

**Workshop: Mensch-Maschine-Interaktion und
Informationsdarstellung
Ruhr-Universität Bochum
20.-21. April 2023**

Abstracts

Künstliche Intelligenz trifft Anthropologie: Menschenbilder im Spiegel der Technik

Menschliche Selbstbeschreibung findet nicht nur auf Basis kultureller Traditionen oder religiöser Überzeugungen statt, sondern auch im Medium der Technik und des technischen Wandels. Das mechanische Zeitalter des *L'Homme Machine* (LaMettrie), der *homo laborans* in der Anthropologie von Karl Marx oder der gegenwärtige *homo digitalis* sagen, so drei Beispiele, nicht nur etwas über Menschen aus und sind nicht nur anthropologische Verhaltungen, sondern spiegeln auch den technischen Entwicklungsstand ihrer Zeit.

Gegenwärtig fordern insbesondere autonome, auf KI und maschinellern Lernen beruhende Techniken, so etwa ADM-Systeme (*automated decision-making*) oder soziale Roboter, das Selbstverständnis von Menschen heraus. Denn diese Techniken sind nicht mehr bloß passiv, denn sie können in bestimmte Kontexte intervenieren bzw. sollen dies explizit. In neuen Mensch/Technik-Konstellationen verschieben sich Subjekt/Objekt-Verhältnisse, in denen der Mensch zunehmend von der traditionellen Subjekt- in eine Objektrolle gegenüber technischen Artefakten gerät, indem er deren Aktionen ausgeliefert wird. Diese neuen Konstellationen bedürfen analytischer und empirischer Aufklärung sowie anthropologischer Reflexion, sowohl zum Zwecke des besseren Verständnisses als auch um die Erkenntnisse in ethischer Hinsicht praktisch nutzen zu können

Im Vortrag werde ich zunächst einen kurzen Streifzug durch die Geschichte von Menschen und Technik unternommen, um die kulturstiftende, soziale aber eben auch anthropologisch bedeutsame Rolle von Technik zu illustrieren. Sodann fokussiere ich auf gegenwärtige Narrative vom Menschen angesichts der von ihm selbst geschaffenen KI-basierten und zusehends autonomen Technik. Das Narrativ vom unterlegenen Menschen und der überlegenen KI steht dabei im Mittelpunkt. Die nähere Betrachtung dieses Narratives führt auf digitale Maschinenmodelle des Menschen als notwendig zu präsupponierende Hintergrundannahme. Abschließend werde ich aus den Ergebnissen Anfragen an eine technikanthropologisch gewendete Philosophische Anthropologie formulieren.

Mangel an emotionaler Granularität? Zu einem Problem von digitaler Gefühlserfassung

Zu den jüngeren Entwicklungen im Feld der künstlichen Intelligenz gehört Affective Computing, wie die Pionierin Rosalind Picard dasjenige Feld der Computerwissenschaften genannt hat, das sich auf Emotionen bezieht – das also Systeme zu entwickeln erlaubt, die Gefühle erfassen, gezielt hervorrufen oder simulieren.

Eine erste konkrete, spezifisch auf AC bezogene Kritik, die man bisher in der Literatur findet, lautet, dass ein weit verbreiteter Einsatz von Gefühlserfassungssystemen dazu führen könnte, dass wir einen Verlust an emotionaler Granularität erleben würden. Mit anderen Worten: dass unsere Fähigkeit, fein unterschiedene Gefühle zu erkennen und womöglich auch überhaupt zu erleben, abnehmen würde. Dieser Kritik gehe ich in diesem Vortrag nach.

Dazu stelle ich diese Hypothese zunächst etwas genauer dar und zeige, auf welchen Prämissen sie beruht. Im Folgenden gehe ich genau auf die Annahme ein, dass es den Systemen selbst an emotionaler Granularität mangelt (inwieweit das dann auch zu einem Verlust solcher Granularität in der Emotionalität von Nutzenden führen kann, lasse ich als empirische Frage offen). Dabei lässt sich zeigen, dass es für ein erstes Verständnis von mangelnder Granularität womöglich eine relativ einfache technische Lösung geben könnte. Doch ich werde anschließend argumentieren, dass emotionale Granularität in einem zweiten Sinn auf andere, und zwar grundlegendere Weise verfehlt wird.

Mein Fazit lautet: Das Problem der gegenwärtigen Grundlagen von Affective Computing-Anwendungen ist nicht nur, dass die Basistheorie zu wenig Emotionstypen vorsieht und nicht nur, dass Ausdruck und Emotionstyp nicht verlässlich und spezifisch zusammenhängen, sondern darüber hinaus, dass weder der Typ noch der Ausdruck etwas über die Intentionalität von bestimmten, manifesten Emotionen aussagt. Eigentliche Granularität der Emotionen und dessen, was zur emotionalen Granularität befähigte Menschen anderen in relevanter Weise voraushaben, hängt nicht nur von Emotionsbegriffen ab, sondern davon, zu erkennen, worauf genau sich eine Emotion bezieht und vor welchem Hintergrund genau sie entstanden ist – es braucht als Erfassen der gesamten Intentionalität. Die Furcht vor einer Prüfung ist nicht nur zu unterscheiden von der Wut über eine nicht bestandene Prüfung, sondern auch von der Furcht vor einem Flugzeugabsturz oder den Folgen des Klimawandels. In allen Fällen ist Gefahr das formale Objekt, aber die anderen beiden Stellen der Intentionalität sind ganz unterschiedlich besetzt. Wie diese Intentionalität von Emotionen mit Affective Computing erfasst werden soll, ist bisher unklar.

Funktioniert - Kaputt. Schwierigkeiten im Umgang mit einer Leitdifferenz der Technik im Kontext von KI

Von Niklas Luhmann stammt das Bonmot, Technik sei das, was kaputt gehen könne (Luhmann 1990: 263). Der Witz von Luhmanns Bemerkung liegt nicht nur darin, dass Technik vorrangig auf das Funktionieren zielt - und erst daher auch "kaputt" sein kann. Die Leitdifferenz spielt vielmehr eine systematische Rolle in der *Konstruktion* und *Gestaltung* von Technik. Dies gilt mit Nachdruck auch für ihre *Verwendung* im Alltag: Das Verständnis, was man mit jeweiliger Technik machen kann, wie man sie nutzt als auch wie man lernt, mit ihr besser umzugehen, sind daran gebunden, dass man *weiß*, wie sie sich verhält, wenn sie funktioniert. Zentrale Begriffe der Mensch-Maschine-Interaktion wie Antizipation, Verstehen und damit auch Interaktion selbst basieren auf dieser Voraussetzung.

Die weit überwiegende Anzahl von Techniken, die wir bislang im Alltag verwenden, erfüllt diese Voraussetzung relativ unproblematisch. Wir können mit Blick auf den Kühlschrank, den Toaster oder das Licht in der Regel ohne weiteres sagen, ob sie funktionieren, da wir eine direkte Einsicht hinsichtlich unserer Interaktion mit dieser haben: Wir haben Erwartungen darüber, was passieren sollte, wenn wir eine Aktion initiieren, wie wir diese Aktion initiieren und eine Einsicht, was darauf hin erfolgt - so dass wir unsere Erwartung daraufhin bilanzieren können, ob sie erfüllt wurde oder nicht. Diese Voraussetzung wird problematischer mit Blick auf technische Systeme, bei denen unklarer ist, (a) auf was die jeweiligen "Inputs" der Interaktion beruhen, (b) welche internen Regeln zur Verarbeitung der "Inputs" bestehen und wie diese Regeln ggf. angepasst werden sowie (c) welche "Outputs" daraufhin erfolgen als auch erfolgen sollen.

IT- und insbesondere KI-Systeme *können* diese Unklarheit z.T. drastisch erhöhen. In diesen Fällen wird es sodann schwieriger zu beurteilen, ob eine jeweilige Technik funktioniert, zu bewerten, wie gut sie funktioniert, und zu antizipieren, wie funktionierende Technik sich verhalten würde. Bislang wurde mit Blick auf KI-Modelle vor allem ihre *epistemische* Opazität (für Expertinnen) betont und untersucht. Die hier thematisierte Problematik ergänzt diesen Blick auf Formen einer *pragmatischen* Opazität, die den Umgang mit Technik im Alltag betrifft.

Errechnete Lagebilder. Sehgewohnheiten und Entscheidungsunterstützung im Rettungseinsatz

An Verbundprojekten mit KI-Komponenten ist üblicherweise eine ethische Begleitforschung beteiligt. Ihre Aufgabe ist die Identifizierung bestehender Wertekonflikte sowie möglicher Risiken in Absprache mit den betroffenen Personengruppen und die Mitgestaltung möglicher Maßnahmen zur Minimierung dieser Risiken. Man kann die Beobachtung machen, dass es in dieser Art von interdisziplinär ausgerichteten Forschung zwei Arten von Risiken und Erfordernissen gibt: die sofort einsichtigen, aus anderen Kontexten bekannten Risiken – Sicherheit, Schutz vor Hacking, keine Gefährdung beteiligter Person durch die eingesetzte Technologie – und die, zu deren Klärung und Verständnis man etwas theoretischen Vorlauf braucht. Zu ersteren zählen mittlerweile auch Risiken im Kontext von KI, die bereits unter ethischen Gesichtspunkten breit diskutiert werden und für deren Einsatz bereits Handlungsempfehlungen erarbeitet und weiterentwickelt werden.

Eine spannende Frage, auch unter dem Aspekt der Abschätzung von Technikfolgen in noch nicht etablierten Formen der Techniknutzung, ist nun, wie man Risiken des letztgenannten Typus speziell im Rückgriff auf erkenntnistheoretische und technikphilosophische Diskurse ermitteln kann. Ich werde in meinem Beitrag an einem Beispiel aus der KI-gestützten Rettungspraxis zeigen, wie dieser theoretische Vorlauf der Risikoidentifikation aussehen kann. In meinem Beispiel geht es um ein UAV (*unmanned aerial vehicle*, „Drohne“), das teilautonom in teileingestürzte Gebäude einfliegt und den außen befindlichen Einsatzkräften eine 3D-Lagekarte mit lokalisierten Lebenszeichen zur Rettungsplanung zu Verfügung stellt. Dieses aus Lidar- und Radarsignalen errechnete Bild (im Gegensatz zu einem klassischen Kamerabild) stellt hinsichtlich Darstellung und Einstellmöglichkeiten hohe Anforderungen an das User Interface. Diese Anforderungen werde ich auf Wahrnehmungsgewohnheiten zweierlei Art beziehen: zum einen auf unsere Sehgewohnheiten im ‚analogen‘ Raum, in dem wir uns selbst leiblich befinden, und zum anderen auf unsere Gewohnheiten bei der Verwendung von bildhaft dargestellter Information.

Abrichten, Konditionieren, Lernen: Künstliche Intelligenz und Behaviorismus

Ausgehend von den Mensch-Maschine-Interaktion Konzeptionen von Georges Canguilhem und Gilbert Simondon, soll in diesem Vortrag der Einfluss des Behaviorismus auf die Künstliche Intelligenz vorgestellt werden.

Canguilhem hat in seinem Aufsatz, *Maschine und Organismus*, versucht zu zeigen, dass das 20. Jahrhundert geprägt ist von einer Biologisierung der Technik, – und nicht einer Mechanisierung des Organismus', wie noch bei Descartes. Maschinen und andere technische Objekte werden ab nun mit biologischen Begriffen à la Anpassung, Evolution, Milieu beschrieben.

Mit der Künstlichen Intelligenz zeigt sich seit ihrem historischen Auftauchen nun, zusätzlich zur Biologisierung, vor allem eine Psychologisierung der Technik. Wichtige Figuren der KI Entwicklung (Turing, Rosenblatt, Simon) übernehmen Konzepte und Modelle von Lernen und Intelligenz (Abrichten, Konditionieren, Versuch-Irrtum usw.) aus dem Behaviorismus, ein Gestus, der sich bis in die heutige KI Forschung erhalten hat. Ich werde sowohl auf die historische Entwicklung als auch auf rezente Beispiele aus der KI verweisen.

Dabei lassen sich markante Mensch-Maschine Differenzen und Unterscheidungen feststellen, die ich mit Hilfe von Simondons Technikphilosophie beschreiben werde.

Sicher ist (nicht) sicher. Der Beitrag der Ethik zur IT-Sicherheit

Die Mensch-Maschine-Interaktion basiert auf digitalen Systemen. Eine Gefährdung der Sicherheit dieser Systeme kann unter Umständen auch zu einer Gefährdung diverser Sicherheitsaspekte für den Menschen, wie seiner Privatsphäre, führen. Zum Ausdruck kommt dies beispielsweise bei der Betrachtung der Nutzung des Smartphones als Navigationssystem und der dabei gesammelten Trackingdaten. Durch eine unvorsichtige Nutzung des Systems, beziehungsweise die leichtfertige Abgabe von Datennutzungsrechten, oder durch kriminelle Angriffe auf das System, bei dem die Daten geklaut werden, erlangen Dritte Zugriff auf die Daten. Diese können wiederum weiterverarbeitet und zur Identifizierung oder sogar Profilierung der Nutzenden verwendet werden. So stellt sich hier eine (nicht neue, aber komplexe und zugleich noch nicht eingedämmte) Bedrohungslage für die Privatsphäre Nutzender dar. Für eine sichere Mensch-Maschine-Interaktion ist die IT-Sicherheit daher relevant.

Es stellen sich jedoch Fragen über die Umsetzung der Sicherheit in IT-Systemen: Welche (ethischen) Werte sollten bei der IT-Sicherheitsforschung im Vordergrund stehen? Wie sollten Forschende mit gefundenen Unsicherheiten in Systemen umgehen? Wer trägt die Verantwortung für Konsequenzen, die aus dem Eingriff in die Sicherheit der Systeme resultieren? Welche Rolle spielt der Mensch in der IT-Sicherheit? Fragen wie diese sind derzeit noch nicht umfassend beantwortet. Im professionellen IT-Sicherheitsbereich fehlt ein stabiles ethisches Fundament, welches sich auch den praktischen Fragen der Umsetzung zureichend widmet. Derzeit sind im Feld der IT-Sicherheit indes seitens relevanter Akteur:innen Entwicklungen zu beobachten, die sich dieser Lücke annehmen. Für die Ethik ergibt sich hier eine Gelegenheit, die Akteur:innen im Feld der IT-Sicherheit bei der Entwicklung einer ethischen Leitlinie (und je nach Definition sogar ihrer eigenen Bereichsethik) zu beobachten und beispielsweise durch die Einbringung von Erkenntnissen aus artverwandten Bereichsethiken oder durch die theoretische Eingrenzung des Themengebiets zu unterstützen. Am Beispiel der IT-Sicherheit - einem basalen und zugleich unterschätzten Bereich des Digitalen - lässt sich so zeigen, dass ein praktischer Beitrag der Ethik zur digitalen Entwicklung in der Unterstützung bei der Erstellung eines ethischen Konzepts und im Sinne einer Bereichsethik in der inhaltlich-thematischen Rahmensetzung bestimmter Arbeits- oder Themenbereiche liegen kann.

Fairness und Bias in Learning Analytics-Systemen

Seit ungefähr 2010 setzen immer mehr Hochschulen Learning Analytics (LA)-Methoden ein, um Erkenntnisse über den akademischen Fortschritt von Studierenden zu erhalten, zukünftiges Verhalten vorherzusagen und mögliche Probleme frühzeitig zu erkennen. Die ethische Diskussion und Literatur dazu umfasst vielfältige Problemfelder, darunter eine Diskussion zur Stereotypisierung und Benachteiligung von Studierenden, in welcher die Befürchtung thematisiert wird, dass Studierende in Interaktion mit akademischem Personal Vorurteilen begegnen könnten.

Dieser Vortrag blickt deshalb darauf, ob LA-Systeme Lehrende in ihrer Interaktion mit Studierenden beeinflussen und inwiefern sich die Sorge der Benachteiligung und der Chancengleichheit von Studierenden bestätigen lässt.

In der Einleitung des Vortrags wird kurz die Grundannahme vorgestellt, dass sowohl in der traditionellen Präsenzlehre als auch in der von Learning Analytics gestützten Lehre die faire Chancengleichheit der Studierenden aufgrund von Cognitive Bias (CB), die die Urteilsfähigkeit und die Benotung der Lehrenden beeinflussen, eingeschränkt ist. Der Hauptteil des Vortrags wird aufzeigen, warum LA-Systeme diese Situation der Chancenungleichheit verschärfen. Die These hierbei ist, dass diese Daten und Informationen an Lehrende weitergeben, die bei diesen möglicherweise CB auslösen können, welche sich in der Präsenzlehre nicht derartig hätten entwickeln können. Dafür wird anhand von drei Beispielen präsentiert, wie die von LA-Systemen bereitgestellten Daten und Informationen in Kombination mit CB wirken können. Die ersten beiden Beispiele zeigen, wie LA die Datengrundlage für *Implicit Bias* bietet. Anhand vom Chrono-Typ wird darauf eingegangen, wie LA die Einordnung von Studierenden in implizite Stereotypen durch die Lehrenden ermöglicht, während anhand von Lernverhalten thematisiert wird, welche Auswirkungen implizite Einstellungen haben können. Das dritte Beispiel betrachtet die Interaktion zwischen Prognosen, Warnsysteme und Empfehlungen von LA-Systemen in Kombination mit dem *Confirmation Bias* sowie der *Behavioral Confirmation*. Zum Abschluss wird auf Grundlage der gewonnenen Erkenntnisse ein Gestaltungsvorschlag vorgestellt, der am Datenfluss zwischen den Stakeholdern ansetzt und somit die Informationsdarstellung von LA-Systemen thematisiert.

Kann KI uns beim persönlichen Wachstum unterstützen? John Deweys Begriff des Wachstums und „AI Coaching“

Künstliche Intelligenz hat ihren Weg in die Coachingwelt gefunden. Der deutsche Zweig der International Coaching Federation (ICF) ruft in seinem Newsletter vom Februar 2023 „Die neue Ära des Coachings mit ChatGPT!“¹ aus. Weiter heißt es dort, dass KI „sicherlich nicht mehr verschwinden, sondern Teil unserer Zukunft als Coaches werden“² wird. Das Unternehmen Rocky Robots Ltd. vertreibt bereits die Coaching-App *Rocky.AI*. In einem Eintrag im Firmenblog wird die eigene App als Ausweg für all diejenigen angepriesen, die das Gefühl haben, in einer Sackgasse festzustecken. An die Sehnsucht, der monotonen Routine zu entkommen und das eigene Potential zu verwirklichen, appellierend, schreibt *Rocky.AI*:

„There is a danger that we never break free from the habits that constrain us, particularly as we might be oblivious to the fact these habits exist. If we don't become self-aware, we cannot free ourselves and move forwards. However, if we become self-aware, we can take our lives into our hands and start deciding which direction to take. We can escape from our monotonous routine and reach our dream destination, where we can fully express our potential and become our best selves.“³

Rocky.AI will durch Einsatz von KI im Coachingprozess Menschen dabei helfen, sich lähmenden Routinen bewusst zu werden und ihnen so ermöglichen, aus diesen auszubrechen. Erklärtes Ziel ist es die Nutzer*innen beim persönlichen Wachstum zu unterstützen.

In meinem Vortrag werde ich im Anschluss an John Deweys Überlegungen zum Selbst eine Auffassung persönlichen Wachstums entwickeln. Für Dewey ist das dynamische Selbst die Voraussetzung für Wachstum, denn nur diesem ist es möglich alte Routinen abzulegen und sich so neuen Situationen anzupassen. Persönliches Wachstum ereignet sich während dieses Anpassungsprozesses. Vor diesem theoretischen Hintergrund stellt sich die Frage, ob Artificial Intelligence Coaching für die an Dewey orientierte Auffassung persönlichen Wachstums nutzbar gemacht werden kann und einem Menschen dabei helfen kann, ein dynamisches Selbst auszubilden. Sollte AI Coaching einen Menschen dabei unterstützen können, ein dynamisches Selbst auszubilden, dann kann AI Coaching auch beim persönlichen Wachstum helfen.

Der Vortrag ist in drei Teile gegliedert. Im ersten Teil stelle ich AI Coaching vor. Im zweiten Teil wende ich mich Deweys Überlegungen zum Selbst zu. Hier werde ich das, was Dewey den *komplexen Charakter des Selbst* nennt, rekonstruieren. Dabei wird sich zeigen, dass das dynamische Selbst die Bedingung für persönliches Wachstum ist. Im letzten Teil prüfe ich, ob AI Coaching einen Menschen dabei unterstützen kann, ein dynamisches Selbst auszubilden.

¹ <https://www.coachfederation.de/infonews/newsletter/newsletter-2023-02/die-neue-ara-des-coachings-mit-chatgpt.html> (13.04.2023)

² Ebd.

³ <https://www.rocky.ai/post/https-www-rocky-ai-post-how-to-boost-self-improvement-and-personal-growth-the-best-digital-tools-for-personal-development> (13.04.2023)

Bedeutung und Struktur von Autonomiekonzepten für die Praxis Künstlicher Intelligenz

Autonomie zählt zurecht zu den Grundreflexionsbegriffen des philosophischen Diskurses, mit weitreichenden Einflüssen in andere Diskursräume hinein, wie dem der Bildungswissenschaft. Dennoch ist der Autonomiebegriff hochumstritten, so weit, dass viele Autoren, wie bspw. Dewey, ihn vollständig vermeiden oder zugunsten anderer Begriffe und Konzeptionen menschlicher Handlungsbestimmung ablegen wollen. Dazu kommt der hochabstrakte Charakter dessen, was Autonomie auf theoretischer und mehr noch auf praktischer Seite sein soll – denn was es heißt, „autonom zu leben“ ist oftmals alles andere als klar. Bezogen auf unseren Umgang mit und die Entwicklung von Technologie ändert sich dieser Umstand kaum. Im Gegenteil scheint die Rede von Autonomie hier noch diffuser zu werden.

Angesichts dieser Lage will dieser Vortrag für folgendes argumentieren:

Erstens soll gezeigt werden, warum wir gute Gründe haben, am Autonomiebegriff festzuhalten und ihn sogar zu stärken. Autonomie ist ein zentrales Phänomen der menschlichen Praxis, das sowohl auf der Ebene des intrapersonellen Selbstbezuges und der Fassung von Handlungsgründen wie auch auf politischer Ebene, besonders in liberal-demokratischen Gesellschaften, notwendig mitgedacht und anerkannt werden muss. *Zweitens* wird die These vertreten, dass der Begriff in seiner praktischen Bedeutung heute in eine neue Hochphase eintritt, die direkt mit dem Advent und der raschen Ausbreitung von Künstlicher Intelligenz zu tun hat. KI-Technologien machen es notwendig, durch ihre besonderen Strukturen und Funktionalitäten, die sie von Nicht-KI-Technologien abheben, mit einem besonders aufmerksamen Blick auf die Fragen um menschliche Autonomie zu schauen. Denn sollte sich Künstliche Intelligenz auch nur halbwegs so entwickeln, wie dies heute viele Experten erwarten, wird eine der zentralen Fragen der Zukunft, im Bezug zur ethischen Reflexion von Technologie und unseres Umgangs damit, jene der „*Bestimmung der Grenzen und Möglichkeiten menschlicher Autonomie im Angesicht von KI-Systemen*“ sein.

Geschichte und Philosophie Maschinellen Lernens. Konflikte zwischen humanistischen und funktionalistischen Lernkonzepten

Im Bereich des maschinellen Lernens wird die kontextbasierte Verarbeitung häufig als Konditionierung bezeichnet. Die Maschine lernt somit, ob nun überwacht, „frei“ oder verstärkend, nach einem bestimmten Modell der Konditionierung, das sich ab den 1910er Jahren im psychologischen Feld etabliert hatte. In dem Vortrag soll es um die Frage gehen, wie sich dieses Konzept einer „klassischen Konditionierung“ in seiner Gründungsphase (also bereits vor dem radikalen Behaviorismus und Skinners verstärkender und abschwächender Konditionierung) epistemologisch und anthropologisch aufgestellt hat. Wie wurde durch Bilder und Argumentation am Beginn des 20. Jahrhunderts ein Kampfplatz zwischen humanistischem Lernen (Einsicht) und funktionalem Lernen (Dressur) geschaffen? Was gab es für Übersetzungsschwierigkeiten? Welche Interessen standen bei der Etablierung im Hintergrund? Wie konnte Lernen vom Menschen auf technische Strukturen übertragen werden?

Moralisches KI-Enhancement und Identität

Informationstechnologien können unser Leben einfacher und effizienter machen. Können sie uns auch moralischer machen? In diesem Vortrag argumentiere ich für einen neuen Ansatz, um diese Frage zu bejahen. Mein Argument verteidigt die zentrale Idee, dass die Entwicklung moralischer Technologien – d. h., Informationstechnologien, die zu dem Zweck entwickelt werden, unsere moralischen Fähigkeiten zu verbessern – auf unsere moralische Identität abzielen sollte, um uns nachhaltig moralischer zu machen.

Eine gängige Sichtweise auf die Auswirkungen neuer Informations- und Kommunikationstechnologien (IKT) betont deren Risiko, zu einem moralischen Fähigkeitsverlust menschlicher Akteure zu führen (Vallor 2015). Es liegt auf der Hand, dass die Auswirkungen der IKT auf unsere moralischen Fähigkeiten in hohem Maße von ihrer Gestaltung abhängen. Dementsprechend befassen sich neuere Arbeiten zu moralischen Technologien mit der Frage, wie IKT gestaltet sein sollten, um uns moralischer machen zu können (Savulescu & Maslen 2015; Klincewicz 2016; Giubilini & Savulescu 2018; Lara & Deckers 2020). Diese Vorschläge sind mit einem grundsätzlichen Dilemma konfrontiert (Volkman & Gabriel, Entwurf): Eine stark präskriptive Funktion muss gegen moralischen Pluralismus abgewogen werden. Technologien, die sich auf das Erstgenannte konzentrieren, bedrohen den moralischen Pluralismus und die eigenständige moralische Entscheidungsfindung und drohen zudem zu moralischem Disengagement oder Fähigkeitsverlust zu führen (Bandura 2002). Technologien, die sich auf die Bewahrung des moralischen Pluralismus konzentrieren, sind jedoch tendenziell dazu gezwungen, die tatsächliche normative Ausrichtung auf richtiges Handeln zu opfern und drohen die Akteure lediglich in ihren bisherigen Ansichten zu bestärken (Klincewicz 2016).

In diesem Vortrag argumentiere ich für einen pluralistischen Ansatz für die Gestaltung moralischer Technologien. Die drei genannten Gefahren stark präskriptiver Systeme sind jeweils schwerwiegend und zudem schwer zu mitigieren. Ein Großteil des Arguments dient der Plausibilisierung, dass die Schwächen pluralismuspräservierender Systeme durch einen spezifischen Ansatz weitestgehend ausgeglichen werden können. Ich argumentiere, dass ein Fokus auf unsere moralische Identität bei der Ausgestaltung moralischer Technologien gelegt werden sollte. Dementsprechend können wir den Erfolg möglicher moralischer Technologien in Hinblick auf deren Auswirkung auf unsere moralische Identität messen. „Moralische Identität“ bezeichnet die Bedeutung oder Wichtigkeit von Moral für die Identität einer Person (Hardy & Carlo 2011). Nach einer dominanten Konzeptualisierung, der „personologischen“, lässt sich moralische Identität als „eigenschaftsähnlicher individueller Unterschied in dem Ausmaß, in dem Moral für die eigene Identität zentral ist und mit den persönlichen Werten und Zielen übereinstimmt“ charakterisieren (Hardy & Carlo 2011, 496, Übers. IT; Blasi 1980, 1983). Es gibt Evidenz, dass die moralische Identität der wichtigste Prädiktor für moralische Handlungen und Selbstverpflichtungen ist (Damon & Hart 1992). Zudem scheint ein Priming der moralischen Identität sowohl moralische Gefühle als auch moralisches Verhalten zu beeinflussen (Aquino et al. 2007; Aquino et al. 2009). Diese psychologischen Erkenntnisse

deuten darauf hin, dass ein erfolgreicher Weg zu moralischem Enhancement über die Stärkung der moralischen Identität der Einzelnen führt.

In Hinblick auf das aufgeworfene Dilemma trägt die Fokussierung auf moralische Identität dem moralischen Pluralismus insofern Rechnung, da eine stärkere moralische Identität nicht darauf hinausläuft, dass man substanziell vordefinierten moralischen Werten folgt, sondern vielmehr moralischen Erwägungen als solchen eine größere Bedeutung zuschreibt. Das Ziel des Vortrags ist, diese Aussage argumentativ zu untermauern sowie aufzuzeigen, dass eine gestärkte moralische Identität auch eine wesentliche moralische Orientierung bieten kann. Zudem werde ich zeigen, wie der beschriebene Ansatz die weiteren Gefahren pluralistischer Ansätze mitigieren kann.

Learning Analytics und Self-Tracking: Feedback-Systeme zwischen Autonomie und Selbstregulation

An vielen Hochschulen wird derzeit ausgelotet, inwiefern (insbesondere KI-basierte) *Learning Analytics*-Anwendungen Studierenden wertvolles Feedback liefern können, die Sie dazu befähigt, ihren Lern- bzw. Studienerfolg zu verbessern. Dabei werden auch vielfach kritische Stimmen laut, die Learning Analytics-Anwendungen als Teil eines breiteren Diskurses zur datenbasierten Selbstregulation bzw. Selbstoptimierung auffassen und darin eine Gefahr insbesondere für die Autonomie der Lernenden sehen. So betonte kürzlich auch der Deutsche Ethikrat (2023: 164), Learning Analytics-Anwendungen müssten auf einen starken Bildungsbegriff bezogen werden; eine Orientierung an „technologischen Optimierungsvorstellungen“ verbiete sich hingegen.

In meinem Vortrag werde ich zeigen, dass es sich mit Blick auf die Frage, wie die Autonomie von Lernenden durch Learning Analytics-Anwendungen auch gestärkt werden könnte, lohnt, eine sehr ähnlich strukturierte Debatte in den Blick zu nehmen, nämlich die Debatte zum sogenannten Self-Tracking. Zwar werden diese beiden Debatten in der Forschungsliteratur bislang kaum und unter ethischen Gesichtspunkten gar nicht verbunden, doch lassen sich hier relevante Überlegungen gewinnbringend übertragen. Konkret werde ich anhand von John Deweys Konzepten der intelligenten Habitualisierung (*intelligent habit*) und des Wachstums (*growth*) argumentieren, dass daten-basiertes Feedback zu den eigenen (Lern-)Gewohnheiten helfen kann, die Autonomie von Lernenden innerhalb ihrer situativen Eingebundenheit zu stärken. Auf dieser Grundlage werde ich argumentieren, dass sich dies als eine technikethische Aufforderung lesen lässt, LA-basierte Feedback-Tools genau so zu gestalten dass sie autonomiestärkende Verwendungsformen nahelegen, für autonomieschwächende Verwendungsformen hingegen wenige oder keine Angebote machen.