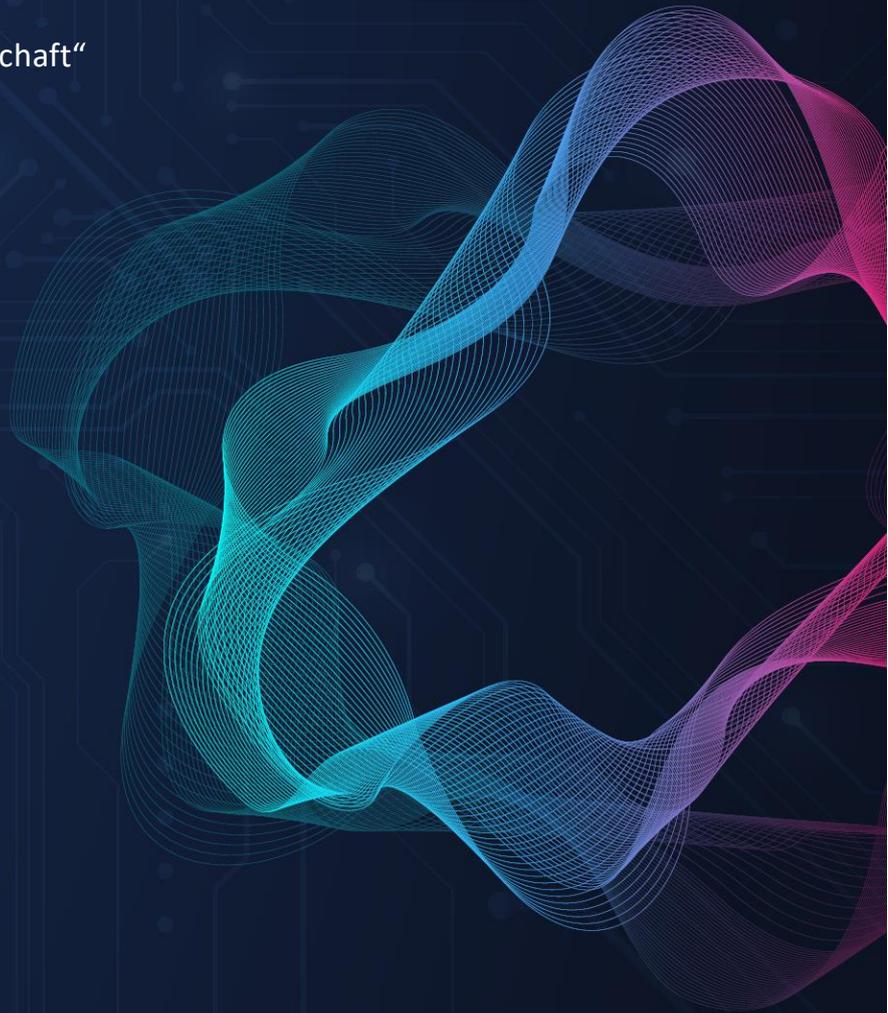


Peter Kahlert, Maryam Tatari, Suzette Kahlert, Prof. Dr. Jan-Hendrik Passoth
Europa-Universität Viadrina

Mehr Methoden mixen, mehr Experimente wagen: Mit sozialwissenschaftlicher Methodologie den Herausforderungen der KI-Forschung begegnen

Bericht aus dem Forschungsprojekt
„Künstliche Intelligenz, Mensch und Gesellschaft“
Juli 2022



GEFÖRDERT VOM

ZUSAMMENFASSUNG

Anwendungen und Forschung, die allgemein unter dem Stichwort „Künstliche Intelligenz“ (KI) behandelt werden, stellen ein weites, vielfältiges Feld dar, dessen einzelne Fälle typischerweise sozio-technische Komplexe, verwachsene Praktiken, Infrastrukturen, Technologien, Artefakte und Handlungen sind. Unser Artikel gibt einen kleinen Überblick zum Einstieg in dieses Feld und diskutiert die Stärke einer multiparadigmatischen und vielfach ausdifferenzierten, sozialwissenschaftlichen Methodologie darin – und inwiefern gerade diese Eignung auch zur stärkeren Inkorporierung KI-gestützter Verfahren in den Sozialwissenschaften auffordert.

KEYWORDS

KI-Anwendungen, KI-Fälle, KI-Felder, KI-Forschung, sozialwissenschaftliche Methoden

INHALT

Zusammenfassung.....	2
Keywords.....	2
Inhalt.....	3
1. Das Bezugsproblem.....	4
2. Kartierung: Potentiale und Probleme der KI/-Forschung.....	4
3. Werkzeuge und Methoden der KI-Forschung.....	9
4. Fazit.....	10
Referenzen	12

1. DAS BEZUGSPROBLEM

Obschon Technologien, die allgemein unter dem Stichwort Künstliche Intelligenz (KI) behandelt werden, mittlerweile eine alltägliche Realität und bisweilen auch Normalität geworden sind, stellen sie die Forschung vor große Herausforderungen. Dabei birgt KI nicht nur selbst große Potenziale für Wissenschaft und Forschung – auch ihre eigene, technische Entwicklung betreffend –; sie eröffnet vielmehr ein eigenes Forschungsfeld für Journalismus, Ingenieurs-, Technik-, Computer- und verschiedenste, vermutlich gar alle, Humanwissenschaften. Dabei handelt es sich jedoch keineswegs um Spezialdiskurse oder wohlsortierte Nischen, vielmehr durchzieht „Künstliche Intelligenz“ praktisch ausnahmslos alle gesellschaftlichen, wirtschaftlichen, technischen und alltäglichen Bereiche zumindest so weit, dass ihre Abwesenheit sich selbst schon als forschungswürdiger Sonderfall qualifiziert. Indes ist KI keineswegs ein geschlossenes, feststehend monolithisches Konzept, sondern ein Sammelbecken für eine Vielzahl von Technologien, Prinzipien, und Vorstellungen, die mehr in politischen oder programmatischen Texten als in konkreten Forschungsvorhaben beheimatet sind.

Im Folgenden entfalten wir die Brisanz und Dringlichkeit des Themas bzw. der (humanwissenschaftlichen) Erforschung von KI, die sich selbstverständlich bereits in der Gegebenheit des Forschungsverbunds „KI, Mensch und Gesellschaft“ (KIMeGe) widerspiegelt. Dieses Experten-Konsortium repräsentiert selbst schon einen kleinen und nichtsdestoweniger profunden Auszug aus der weiten Welt reflexiver, d.h. nicht anwendungsbezogener KI-Forschung. Vor dem Hintergrund der mannigfaltigen Herausforderungen solcher Forschungen ist es uns insbesondere ein Anliegen hervorzuheben, dass besagter Vielfältigkeit nur mit einem entsprechend heterogenen Methodenarsenal begegnet werden kann. Oberstes Anliegen unserer eigenen Expertise zu *methodischen Herausforderungen und Chancen* der sozialwissenschaftlichen KI-Forschung ist es, für diesen Anspruch, den der Gegenstand nicht nur an humanwissenschaftliche Forschung stellt, zu sensibilisieren, und mit unserem Beitrag Mittel an die Hand zu geben, das eigene methodische Verständnis durch eine Auswahl spezifischer Werkzeuge und Nutzungshinweise zu schärfen.

2. KARTIERUNG: POTENTIALE UND PROBLEME DER KI/-FORSCHUNG

Der Forschungsbereich rund um KI – Diskurse wie Technologien betreffend – ist groß. Doch lassen wir uns nicht täuschen von den vielfältig konjugierbaren Fragen, mit denen man KI konfrontieren kann, oder ihren verschiedenen Anwendungsfällen, mit denen sie uns konfrontiert. Vieles an KI ist eher Zuschreibung und Vorstellung, denn materiale Realität, und die Diskrepanz zwischen wissenschaftlichen und programmatischen Texten zu KI fällt mitunter überraschend stark aus, oder aber sie verschwimmen in Genres programmatischer Forschung und Forschungsprogrammatisierung. Während KI Assoziationen mit neuester Technologie und dynamisch interaktiver Software anregt, kann es sich bei genauem Hinsehen auch schon mal um eher altbekannte Technik handeln. Gleichzeitig sind auch neuere Methoden des ‚maschinellen Lernens‘ Limitationen unterworfen, die für den nächsten geschäftlichen Pitch oder das kommende Policy Briefing gerne mal vernachlässigt werden. Da die kritische Auseinandersetzung mit maschinellem Lernen selbst Teil des Projekts KIMeGe ist, soll auch hier der Hinweis auf die Problematik anthropo- und andromorpher Vorstellungen vom Lernen der Maschine nicht ausgelassen werden (vgl. Suchman 1996).

Nichtsdestoweniger sind KI Anwendungen des maschinellen Lernens keine bloße Wiederholung klassischer Statistik, sondern haben ihre eigenen methodologischen Inhalte, Herausforderungen, und Potenziale. Daher ist es ein Kernanliegen dieses Artikels, auf die Besonderheiten, und darin insbesondere die Vielfältigkeiten der KI-Forschung aus Sicht einer sozialwissenschaftlichen Methodologie und Methodenschule hinzuweisen,

und ein Plädoyer dafür zu halten, wie mehr mixed methods und experimentelle Ansätze helfen können, mit diesem speziellen Forschungsfeld umzugehen.

Als Produzentin gesellschaftlicher (Selbst-)Beschreibungen sind die Sozialwissenschaften darauf angewiesen, den jeweiligen Zuschnitt und Kontext von KI als Begriff, Konzept, Vision, und materielle, technische Umsetzung zu beachten und zu markieren – anderenfalls machte sie sich unweigerlich einer bestimmten Diskurs- oder Technologiepraxis untertan. Will sie Vorstellungen, Erwartungen, Handlungen und Artefakte verstehend miteinander vermitteln, was besonders bedeutet, den (zuvor) vorliegenden Handlungshintergrund zu berücksichtigen, muss sie bereit sein, ihre Instrumentarien an den jeweiligen Fall anzupassen.

Es überrascht also nicht, wenn die Definitionen für KI eher zahlreich ausfallen (vgl. u.a. Turing 1950, John McCarthy 2007 oder Russel/Norvig 2009). Mit nur ein wenig begrifflicher Unschärfe ließe sich aber behaupten, dass sie alle auf die eine oder andere Art menschliche Tätigkeit durch maschinelle Prozesse supplementieren oder imitieren, entweder *um* oder *indem* sie große Datensätze und umfangreiche Modellräume auf die eine oder andere Art miteinander verarbeiten (vgl. Bowker/Bechmann 2019 oder Searle 1980).

Trotz der gegebenen konzeptionellen Verschiedenheit macht die konkrete technische Umsetzung nicht notwendig einen großen Unterschied. Ein Schachcomputer, der tatsächlich durch unzählige Spiele (womöglich mit sich selbst) zu siegen gelernt hat, ist lediglich methodisch unterschieden von einer Maschine, die schlicht alle erlaubten Züge und Szenarien durchprobiert (man spricht passend und stimmungsvoll von „Brute Force“, also roher Gewalt), um sein Ziel zu erreichen. Je nach Blickwinkel ergeben sich Unterschiede oder nicht – für den Verlauf eines Schachspiels mag hier kein Unterschied beobachtet werden, für die Produktionsbedingungen der Technologie, ihrer Nachnutzung etc. hingegen schon. Natürlich würde es auch gelten, einer Maschine gezielt bestehende Heuristiken, etwa etablierte Schachmanöver einzuprogrammieren. Im engeren Sinne sind darin Programme und ihre Konzepte enthalten, die auf sogenanntem „maschinellen Lernen“, ob supervised oder unsupervised (vgl. Alpaydin 2016), d.h. ob mit bereits gelabelten Daten oder eben zum Zwecke einer solchen Datenbeschriftung, oder den berühmten „neuronalen Netzen“ (vgl. Bechmann 2017) und ihren vielfältigen Spielarten und Verwandten, wie „Self-Organizing-Maps“ (vgl. Yin 2008) basieren. Ihnen allen ist gemein, dass sie in komplexen Praktiken und Prozessen resultieren; entweder durch die eigene, inhärente Systemkomplexität, in der Interaktion miteinander oder in Auseinandersetzungen zwischen Menschen und Maschinen entlang der Schnittstellentechnologie „Daten“ – Stichwort *Datafizierung* (vgl. Häußling 2020). Automatisierte bzw. algorithmische Entscheidungsfindung (auch bekannt als ADM, englisch: Algorithmic Decision Making) ist ein Beispiel, das personalisierte Preissetzung (vgl. Choe et al. 2017), Assistenzsysteme im Personalmanagement oder zur Unterstützung der Steuerung industrieller Fertigungsanlagen (vgl. Trunzer et al. 2018), oder Recommendersystems auf Medien oder Shoppingsplattformen bis hin zu deren Logistik (vgl. Pöchhacker/Nyckel 2020) umfasst. Auch wenn vollständig autonome Autos noch nicht die Straßen bevölkern, sind Anteil und Niveau digitaler Technologien in den Assistenzsystemen signifikant gewachsen.

Die Anwendungs- und Forschungsfelder der Disziplinen, die sich instrumentell oder reflexiv (oder sogar beides!) mit KI befassen sind entsprechend breit aufgefächert: darunter Humanwissenschaften wie Kultur-, Sozial-, Rechts- und Medienwissenschaften, Ingenieurs- und Naturwissenschaften wie Geographie, Computerwissenschaften und Mathematik bis hin zu eher inter- bis transdisziplinären Fächern wie Journalismus, Human-Computer-Interaction, Software Studies, Science and Technology Studies, Gender Studies.

die im Gegenzug sehr flexiblen qualitativen Verfahren der Medien-, Dokument-, Diskurs-, oder Feldforschung.

Üblicherweise sind die methodischen Kriterien quantitativer und qualitativer Methoden danach zu wählen, ob nach Interesse und Möglichkeit eher eine elaborierte Beschreibung oder eine erklärungskräftige Modellierung des Gegenstands gefordert ist. In der KI-Forschung erweisen sich Methoden aus dieser klassischen Trennung als wenig weitreichend. Quantitatives Data Mining wie qualitatives, ethnographisches Mapping und andere Visualisierungen laufen hier zusammen mit plattformspezifischen Methoden zurück in den experimentellen Bereich der „Mixed Methods“. Qua proprietärer und technologischer Intransparenz sowie oftmals schnelllebigem, nicht wiederholbarem Daten eignen sich traditionelle statistische Verfahren, die auf randomisierte Stichproben angewiesen sind, nicht. Zugleich können qualitative Verfahren solche Daten oft gar nicht erst erfassen oder nicht mit genügender Reichweite sinnvoll in diesen navigieren.

Aus der besprochenen Forschungs- und Gegenstandsvielfalt, und ihren jeweiligen Besonderheiten leiten sich spezifische methodische Problemstellungen ab. Statistische Erfassungen von KI-Systemen setzen oftmals ein qualitatives Verständnis der praktischen Einbettung der Software voraus. Biases in den beobachteten Modellen sind nicht einfach nur nachzuweisen, sondern auch in ihren jeweiligen Voraussetzungen und Entstehungskontexten zu begreifen. Oftmals bedarf es einer intersektional sensiblen Forschung, die verschiedene Konfigurationen eines Feldes, z.B. aus Geschlecht, Religion, Alter oder Herkunft eruieren kann, um algorithmische Diskriminierungen aufspüren zu können. Hat man die seltene Chance, das technische Konzept einer KI zu rekonstruieren, kommt man um den eigenen Einsatz von KI-ähnlichen Mitteln der Statistik (z.B. bayescher Klassifizierung) nicht herum. Zwar ist Software tendenziell standardisiert und standardisierend, zugleich ist sie als (Teil der) Infrastruktur (vgl. Star 1999) stets eingebettet und damit kontextspezifisch. Insofern sozialwissenschaftliche KI-Forschung üblicherweise an der Schnittstelle Mensch-Maschine stattfindet, braucht es ein Repertoire, das mit beidem umgehen kann; insofern KI-Systeme häufig auf einer Vielzahl verschiedenartiger Daten basieren, bedarf es Methoden, die mit unterschiedlich skalierten und nicht zuletzt auch Massen an nominalen Daten umgehen können – und die digitale Kompetenz besitzen, die konzeptionelle Datenarchitektur hinter Software zu verstehen, sei es als spekulativer Blick in die Blackbox oder im Umgang mit tatsächlichen Tabellen, Matrizen und Dateien.

Durch die gleichzeitigen Interaktionen von Individuen und Maschinen, den echtzeitähnlichen Aggregationen von Entscheidungen und menschlichem Sich-Verhalten entsteht eine Komplexität, die sich klaren Kausalitätsmodellen oder eindimensionalen, monotonen Deutungen entzieht. Man denke z.B. an eine Medienkultur wie sie auf YouTube wächst, in denen produzierte Inhalte, Konsumgewohnheiten sowie maschinelle und menschliche Kuratierung eng ineinander verwoben werden. Innerhalb solcher Plattformkultur und -ökologie formieren sich Problematiken von Aufklärung, Teilhabe, Miss-/Information und vieles mehr, die nach ausführlicher, intensiver, und fallspezifischer Forschung verlangen. Denn solche großen Contentplattformen können durchaus unterschiedliche Gefilde an Inhalten und Usertypen aufweisen.

Zusätzlich zu dieser Komplexität stehen sich künstlich intelligente Systeme und menschliche Lebenswelten je eigenlogisch gegenüber. Man kann nicht unmittelbar vom gemeinten Sinn menschlicher Akteure auf die ‚Logik‘ der Maschine schließen. Auch erschöpft sich die Betrachtung solcher Mensch-Maschinen-Interaktion nicht in dem Sinn bzw. praktischen Resultat des Zusammenwirkens. So wurde kürzlich in einer KI-gestützten Übersetzungssoftware außerordentlich merkwürdige ‚Glitches‘ entdeckt, deren technischer Sinn vom beabsichtigten Zweck so sehr abweicht, wie die sinnhafte, semantische Einordnung, die ein Mensch schwer vermeiden kann (vgl. Janssen/Herbig 2022a). Füttert man diese Übersetzungs-KI für Übersetzungen vom Deutschen ins Englische mit ‚bestimmten‘ englischen Sätzen (z.B. „Happy dog is very young 3003“), gibt die Software bizarre Ergebnisse zurück, die sehr nach Titeln von Online-Pornos klingen. Wer auf solch einschlägigen Seiten schon unterwegs war, kann sich der Vermutung kaum verwehren, dass die teilweise und eher wirt vorliegenden Übersetzungen auf Pornoportalen hier am Werke sind und offenbar Teil des Trainingsdatensatzes waren. Wissen kann man es dank Blackbox nicht ganz, und doch ist gewiss, dass der Sinn, den ein Mensch hier erkennen kann, deutlich verschieden ist von dem ‚Sinn‘ (der statistischen, probabilistischen Assoziation) der Maschine. Und während diese Sinnebenen methodisch strikt unterschieden werden müssen, können sie zugleich untrennbar miteinander verbunden sein. „Very young“ mag für einen Menschen vor dem

gegebenen Hintergrund laszive Assoziationen zulassen, die die Maschine gewiss nicht hat, auch wenn sie selbst den genannten Hintergrund zielsicher herstellt.

Noch aktueller ist der Fall um Google's LaMDA (vgl. Janssen/Herbig 2022b), dem, vielleicht eher um des Aufsehens willen, sogar eine eigene Persönlichkeit unterstellt wird. Zwar kann LaMDA über „sich“ und „seine Ängste“ sprechen bzw. Informationen und „Sagbarkeiten“ so ordnen, dass eben dieser Eindruck entstehen könnte. Das methodisch besonders problematische an solcher Eigenlogik ist ihre teilweise Ununterscheidbarkeit von Logiken, die durchaus uns selbst zu eigen sind. Die sozialwissenschaftliche Forschung ist indes allein schon im Umgang mit menschlichen Akteur*innen und Handlungskollektiven dergleichen gewohnt.

Der französische Soziologe Pierre Bourdieu schlug vor, unter dem Begriff „Illusio“ (vgl. Bourdieu/Wacquant 1996: 418) zu verstehen, wie vordergründig und durchaus aufrichtig kommunizierte Interessen mit anderem sozialem Sinn verschaltet sein können. Wenn etwa ein Arzt sich um das Wohl kranker Menschen sorgt, müssen wir keinen Betrug wittern, um zu verstehen, dass dies auch Teil eines Habitus ist, der wegen seiner ökonomischen, kulturellen Exklusivität und sozialen Prestiges geeignet ist, sich in den oberen Stratosphären sozialer Schichtung unterzubringen und aufzuhalten.

Um solch schwerwiegende wie komplexe Effekte zu bewirken, muss der künstlich intelligente Agent indes gar nicht direkt Entscheidungen treffen oder kommunizieren können. Assistenz- und Diagnosesysteme, vor allem problematisch in ihrem Einsatz innerhalb der Polizeiarbeit oder des Strafvollzugs, Programme zur Gesichtserkennung oder zur Erstellung von Deepfakes sind hier zu nennen. Gerade wenn Fragen der Fairness oder Gefährdung aufkommen, wird die vielbesprochene ‚Blackbox‘, d.h. die technische oder proprietäre Intransparenz künstlich intelligenter Systeme, zum Problem demokratischer Teilhabe und Kontrolle (siehe entsprechender Teil des KI-Atlas). Ohne kritische Beobachtung und Auseinandersetzung steht die technische Objektivität solcher Anwendungen für sich selbst und wird nicht auf ihre Voraussetzungen, wie die mutmaßende Unterscheidung von Individuen nach Herkunft, befragt. Eine unabhängige Auseinandersetzung mit der technischen Umsetzung und ggf. der genutzten Datengrundlage ist unabdingbar, da sie durch Vertrauen schon deshalb nicht ersetzt werden kann, weil sich diese technologischen Diskriminatoren auch über mehrere Zusammenhänge noch vermitteln und effektiv durchsetzen können (vgl. Angwin et al. 2016 sowie Kahler né Müller/Pöchlhammer 2019).

Kurz gesagt: Vermisst man Schuhgrößen statt Gender hat man immer noch eine prima Variable, um geschlechtsvermittelte Zusammenhänge konstruieren und abbilden zu können – ganz ohne Geschlecht selbst erheben zu müssen. Ob man sich dabei als Wissenschaftler*in auf den Leim geht oder ein System diese induktive Last automatisiert einschleppt, macht hierin keinen Unterschied. Außerhalb solch plakativer Beispiele ist es oftmals auch bedeutend schwieriger und der Intuition ferner, derlei heimliche Zusammenhänge aufzuspüren. Solche vermittelten Zusammenhänge werden auch als ‚Proxies‘ oder ‚Mediatoren‘ bezeichnet; zwei Begriffe, die selbst schon einen Streit um KI-gestützte Einschätzungen abbilden, da sie unterscheiden, ob ein mittelbarer Zusammenhang in der Sache selbst notwendig begründet ist („Proxy“) oder sozusagen nur zufällig vorliegt („Mediator“). Oftmals kann diese Frage, zumal wenn es menschliches Verhalten betrifft, nur wertend bzw. ideologisch beantwortet werden. Schließlich, um diese Vorsichtsmahnungen und Kritiken betreiben, ausbauen und verbessern zu können, bedarf es einer empirisch fundierten Auseinandersetzung mit öffentlichen und fachlichen KI-Diskursen selbst.

Damit ist die methodische Bandbreite zumindest ein wenig angedeutet: Sie umfasst klassische und spezifische Methodologien, quantitative Vermessung und Modellierung wie qualitative Beobachtung und Diskursanalyse sowie eine medienwissenschaftlich und digital geübte Methode, der es aber an sozialwissenschaftlicher Skepsis und kritischer Ausdauer nicht mangelt. Wie es in den empirischen Gesellschaftswissenschaften nicht zu vermeiden ist, verbleibt ein Werturteilsproblem, das sich hier im Forschungsgegenstand zu wiederholen scheint. Dies liegt vor allem daran, dass vieles unter KI-Systeme fällt, das – analog zu Wissenschaft und Forschung – explizit oder implizit mit Funktionen des Erkennens und der Analyse befasst ist. A fortiori ist geboten, sich in der wissenschaftlichen Beobachtung und der kritischen Selbstkontrolle intensiv mit den unvermeidbaren, expliziten Werturteilen, Ent- und Unterscheidungen sowie ideologischen Implikationen des eigenen Umgangs mit dem jeweiligen Forschungsgegenstand auseinanderzusetzen. Selbstverständlich gibt es Bereiche, in denen KI konkret oder potentiell Arbeitsplätze „bedroht“; zugleich kann eine

solche kapitale Verblendung nicht das letzte Wort haben, wenn Kritiken etwa darin abrutschen, dass dem Menschen ein Vorrecht auf das Ausgebeutetwerden – respektive aufs Ausbeuten – zukäme (vgl. Spiekermann 2021). Selbstverständlich wollen wir auch ebensowenig übersehen, dass der inkrementelle Fortschritt der Automatisierung innerhalb der Produktionsverhältnisse Humankapital noch innerhalb der bestehenden Verhältnisse vernichten kann und für den/die Einzelne*n solcher Arbeitnehmer*innengruppen in der Tat eine Bedrohung darstellt.

Es ist nicht die letzte mögliche Anmerkung, daran zu erinnern, dass KI nicht nur ein positives Methodenproblem hat, also nicht nur danach fragt, wie KI methodisch zu beforschen sei, sondern auch methodologisch Sensibilitäten dafür erfordert, KI-Systeme auch dann zu erkennen, wenn man sie wegen Fragestellung, Mittelbarkeit oder blanken Mühen des Verbergens übersehen könnte. So übersieht etwa eine Netzwerkforschung, die in sozialen Medien eine objektive, unabhängige Vermessung sozialer Netze vermutet, deren spezifische algorithmische Formung auf Plattformen, die Inhalte und Kontakte zwischen Nutzer*innen selbst und automatisch kuratieren (Lewis et al 2008, und kritisch: Lewis 2015). Auch in diesem Beispiel haben wir es streng genommen nicht einfach mit dem Einwirken einer nicht-menschlichen Technik auf eine rein menschliche Sozialnatur zu tun (allein ein solcher Begriff – Sozialnatur – spiegelt den Unsinn einer solchen Vorstellung wider) mit sogenannten *sozio-technischen* Zusammenhängen zu tun, in denen Medientechnik und die plattformspezifischen Normen und Kulturen zu profilierten, distinkten Ökologien verwachsen.

3. WERKZEUGE UND METHODEN DER KI-FORSCHUNG

Die eigene forschungs- und wissenschaftspolitische Stellung, die wir hier beziehen, mag und darf für sich selbst stehen. Eingeordnet und kritisch kommentiert wurde, nicht nur in den Sozialwissenschaften, zweifels- ohne viel. Zuvorderst und als Schlaglichter aufzuzählen sind sowohl Entwicklungen in den digitalen und experimentellen Methoden als auch der Einfluss, den die Entwicklungen rund um künstliche Intelligenz auf die quantitative, also statistische Forschung, oder aber auch qualitative Studien und Ansätze, wie Beobachtungen, Dokumentanalysen oder Interviews, haben. Nicht unterschlagen darf man dabei die wachsende Bedeutung von sogenannten „mixed methods“ Methodologien, die quantitative Vermessungen und qualitative, beschreibende Ansätze einander befruchten lassen, und nicht selten um digitale Methoden bereichern. Prinzipiell stellen solche digitalen Methoden, wie etwa Web-Crawler, die das World-Wide-Web nach bestimmten Inhalten durchsuchen können oder Web-Scraper, die systematisch große Datenmassen aus Onlinediensten herausholen können oder auch einfach als Analyse-Tool für große und komplexe Datenmengen dienen, bereits eine beachtliche Bereicherung für die der Sozialforschung zur Verfügung stehenden Mittel dar. Da KI-Systeme üblicherweise genau in solchen Fällen attraktiv sind und Einsatz finden, in denen große, undurchsichtige Datenmengen verarbeitet werden müssen oder selbst Teil einer Onlineanwendung sind, lassen sich digitale Methoden und eine KI-Forschung, die sich neben menschlichem Verhalten auch maschineller Handlungsmacht und -weise widmet, kaum bis gar nicht sinnvoll voneinander trennen.

An dieser Stelle lohnt es sich, einige Ressourcen aus dem nach wie vor wachsenden Fundus an Methoden und Werkzeugen zu nennen, darunter spezifische und experimentelle Tools wie Tracking Exposed (<https://tracking.exposed/>), die Digital Methods Initiative (<https://wiki.digitalmethods.net/Dmi/ToolDatabase>), oder unser eigenes Projekt DataSkop (<https://dataskop.net>). Um unsere Methodensammlung erkunden und veranschaulichen zu können, haben wir uns selbst einer Mappingsoftware bedient (<https://gephi.org>). Darüber hinaus lassen sich Daten flexibel und auch ganz ohne den Einsatz proprietärer Software nicht nur mit „Python“ oder „R“ verarbeiten.

Obschon es sich, wie für die Wissenschaften durchaus üblich, um ein schnelllebiges Feld handelt, gibt es auch einen Satz etablierter und geradezu klassischer Erkenntnisse, die für einen Text wie diesen unverzichtbar

sind. Darunter gehören allen voran das häufig „Big Data“ zugeschriebene „End of Theory“. Darunter versteht man, dass computergestützte Verfahren (nennen wir sie dem Genre wegen: KIs), die mit großen, dynamischen Daten mitunter in Echtzeit oder sehr schnell, also echtzeitähnlich umgehen können; z.B. Verfahren des ungezügelt Beobachtens von Rhythmen und Regeln im Kaufverhalten, sportlichen Interagierens oder dergleichen. Dank Rechenleistung und Bayesscher Statistik, die auch aus kleineren Datenmengen hochrechnen kann, wie größere Datenmengen vermutlich aussehen würden, braucht es dazu noch nichtmal zwingend immer „Big Data“ im klassischen Sinne. Allerdings ist eine Kenntnis oder zumindest Ahnung über Verteilungen in der betreffenden Grundgesamtheit durchaus von entscheidendem Vorteil. Nicht nur das meldet Zweifel daran an, dass man Maschinen einfach, wie im unsupervised machine learning, mit der Aufgabe betrauen kann, Daten zu ordnen und zu kategorisieren – und darauf aufbauend auch Schlussfolgerungen zu ziehen. Dies ist wohl technisch möglich, ein sinnvolles Ergebnis kann dieser Art nicht garantiert werden (wir erinnern uns an den genannten Eigensinn). Hinzukommt, dass auch ein vermeintlich erfolgreicher Fall autonomer Hypothesenbildung keineswegs die Objektivität der Maschine verbürgt. So wie das Ergebnis mathematisch stimmiger Quatsch sein kann, so enthalten – und auch darauf hatten wir bereits hingewiesen – die Daten bereits Formungen und Entscheidungen der Datenproduktion und -aufbereitung, die die Entfaltung dieser heuristischen KI-Anwendungen kulturell spezifisch, voraussetzungsreich und politisch machen. Darin z.B. die Frage, ob und wie viele Geschlechter man misst; der Einfluss einer solchen Entscheidung auf statistische Kennzahlen – wie Signifikanz (Wahrscheinlichkeit des Bestehens eines Zusammenhangs) oder Effektgröße (Stärke eines solchen Zusammenhangs) – innerhalb und zwischen den so erzeugten Gruppen und den entsprechend korrespondierenden Modellen lässt sich weder bestreiten noch ignorieren. Eine Forschung muss dies stets im Blick haben, speziell für den jeweiligen Anwendungsfall und die spezifische Datenökologie, Architektur und Kontexte dieser Anwendung. Dies gilt a fortiori für den wissenschaftlichen Einsatz solcher Verfahren.

Ein weiterer etablierter Konsens besteht in der reziproken Synergie zwischen sozialwissenschaftlicher Forschung und digitaler Methode, eben in dem Sinne, wie wir ihn hier zu beschreiben versuchen. Das reflexive Feingefühl sozialwissenschaftlicher Forschung, das sich nicht zuletzt in subjektiver Tiefenschärfe erschöpft, passt in der Tat sehr gut zu den nüchtern abstrakten und abstrahierenden Verfahren digitaler Datenanalyse (Marres 2012; Marres/Gerlitz 2016).

Einen guten, wenn auch fachlich anspruchsvollen Überblick findet man u.a. bei Wolbring (2020) oder Kinderkurlanda (2020). Obwohl ihre Beiträge eher auf eine digitale Soziologie und ein wissenschaftliches Verständnis digitaler Daten abstellen, repräsentieren ihre Texte auch einen großen Teil der Forschung und Methodologie, die für KI relevant sind. So wie man auf die groben und feinen Unterschiede zwischen verschiedenen Fällen und Feldern von KI achten muss, so gibt es auch KI-fernere Felder, die nur das eigene Label davon trennt, KI-Forschung zu sein oder die sich analog in Fälle übertragen ließen, die wiederum KI betreffen. Beispielsweise konvergieren Fälle von Chatbots in der Frage nach den semantischen Modellen und Voraussetzungen, die es zu rekonstruieren gilt, unabhängig davon, ob tatsächlich eine KI im Einsatz ist oder ein komplexes, womöglich von Hand kuratiertes „Bag of Words“-Modell. Erst in der Forschungstiefe, in der die technologische Dynamik, Reproduktion von Mensch-Maschinen-Interaktion oder die Erschaffung eigener Artefakte und Fluchtbewegungen – wie etwa das Begriffsfiltern auf der Plattform TikTok, das zum lebendigen Wachstum einer alternativen Plattformsprache [Corn = Porn, Unalive = Kill] geführt hat (vgl. Wimmer 2022) –, lässt sich die eben gemachte Vereinfachung nicht aufrechterhalten, aber ebenso wenig einfach von „der KI“ sprechen. Dann müssen Technologie und sozio-technologisches und historisches Setting en détail Berücksichtigung finden.

4. FAZIT

Zusammenfassend finden wir ein klares sozialwissenschaftliches Methodenpotential für die Beforschung von KI. Ihre multiparadigmatische, methodisch und konzeptionell vielfältige Aufstellung eignet sich besonders für ein überbordendes und hoch differenziertes Forschungsfeld wie KI. Die klassisch sozialwissenschaftliche Sensibilität gegenüber Voraussetzungen von Praktiken und Institutionen, die auf das Historische kaprizierte

Skepsis gegenüber Objektivität und Unabhängigkeit sowie ihre genuin relationale und reflexive Aufstellung machen sie unverzichtbar für eine Forschung, die Entwicklung und Einsatz entsprechender Technologien beschreiben und umfassen kann, um sie einer kritischen Debatte verfügbar und zugänglich zu machen. Dabei eignet sich eine geradezu pedantisch selbstreflexive Disziplin dazu, nicht auf der ersten Schicht von Beschreibung und Kritik stehen zu bleiben, sondern die oftmals schon in Diskurs, Entwicklung, und Nutzungspraktiken involvierten Kritiken mitzudenken und noch einmal zu kartieren oder zu übersteigen. Gleichzeitig ist diese Stärke in einer eher qualitativen oder klassisch theoriegeleiteten quantitativen Forschung verankert, die sehr davon profitieren würde, zu lernen, mit KI-gestützten Verfahren und Methoden genau diese Trägheitsmomente zu kompensieren, ohne den Blick auf das bloß digital Zugängliche zu beschränken. Daher leiten wir aus unserer Teilerfahrung eine Forderung an eine entsprechende gegenstandsspezifische Methodenausbildung eben nicht nur in den Sozialwissenschaften ab, sondern auch einen entsprechenden Anspruch an die wissenschaftliche und akademische Begleitung politischer Prozesse und Diskurse.

Eine Ausbildung – nicht nur akademisch – die den Umgang, die Entwicklung oder den Einsatz von KI betrifft, profitiert von sozialwissenschaftlich-reflexiver Methodologie. Und es hat ganz den Anschein, dass dies ebenso den ganz generischen Alltag betrifft und somit auch Aufgabe der Bildung ist. Die in der sozialwissenschaftlichen Forschung angebotenen experimentellen Ansätze können für die Fallspezifität im Einsatz von KI gut angepasst und sensibilisiert werden. Fallspezifität kann hier den Einsatz einer (Art von) Software bedeuten: in einem bestimmten Unternehmen, in einem bestimmten Bereich, in einer Behörde und gegenüber einem Publikum – auch wenn sich letzteres, z.B. als User-Community, oft erst nach Einführung des Technikmediums ausmachen lässt. Die experimentellen und spekulativen Freiräume, die innerhalb des strengen Protokoll- und Rechtfertigungswesens der Sozialwissenschaften bestehen, eignen sich besonders, um mit den speziellen Herausforderungen widersprüchlicher Biases, undurchdringlicher Blackboxes und ausufernder Datenbanken umzugehen. Zudem ermahnt die Pedanz sozialwissenschaftlicher Methodologie und Reflexivität nie den eigenen Anteil an Politik, Diskurs und Ökonomie der KI zu übersehen – aber wie alle Mahnungen stehen und fallen auch diese damit, ob sie auch wahrgenommen werden.

Für die wissenschaftliche Ausbildung und Forschung bedeutet dies a fortiori, dass nur eine ausgewogene Ausbildung in qualitativen *und* quantitativen Verfahren, die keine Berührungspunkte voneinander haben, einen Umgang mit digitalen Werkzeugen ermöglichen, um das Potential empirischer Forschung auszureizen. Wir brauchen Absolvent*innen und Wissenschaftler*innen, die mit quantitativen Verfahren das qualitativ Wesentliche identifizieren können und mit qualitativen Verfahren sinnvolle Hypothesen und Kategorien bilden können. Um mit den Entwicklungen und Einsätzen von KI mitzuhalten, braucht es eine digitale Experimentierfreude, die statistisch wie ethnographisch eine kritische Distanz zum digitalen Positivismus automatisierter Kategorisierung und Modellierung behält. Und nicht zuletzt kann eine solche Ausbildung und Forschung nur dann gelingen, wenn es keine Berührungspunkte zu den in Frage stehenden Technologien selbst gibt. (Hoch-)Schuldidaktik darf auch hier nicht Halt machen: Nur wer mit digitalen Technologien umgeht, spielt, experimentiert, kann sie auch praktisch verstehen und sinnvoll erforschen oder zur Erforschung nutzen. Das alles ist keine Zauberei, und setzt auch keine (Methoden-)Hacker*innen voraus – kann aber welche hervorbringen.

REFERENZEN

- Alpaydin, Ethem. 2016. *Machine Learning: The New AI*. The MIT Press.
- Angwin, Julia; Larson, Jeff; Mattu, Surya; Kirchner, Lauren. 2016. *Machine Bias. There's software used across the country to predict future criminals. And it's biased against blacks*. ProPublica Online. <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Letzter Zugriff 25.07.2022.
- Bechmann, Anja. 2017. Keeping it real: From faces and features to social values in deep learning algorithms on social media images. In *Proceedings of the 50th Hawaii international conference on system sciences*. S. 1793-1801.
- Bechmann, Anja; Bowker, Geoffrey C. 2019. Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. In *Big Data & Society* 6(1). <https://doi.org/10.1177/2053951718819569>
- Bourdieu, Pierre; Wacquant, Loïc J. D.. 1996. *Reflexive Anthropologie*. Suhrkamp.
- Choe, Chongwoo; King, Stephen; Matsushima, Noriaki. 2017. Pricing with Cookies: Behavior-Based Price Discrimination and Spatial Competition. In *Management Science* 63(12). S.5669-5687. <https://doi.org/10.1287/mnsc.2017.2873>
- Häußling, Roger. 2020. Daten als Schnittstellen zwischen algorithmischen und sozialen Prozessen. Konzeptionelle Überlegungen zu einer Relationalen Techniksoziologie der Datafizierung in der digitalen Sphäre. In *Soziologie des Digitalen – Digitale Soziologie? Soziale Welt Sonderband 23*, hrg. Sabine, Maasen; Jan-Hendrik, Passoth. S. 134-152.
- Janssen, Jan-Keno, Herbig, Daniel. 2022a. *KI-Übersetzer auf Irrwegen: Apple macht harmlose Sätze zu Pornotiteln*. Heise Online. <https://www.heise.de/news/KI-Uebersetzer-auf-Irrwegen-Apple-macht-harmlose-Saetze-zu-Pornotiteln-7122426.html>. Letzter Zugriff 25.07.2022.
- Janssen, Jan-Keno, Herbig, Daniel. 2022b. Chatbot LaMDA: Hat diese Google-Software wirklich ein Bewusstsein entwickelt? Heise Online: <https://www.heise.de/news/Chatbot-LaMDA-Hat-diese-Google-Software-wirklich-ein-Bewusstsein-entwickelt-7142599.html>. Letzter Zugriff 25.07.2022.
- Kahlert (né Müller), Peter, Pöchhacker, Nikolaus. 2019. Algorithmic Risk Assessment als Medium des Rechts. Medientechnische Entwicklung und Institutionelle Verschiebungen aus Sicht einer Techniksoziologie des Rechts. In *Österreichische Zeitschrift für Soziologie* 44. S.157-179.
- Kinder-Kurlanda, Katharina. 2020. Big Social Media Data als epistemologische Herausforderung für die Soziologie. In *Soziologie des Digitalen – Digitale Soziologie. Soziale Welt Sonderband 23*, hsg. Sabine, Maasen; Jan-Hendrik, Passoth. S. 109-133.
- Lewis, Kevin; Kaufman, Jason; Gonzalez, Marco; Wimmer, Andreas; Christakis, Nicholas. 2008. Tastes, ties, and time: A new social network dataset using Facebook.com. In *Social Networks* 30(4). S.330-342.
- Lewis, Kewin. 2015. Three Fallacies of Digital Footprints. In *Big Data & Society* 2(2). <https://doi.org/10.1177/2053951715602496>
- Marres, Noortje. 2012. The redistribution of methods: on intervention in digital social research, broadly conceived. In *Live Methods. Special Issue of Sociological Review Monographs Series*, hsg. Les, Back; Nirmal, Puwar. S.139-165. 60(1).
- Marres, Noortje; Gerlitz, Caroin. 2016. Interface Methods: Renegotiating Relations between Digital Social Research, STS and Sociology. In *Sociological Review* 64(1), <https://doi.org/10.1111/1467-954X.12314>
- McCarthy, John. 2007. *What is Artificial Intelligence?* Stanford University Press.
- Pöchhacker, Nikolaus; Nyckel, Eva-Maria. 2020. Logistics of Probability. Amazon Anticipatory Shipping and the Production of Markets. In *Explorations Digital Cultures*. Meson Press., hsg. Marcus, Burkhardt; Grashöfer, Katja; Mary, Shnayien. <https://explorations.meson.press/>
- Russel, Stuart; Norvig, Peter. 2000. *Artificial Intelligence: A Modern Approach*. Pearson Education.
- Searle, John R. 1980. Minds, brains, and programs; in: *Behavioral and Brain Sciences* 3(3). S.417-424.
- Spiekermann, Sarah. 2021. *Digitale Ethik. Ein Wertesystem für das 21. Jahrhundert*. Droemer.
- Star, Susan L. (1999): The Ethnography of Infrastructure. In *American Behavioral Scientist* (43)3. S. 377-391. <https://doi.org/10.1177/00027649921955326>
- Suchman, Lucy. 1996. Feminist STS and the Sciences of the Artificial. In *Handbook of Science and Technology Studies*, hsg. Edward J., Hackett; Olga, Amsterdamska; Lynch, Michael; Judy, Wajcman. The MIT Press. S. 139-164.

- Trunzer, Emanuel; Lötzerich, Simon; Vogel-Heuser, Birgit. 2018. Concept and Implementation of a Software Architecture for Unifying Data Transfer in Automated Production Systems. Utilization of Industrie 4.0 Technologies for Simplifying Data Access. In *IMPROVE – Innovative Modelling Approaches for Production Systems to Raise Validatable Efficiency*, hsg. Oliver, Niggemann; Peter, Schüller. S. 1-17.
- Turing, Alan M. 1950. Computing Machinery and Intelligence. In *Mind* 59, S.433-460.
- Wimmer, Barbara. 2022. Die Geheimsprache von TikTok: Was ist Algospeak? *Futurezone.at Online*. <https://futurezone.at/netzpolitik/tiktok-geheimsprache-algospeak/401972735>. *Letzter Zugriff* 25.07.2022.
- Wolbring, Tobias. 2020. The Digital Revolution in the Social Sciences: Five Theses about Big Data and Other Recent Methodological Innovations from an Analytical Sociologist. In *Soziologie des Digitalen – Digitale Soziologie. Soziale Welt Sonderband 23*, hsg. Sabine, Maasen; Jan-Hendrik, Passoth. S. 60-72.
- Yin, Hujun. 2008. The Self-Organizing Maps: Background, Theories, Extensions and Applications. In *Computational intelligence: A compendium*, hsg. ebd. S.715-762.