt in die unterschiedlichen Arrangements von KI und Risiko im Rahmen des Robo-Advisory reinzuzoomen, um dann im nächsten Schritt diese Arrangements als Teil weiter gefasster Assemblagen zu diskutieren. Darauf aufbauend lassen sich alternative Problematisierungen anbieten, als sie in gängigen dominanten Diskursen und Regulierungsansätzen zu finden sind.

Wir haben den Robo-Advisor-Prozess in zwei analytische Kontexte unterteilt, in welchen jeweils unterscheidbare Konfigurationen von KI sichtbar werden (vgl. Suchman 2012). Beim *(1) Risk-Profiling* steht in erster Linie die Beziehung zwischen Anleger*in und Robo-Advisor im Mittelpunkt. Dabei legen wir den Fokus auf die automatische Einordnung von Anleger*innen in Risikoprofile durch ein dafür vorgesehenes Webinterface. Die hier verwendeten Technologien basieren auf vordefinierten Logikregeln und Algorithmen. Der Kontext des *(2) Rebalancing* konzentriert sich auf die automatisierten Markttransaktionen des Robo-Advisor. Im Kern handelt es sich dabei um eine Form des algorithmischen Handels, welcher vollautomatisiert Aktien und andere Finanzprodukte kauft und verkauft. Uns interessiert dabei, wie der Robo-Advisor Entscheidungen trifft und auf welchen Modellen des Finanzmarktes diese Entscheidungen basieren.

KI-Systeme und andere Formen der statistischen Modellierung gelten speziell im Finanzmarkt als etablierte Methoden und eben nicht, wie in öffentlichen Diskursen und regulatorischen Ansätzen konstatiert, als potenziell disruptiv und riskant. Handelsstrategien unter Verwendung von generativen Algorithmen und *deep neural networks* gehen einen Schritt weiter und berücksichtigen nicht nur andere Informationsströme, sondern basieren auch auf komplexen, mehrdimensionalen Informationsverarbeitungsprozessen. Tatsächlich werden dabei Handelsstrategien weitgehend ohne jene menschliche Vorannahmen wie ökonomische Theorien und parametrische Modelle erzeugt. Im Gegensatz zu konventionellen Methoden des Risiko-Managements kann keine kausale oder sequenzielle Beziehung zwischen einem möglichen Risiko und dessen Resultat hergestellt werden. *Deep neural networks*, in Adrian Mackenzies Worten, funktionieren durch Prozesse des »transforming, constructing or imposing some kind of shape on the data« (Mackenzie 2015: 432). Diese Formierung der Daten erlaubt es, Funktionen zu

finden und dadurch bestimmte Muster im Datensatz vorherzusagen. Bevor ein zufriedenstellender Output erzeugt wird, bedarf es eines kontinuierlichen Prozesses des Testens, der Anpassung, und der Validierung durch Softwareingenieur*innen, Data-Scientists und Trader. Es handelt sich hierbei eben nicht um eine Black Box, die durch Transparenz ihre Risikobeziehungen offenlegt. Stattdessen lassen sich Risikobeziehungen im Falle des Robo-Advisors nur in den unterschiedlichen Kontexten situieren, welche mit unterschiedlichen Logiken, Berechnungsmethoden und Graden menschlicher Kontrolle einhergehen, wie wir anhand der analytischen Kontexte des *Risk-Profiling* und des *Rebalancing* verdeutlichen werden.

Risk-Profiling

Dieser erste analytische Kontext befasst sich mit der Risikobeziehung zwischen Kund*innen und dem Robo-Advisor. Zentral in diesem Zusammenhang ist ein Multiple-Choice-Fragenkatalog, der über eine Website ausgefüllt werden kann. Dieser Prozess wird in der Branche *Onboarding* oder *Exploration* genannt. Der Fragenkatalog umfasst allgemeine Fragen zur Person, dem Einkommen sowie zu persönlichen Anlagezielen, aber auch Fragen bezüglich der Erfahrungen mit Geldanlagen oder dem persönlichen Verhältnis zu finanziellen Risiken tauchen darin auf. Die Antworten werden automatisiert ausgewertet und darauf basierend ein Risikoprofil erstellt. Aus der Perspektive von Anleger*innen werden finanzielle Risiken auf diese Weise individualisiert und objektiviert. Durch die Zuteilung eines Risikoprofils, das auf ökonomischen Theorien und Mathematik basiert, soll das Kapital im Rahmen der individuellen Risikoaffinität und finanziellen Möglichkeiten vor den Risiken des Finanzmarktes geschützt bzw. diesen ausgesetzt werden. Ein Industriestandard für solche Risikoprofile heißt *Value-at-Risk* (VaR). Es ist ein statistisches Maß für potenzielle finanzielle Verluste innerhalb eines Portfolios. Auf Basis des Fragenkatalogs gibt der Robo-Advisor ein Risikoprofil mit quantifiziertem VaR-Wert aus und zeigt eine Simulation, wie sich ein Portfolio in der Zukunft entwickeln kann. Dadurch wird Risiko stabilisiert und für Anleger*innen überschaubar und kalkulierbar. Wir argumentieren, dass Risiko in diesem Zusammenhang vor allem in der Beziehung zwischen riskanten Märkten und dem gefährdeten Kapital situiert ist. KI fungiert in diesem Zusammenhang nicht selbst als potenziell riskante Technologie, sondern als Managerin einer unsicheren Zukunft. Wir verstehen diesen Prozess der Risikoprofilierung als algorithmische Antizipationspraktik (vgl. Adams

et al. 2009), bei der Risiko mit der Logik des konventionellen Risikomanagements antizipiert wird. Im Vordergrund steht die Vermeidung finanzieller Verluste durch die VaR-Metrik.

Allerdings sehen wir in diesem Prozess einer Kategorisierung und Profilierung von Anleger*innen noch ein zweites Risiko-Verhältnis aufscheinen: Vor dem Hintergrund, dass solche Formen des Portfoliomanagement zunehmend auch als private Rentenvorsorge beworben und genutzt werden, ließe sich diese Profilbildung in Verbindung mit Anlagestrategien durchaus als relevant für Einschränkungen von Grundrechten verstehen. Laut *Statista* beträgt das durch Robo-Advisor verwaltete Anlagevolumen in Deutschland im Jahr 2023 bereits über 120 Milliarden Euro mit einer steigenden Tendenz. Bis 2027 sieht das Marktforschungsunternehmen die Zahl der Nutzer*innen sogar auf über sieben Millionen steigen.[6] Wer erhält basierend auf welchen Daten Zugang zu welchen Finanzmarktprodukten? Tatsächlich erhält diese Art der automatisierten Profilierung für private Geldanlagen in der KI-Verordnung selbst und den entsprechenden Debatten keine Aufmerksamkeit. Während Diskussionen mit Vertreter*innen aus Recht, Ethik und Industrie zur Frage der Regulierung von Robo-Advisory wurde zwar immer wieder auf den Kontext des *Risk-Profiling* Bezug genommen, es ging aber vielmehr darum, ob ein Multiple-Choice Fragebogen die gleichen rechtlichen Anforderungen erfüllt wie ein konventionelles persönliches Beratungsgespräch. Die Regulierung sieht hier in erster Linie ein Risiko für den Verbraucherschutz durch die Anwendung von KI. Geht es stattdessen um den Zugang zu Krediten, werden die Risiken für Grundrechte und der Schutz vor Diskriminierung ernst genommen. Deutlich wird dies vor allem angesichts der Debatten um den Einsatz von KI im Bereich des *Creditscoring*.[7] Diese Methoden der automatisierten Kategorisierung von Kreditwürdigkeit werden in der KI-Verordnung explizit als Hochrisiko-Anwendungen verstanden und entsprechend reguliert. Wir argumentieren, dass der Robo-Advisor und das Risiko-Verhältnis zwischen Anleger*innen

6 Statista GmbH (o.J.): »Robo-Advisors – Deutschland«, in: Statista. https://de.statista.com/outlook/dmo/fintech/digital-investment/robo-advisors/deutschland (letzter Aufruf: 25.4.2024).

7 Creditscoring ist eine etablierte Praxis bei der privaten Kreditvergabe durch Banken. In Deutschland speziell werden Debatten um diese Praxis immer wieder im Kontext der SCHUFA geführt, welche personengebundene finanzielle Daten zentral sammelt und das sogenannte SCHUFA-Scoring als Dienstleistung zu Verfügung stellt.

und KI aus dem sozioökonomischen Kontext einer zunehmend neoliberalen Rentenpolitik herausgelöst sind.

Rebalancing

Als zweiten analytischen Kontext des Robo-Advice-Prozesses betrachten wir die KI-basierte Interaktion mit den Finanzmärkten. Der Grad der Automatisierung sowie der hier angewandten Methoden und KI-Verfahren variiert zwischen verschiedenen Anbietern. Wichtig ist, dass beim Kontext des *Rebalancing*, den wir hier beschreiben, mehrere automatisierte Prozesse zusammenspielen. Im ersten Schritt erstellt der Robo-Advisor ein Risiko-Rendite-Profil für jeden Vermögenswert, der Teil der Handelsstrategie ist. Dabei handelt es sich unter anderem um ETFs und Anleihen, aber auch Aktien und andere Wertpapiere können zuvor vom Menschen vorselektiert werden. Die Risiko-Rendite-Profile basieren auf einer Reihe fest programmierter Regeln sowie auf der Analyse historischer Marktdaten. Im zweiten Schritt werden die vorselektierten Wertpapiere den oben beschriebenen Risikoprofilen der Anleger*innen zugeordnet. Diese Zuordnung wird in der Industrie als *Matching* bezeichnet und dabei wird ein Aktienportfolio erzeugt, dessen Zusammensetzung genau dem VaR-Wert des Risikoprofils entspricht. Der Robo-Advisor führt Markttransaktionen wie den Kauf und Verkauf von Wertpapieren autonom durch. Diese automatischen Transaktionen bezeichnet man als *Rebalancing*, da jegliche Veränderungen an den Märkten das Gleichgewicht von Risiko und Rendite innerhalb der Aktienportfolios verschiebt. In der Praxis hat das zur Folge, dass während volatiler Marktphasen der Anteil weniger riskanter Anlagen automatisch erhöht wird und vice versa.

Regulierungsinstitutionen wie die *Bundesanstalt für Finanzdienstleistungsaufsicht* (BaFin) beschäftigen sich mit den finanzpolitischen und systemischen Risiken, die vom Robo-Advisor ausgehen. Aus der Perspektive der Finanzmarktregulierung werden zurzeit nur Formen des algorithmischen Handels wie zum Beispiel der Hochfrequenzhandel expliziert reguliert. Diese spezielle Form der Transaktion unterliegt in Deutschland und der EU besonderen Richtlinien auf Grund seiner potenziell systemischen Risiken für die Märkte. Im Rahmen von Expert*inneninterviews mit Vertreter*innen aus der Industrie wurde die Frage diskutiert, ab wann Robo-Advisory ebenfalls eine Form des algorithmischen Handelns darstellt, und somit speziellen Regularien unterliegen würde. Das Risiko-Verhältnis, das hier diskutiert wird, bezieht

sich auf die einzelnen Marktransaktionen, welche der Robo-Advisor autonom durchführt und nur einen Teil des *Rebalancing* ausmachen.

Wir sehen im Kontext des *Rebalancing* wiederum weitere Risiko-Verhältnisse, die nicht in den einzelnen automatisierten Marktransaktionen situiert sind, sondern in der besonderen Rolle von KI beim Simulieren und Prognostizieren des Marktes. Dabei kommen Simulationen zum Einsatz, um die Performance eines Portfolios über eine bestimmte Dauer zu prognostizieren. Diese Simulationen basieren auf historischen Marktdaten und erzeugen die Signale, die der Robo-Advisor benötigt, um Markttransaktionen auszuführen. Da es zu komplex wäre, jedes einzelne Wertpapier zu modellieren, werden Portfolios immer als Einheit simuliert. Risiko wird so quantifizierbar, bleibt aber hochgradig relational und kann nur innerhalb eines gesamten Portfolios aus unterschiedlichen Wertpapieren hinreichend prognostiziert werden. Wir argumentieren hier mit Mackenzie, dass diese Simulationen keine singulären und eindeutigen Prognosen und entsprechende Risikoverhältnisse mehr produzieren, sondern eine ganze Reihe mehrdeutiger und spekulativer Prognosen (vgl. Mackenzie 2013). Erst in einem experimentellen Prozess der Evaluierung und In-Beziehung-Setzen dieser »contingencies and open-ended solutions« (Amoore 2020: 12) kann der Robo-Advisor Signale erzeugen, welche nicht bloße Lösungen mathematischer Probleme darstellen, sondern »actionable propositions« (ebd.) für mögliche Transaktionen, die das Portfolio optimieren. Wir sehen hier mit Mackenzie die Entstehung neuer Vorhersagemodi, also Praktiken der Voraussage innerhalb von Antizipationsregimen, die nicht mehr darauf ausgelegt sind, Zukunftsszenarien einzugrenzen und deren Eintritt statistisch vorherzusagen, sondern zukunftsorientierte Interventionen in der Gegenwart zu optimieren. Mackenzie (2013) bezeichnet diese Vorhersagemodi als *predictivity*, um den prozesshaften Charakter gegenüber *prediction* hervorzuheben. *Predictivity* nimmt die verschiedenen Datenflüsse, Infrastrukturen, Codierungs- und Finanzmarktpraktiken von KI-Konfigurationen ernst und produziert damit bestimmte Risikoverhältnisse zwischen Wertpapieren, Portfolios und dem Markt. Im Vordergrund stehen dabei weniger die individuellen Risiken von Anleger*innen durch neue Anlageprodukte, sondern die systemischen und gesellschaftlichen Risiken neuer Wahrheitsregime und Vorhersagemodi.

Für eine zweifelnde Wissenschaft

Wie wir anhand der Diskussion unserer Fallstudie zum Robo-Advisory herausgearbeitet haben, lässt sich mit dem Ansatz einer *assemblage ethnography* die Rolle von KI in den verschiedenen Bereichen des Robo-Advisory und als Teil unterschiedlicher Risikoverhältnisse relational und als grundlegend emergent denken. Anders als mit der Einteilung von KI als Produkt in eindeutige Risiko-Kategorien, wie sie der AIA vorschlägt, wird dabei zum einen KI nicht einfach nur als potenziell riskante Technologie sichtbar, sondern wird in ihrer Rolle als Risiko-Managerin im Finanzmarktgeschehen ernst genommen. Die Automatisierung des Finanzmarktgeschehens ist mittlerweile so etabliert, dass eine Rahmung von KI als disruptiv und riskant, wie es in öffentlichen Diskursen problematisiert wird, nur bedingt Mehrwert generiert. Unser Vorschlag KI als Assemblage zu denken, ermöglicht es hingegen in die spezifischen Kontexte hineinzuzoomen und dadurch eine differenzierte Analyse und Multiplizierung von KI-Risiko-Verhältnissen herauszuarbeiten. Gleichzeitig lässt sich mit diesem relationalen Ansatz der Blick auf KI um weitere Formen der Problematisierung ergänzen, wodurch andere, gesellschaftlich relevante Risikoverhältnisse sichtbar werden. So lassen sich, wie wir mit dem Beispiel des Risikoprofils angemerkt haben, die Auswirkungen von automatisierter Anlageberatung durchaus als grundrechtsrelevant diskutieren. Wer erhält Zugang zu welchen Finanzprodukten und kann wie in welcher Höhe für die Rente vorsorgen? Am Beispiel der Simulationen und Prediktionsmechanismen wiederum lässt sich diskutieren, wie mit dem Einsatz von KI im Finanzmarkt und im Handel neue Formen der Antizipation von Zukünften zum Tragen kommen. *Predictivity* zielt auf eine andere Form der kontinuierlichen, datengetriebenen, simulierenden Optimierung von Zukünften in der Gegenwart ab und verändert damit auch das Risiko-Verhältnis in Finanzmarkttranskationen maßgeblich. Eine zweifelnde Wissenschaft, wie sie Amoore vorgeschlagen hat, muss dabei die epistemischen und technischen Grundlagen der vielfältigen Anwendungen von KI in spezifischen Kontexten ernst nehmen, um daran anknüpfend kritische Anmerkungen liefern zu können.

Literatur- und Quellenverzeichnis

Adams, Vincanne/Murphy, Michelle/Clarke, Adele E. (2009): »Anticipation: Technoscience, life, affect, temporality«, in: Subjectivity 28 (1), S. 246–265.

Amelang, Katrin (2022): »(Not) Safe to Use: Insecurities in Everyday Data Practices with Period-Tracking Apps«, in Andreas Hepp/Juliane Jarke/Leif Kramp (Hg.), New Perspectives in Critical Data Studies. The Ambivalences of Data Power (= Transforming Communications – Studies in Cross-Media Research), Cham: Palgrave Macmillan, S. 297–321.

Amoore, Louise (2020): Cloud Ethics: Algorithms and the Attributes of Ourselves and Others, Durham: Duke University Press.

Brożek, Bartosz/Furman, Michał/Jakubiec, Marek/Kucharzyk, Bartłomiej (2023): »The Black Box Problem Revisited. Real and Imaginary Challenges for Automated Legal Decision Making«, in: Artificial Intelligence and Law 32, S. 427-440.

Burrell, Jenna (2016): »How the Machine ›Thinks‹: Understanding Opacity in Machine Learning Algorithms«, in: Big Data & Society 3 (1).

Collins, Harry M. (1992): Artificial Experts: Social Knowledge and Intelligent Machines, Cambridge: The MIT Press.

Egan, Matt (2023): »Exclusive: 42 % of CEOs Say AI Could Destroy Humanity in Five to Ten Years«, in: CNN, https://edition.cnn.com/2023/06/14/business/artificial-intelligence-ceos-warning/index.html (letzter Aufruf: 25.4.2024).

European Commission (2020): Shaping Europe's Digital Future. Publications Office of the European Union, https://eufordigital.eu/library/shaping-europes-digital-future/ (letzter Aufruf: 25.4.2024).

Farías, Ignacio/Bender, Thomas (Hg.) (2011): Urban Assemblages: How Actor-Network Theory Changes Urban Studies, London: Routledge.

Forsythe, Diana E. (2001): Studying Those Who Study Us: An Anthropologist in the World of Artificial Intelligence, Stanford: Stanford University Press.

Haraway, Donna (1985): »A Cyborg Manifesto: Science, Technology, and Socialist Feminism in the 1980s«, in: Socialist Review 80, S. 65–108.

Hawthorne, Susan/Klein, Renate (1999): Cyberfeminism: Connectivity, Critique and Creativity, North Melbourne: Spinifex Press.

Hayles, N. Katherine (2006): »Unfinished Work: From Cyborg to Cognisphere«, in: Theory, Culture & Society 23 (7–8), S. 159–166.

Klausner, Martina/Niewöhner, Jörg/Seitz, Tim (2023): » Curating the Widerstandsaviso: Three Cases of Ethnographic Intravention in R&D Consortia«, in: Science as Culture 32, S. 1–24.

Koch, Gertraud (2005): Zur Kulturalität der Technikgenese: Praxen, Policies und Wissenskulturen der künstlichen Intelligenz, St. Ingbert: Röhrig Universitätsverlag.

Latour, Bruno (2002): Die Hoffnung der Pandora: Untersuchungen zur Wirklichkeit der Wissenschaft, Frankfurt a.M.: Suhrkamp.

Mackenzie, Adrian (2013): »Programming Subjects in the Regime of Anticipation: Software Studies and Subjectivity«, in: Subjectivity 6 (4), S. 391–405.

Mackenzie, Adrian (2015): »The Production of Prediction: What does Machine Learning Want?«, in: European Journal of Cultural Studies 18 (4–5), S. 429–445.

Mackenzie, Adrian (2017): Machine Learners: Archaeology of a Data Practice, Cambridge: The MIT Press.

Marcus, George. E. (1995): »Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography«, in: Annual Review of Anthropology 24 (1), S. 95–117.

Mathar, Thomas (2010): Der digitale Patient: Zu den Konsequenzen eines technowisenschaftlichen Gesundheitssystems, Bielefeld: transcript.

Ong, Aihwa/Collier, Stephen (2005): Global Assemblages: Technology, Politics, and Ethics as Anthropological Problems, Oxford: Blackwell.

Rabinow, Paul (2003): Anthropos Today: Reflections on Modern Equipment, Princeton: Princeton University Press.

Suchman, Lucy (1987): Plans and Situated Actions: The Problem of Human-Machine Communication, Cambridge: Cambridge University Press.

Suchman, Lucy (2008): »Feminist STS and the Sciences of the Artificial«, in: Edward J. Hackett/Olga Amsterdamska/Michael E. Lynch/Judy Wajcman (Hg.), The Handbook of Science and Technology Studies, 3te Ausgabe, Cambridge: The MIT Press, S. 139–164.

Suchman, Lucy (2012): »Configuration«, in: Celia Lury/Nina Wakeford (Hg.), Inventive Methods: The Happening of the Social, London: Routledge, S. 48–60.

Wahlberg, Ayo (2022): »Assemblage Ethnography: Configurations Across Scales, Sites, and Practices: Post-Structuralism«, in: Maja H. Bruun/Ayo Wahlberg/Rachel Douglas-Jones/Cathrine Hasse/Klaus Hoeyer/Dorthe B. Kristensen/Brit R. Winthereik (Hg.), The Palgrave Handbook of the Anthropology of Technology, Singapore: Palgrave Macmillan, S. 125–144.

Welz, Gisela (2021): »More-Than-Human Futures: Towards a Relational Anthropology in/of the Anthropocene«, in: Hamburger Journal für Kulturanthropologie (HJK) 13, S. 36–46.