

RAINER MÜHLHOFF

# MENSCHENGESTÜTZTE KÜNSTLICHE INTELLIGENZ

## Über die soziotechnischen Voraussetzungen von «deep learning»

### I. Einleitung

Nach den 1950er und 1970er Jahren gibt es zurzeit wieder einen Hype um das Thema Künstliche Intelligenz (KI). Diesmal scheint KI sogar das Potenzial zu haben, durch Anwendungen in Industrie, Konsumwelt, Politik und Verwaltung spürbare Auswirkungen im Alltagsleben vieler Menschen zu zeigen. Diese seit spätestens 2016 sichtbare Konjunktur von KI geht weitestgehend auf die Erfolge *eines* bestimmten Ansatzes innerhalb des heterogenen Feldes der KI zurück: auf *deep learning* (DL), ein auf simulierten künstlichen neuronalen Netzwerken (KNN) beruhendem Verfahren im Bereich des maschinellen Lernens.<sup>1</sup> Einige bisher als schwierig geltende informatische Probleme können mit diesem Verfahren nun automatisiert werden, darunter Objekterkennung in Bildern, Transkription gesprochener Sprache, Identifikation von Mustern und Korrelationen in großen Datensätzen.

Während der Durchbruch von DL oft als «Revolution» für Industrie, Wissenschaft, Politik und Gesellschaft gehandelt wird, betont die medienwissenschaftliche Debatte, dass es sich hierbei nur um eine momentane, vor allem ökonomische, Vormachtstellung eines von mehreren Ansätzen innerhalb der KI-Forschung handelt. So folgt DL einem statistischen, auf der Aggregation von Erfahrungswissen basierenden Ansatz, der etwa dem Paradigma der symbolischen KI, das Intelligenzvermögen als Zeichenkettenmanipulationsvermögen versteht,<sup>2</sup> schon seit Alan Turing gegenübersteht. Andreas Sudmann führt aus, dass es zur aktuellen Dominanz der KNN-basierten Künstlichen Intelligenzen vor allem durch eine *medientechnologische* Revolution gekommen sei, die von der Entwicklung des parallelen Rechnens auf *graphical processing units* (GPUs) angestoßen wurde.<sup>3</sup>

Während damit vor allem die Rolle der Hardware betont wird, möchte ich im vorliegenden Artikel zeigen, dass der aktuelle Erfolg von DL auch eine soziale und medienkulturelle Seite hat. Ich werde argumentieren, dass der Durchbruch von DL einen grundlegenden soziotechnischen Strukturwandel forderte: DL-basierte Künstliche Intelligenzen beruhen heute auf Strukturen

<sup>1</sup> Vgl. Yann LeCun, Yoshua Bengio, Geoffrey Hinton: Deep Learning, in: *Nature*, Vol. 521, 2015, 436–444; Ian Goodfellow, Yoshua Bengio, Aaron Courville: *Deep Learning*, Cambridge, London 2016.

<sup>2</sup> Vgl. John Haugeland (Hg.): *Mind Design*, Cambridge 1981.

<sup>3</sup> Vgl. Andreas Sudmann, Christoph Engemann: Einleitung, in: dies. (Hg.): *Machine Learning. Medien, Infrastrukturen und Technologien der Künstlichen Intelligenz*, Bielefeld 2018, 9–36.

zur *Einbindung* von Menschen in distribuierte, hybride Mensch-Maschine-Rechennetzwerke, die im Ganzen die Intelligenzleistung vollbringen, die als KI häufig allein dem Computersystem zugeschrieben wird. Die *knappe Ressource*, von der das Gelingen von DL-Projekten in Industrie oder Forschung meist abhängt, ist die Verfügbarkeit von Trainings- und Verifikationsdaten, die letztlich durch menschliche Mitarbeit gewonnen werden. Die Frage der Gewinnung von Trainingsdaten ist für diese Technologie so wesentlich, dass im Herzen jedes DL-basierten KI-Ansatzes ein *charakteristisches Problem des Designs von Mensch-Maschine-Interaktion* steht: Wie lassen sich ein Interface, eine Plattform oder ein *use case* konstruieren, die zugleich als *Infrastruktur* zur Gewinnung von Daten durch freiwillige menschliche Mitarbeit dienen können?

Der kommerzielle Durchbruch von KI, so die medieneurologische Grundthese dieses Textes, hängt eng mit zentralen Entwicklungen im Bereich *human-computer-interaction* (HCI) und des kommerziellen *user experience design* (UX-Design) zusammen.<sup>4</sup> Mit diesem Ansatz möchte ich eine These aufgreifen und auf den Fall von KI ummünzen, die in den Interfacetheorien wohlbekannt ist: Das digitale Interface ist ein *Werkzeug des sanften Regierens*, das auf spezifische Weise Nutzersubjekte hervorbringt.<sup>5</sup> In den letzten zehn Jahren ist eine Fülle holistischer medientechnologischer Interfaces – soziale Netzwerke, Klick-Arbeit-Plattformen, *internet of things* – entstanden, die das Problem der Datengewinnung mit Mitteln der HCI lösen. Im Kontext dieses Mediendispositivs bildeten sich spezifische Gewohnheiten, Subjektivierungs- und Sozialisierungsweisen heraus, die es den vernetzten Medien erlauben, sich zur Datengenerierung tief in menschliche Lebenswelten und Alltagsvollzüge einzuschreiben.<sup>6</sup>

Ich werde die historische These im Folgenden zunächst anhand zweier Schlaglichter ausführen – Vorträge einschlägiger Expert\_innen aus den Jahren 2006 und 2017, die als Videos vorliegen und in exemplarischer Weise die hier genannte Entwicklung über zehn Jahre sinnfällig machen (Abschnitte II und III). Sodann werde ich den soziotechnischen Strukturwandel, auf dem der Erfolg von DL beruht, in verschiedene Teilaspekte auftrennen, indem ich vier Formen der Einbindung von User\_innen in hybride Mensch-Maschine-Rechennetzwerke unterscheide (Abschnitt IV). Zum Schluss werde ich eine sozialtheoretische Analyse dieser Konstellation und die Konsequenzen dieser Überlegungen für den Begriff der (Künstlichen) Intelligenz andeuten (Abschnitt V).

## II. «Human Computation» (Rückblende 2006)

Im Jahr 2006 hält der Informatiker Luis von Ahn, Pionier des Prinzips Crowdsourcing und Gründer des Unternehmens reCAPTCHA, unter dem Titel «Human Computation» einen *Tech Talk* bei Google.<sup>7</sup> Das Forschungsvorhaben, das er dort vorstellt, sei mit der Idee gestartet, dass das menschliche Gehirn doch eigentlich «eine ziemlich leistungsfähige Recheneinheit ist», die nämlich «Probleme lösen kann, die für Computer bislang unlösbar sind» – zum Beispiel

<sup>4</sup> Vgl. Rainer Mühlhoff: Digitale Entmündigung und «User Experience Design». Wie digitale Geräte uns nudgen, tracken und zur Unwissenheit erziehen, in: *Leviathan – Journal of Social Sciences*, Vol. 46, Nr. 4, 2018, 551–574.

<sup>5</sup> Vgl. Florian Hadler, Joachim Haupt: Towards a Critique of Interfaces, in: dies. (Hg.): *Interface Critique*, Berlin 2016, 7–13, hier 9. Vgl. auch Jan Distelmeyer: *Machtzeichen. Anordnungen des Computers*, Berlin 2017; *Navigationen. Zeitschrift für Medien- und Kulturwissenschaften*, Jg. 17, H. 2, 2017: Medien, Interfaces und implizites Wissen, hg. v. Christoph Ernst, Jens Schröter.

<sup>6</sup> Zum Aspekt der *everydayness* siehe Sabine Wirth: Between Interactivity, Control and «Everydayness» – Towards a Theory of User Interfaces, in: Hadler u. a. (Hg.): *Interface Critique*, 17–35; Timo Kaerlein: *Smartphones als digitale Nahkörpertechnologien. Zur Digitalisierung des Alltags*, Bielefeld 2018. Für einen interfacetheoretischen Zugriff auf das Thema Subjektivierung in Anschluss an Deleuze siehe Alexander Galloway: *The Interface Effect*, Cambridge, Malden 2012.

<sup>7</sup> Vgl. Luis von Ahn: Human Computation, Google Tech Talk vom 26.6.2006, dort datiert 22.8.2012, [www.youtube.com/watch?v=txo82gDwGcM](http://www.youtube.com/watch?v=txo82gDwGcM), gesehen am 23.5.2019.



Abb. 1 Aufgabenstellung im ESP Game

Objekte in Bildern zu erkennen oder Sprache zu transkribieren.<sup>8</sup> Überdies komme es in der Welt alltäglich zu einer «immensen Verschwendung menschlicher Hirnzyklen»;<sup>9</sup> man denke allein «an die 9 Mrd. Stunden, die Menschen im Jahr 2003 weltweit Solitär gespielt haben».<sup>10</sup> Menschen seien also gute Rechenmaschinen *und* ihre Rechenleistung sei auch noch im Überfluss verfügbar – aus diesen beiden Prämissen setzt von Ahn das Programm seiner Forschung im Bereich der HCI zusammen: «Ich werde ein Computerprogramm in menschlichen

Hirnen anstatt auf Siliziumprozessoren laufen lassen.»<sup>11</sup> Zu diesem Zweck «werden wir jetzt die gesamte Menschheit als eine extrem leistungsfähige, verteilte Rechenmaschine betrachten, welche in großem Maßstab Probleme lösen kann, die für Computer aktuell noch unlösbar sind.»<sup>12</sup>

Eines der ersten Projekte von Ahn und seiner Kollegin Laura Dabbish war das ESP Game<sup>13</sup> – es wurde später von Google akquiriert und ist als Google Image Labeler bekannt geworden. Sein Zweck ist die Gewinnung qualitativ hochwertiger Labels zur Beschreibung von Bildinhalten durch die kostenlose Mitarbeit von Menschen im Internet. ESP ist ein Zwei-Personen-Online-Spiel, in dem man mit einem zufällig zugeordneten Spielpartner\_in zusammenspielt. Ein Spielzyklus besteht darin, dass beiden Spielern\_in das gleiche Bild angezeigt wird und sie aufgefordert werden, Stichworte einzutippen, die dieses Bild beschreiben. Man kann nicht sehen, was der/die andere tippt, aber wenn beide Spielern\_in das gleiche Stichwort eingeben, bekommen sie Punkte. Quasi als Nebeneffekt können diese Stichworte dann als Labels für das Bild verwendet werden.

Dieses Spiel hat im Jahr 2003 innerhalb kürzester Zeit große Popularität gewonnen. Die Datenbank von Google Image Search umfasste damals rund 425 Mio. Bilder; von Ahn und Dabbish schätzten, dass ihr Spiel bei realistischer Auslastung diesen Bestand innerhalb von sechs Monaten durch die kostenlose Arbeit der Spielern\_innen komplett indizieren könnte.<sup>14</sup> Die Labels könnten dann als Stichworte für Googles Bildersuche verwendet werden – die zu der Zeit noch ausschließlich aus dem Dateinamen der Bilder, HTML-Captions und dem umgebenden Text auf den Websites gewonnen wurden.

Von Ahn hat mit Spielen wie diesem ein Paradigma begründet, das er als «Human Computation» bezeichnet<sup>15</sup> und das heute in den Bereich Gamification fällt. Im Unterschied zu dem damals gerade neu eingeführten Service Amazon Mechanical Turk (kurz: Mturk), der es erlaubt, repetitive, aber einfache Aufgaben an bezahlte Klickarbeiter\_innen auszulagern (siehe unten), besteht von Ahns Vision dezidiert darin, eine «extrem langweilige Aufgabe» in ein Spiel zu verwandeln, das Spaß macht.<sup>16</sup> Das wohl bekannteste Produkt der Forschungen Luis von Ahns ist reCAPTCHA – ein von ihm gegründetes Unternehmen, das später von Google übernommen wurde. reCAPTCHA verknüpft die Idee des Captcha<sup>17</sup> mit

<sup>8</sup> Ahn: Human Computation, bei 6' 40", im Folgenden alle Übers. RM.

<sup>9</sup> In Anlehnung an die «Prozessorzyklen» in der Informatik ist dies ein Wortspiel für eine fiktive Grundeinheit der Informationsverarbeitungsleistung des Gehirns.

<sup>10</sup> Ahn: Human Computation, bei 7'.

<sup>11</sup> Ebd., bei 25'.

<sup>12</sup> Ebd., bei 8'.

<sup>13</sup> Vgl. Luis von Ahn, Laura Dabbish: Labeling Images with a Computer Game, in: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York 2004, 319–326.

<sup>14</sup> Vgl. ebd.

<sup>15</sup> Vgl. Luis von Ahn: *Human Computation*, Dissertationsschrift, School of Computer Science, Carnegie Mellon University, Pittsburgh 2005.

<sup>16</sup> Ahn: Human Computation, 32' 40".

<sup>17</sup> Luis von Ahn, Manuel Blum, Nicholas J. Hopper u. a.: CAPTCHA. Using Hard AI Problems for Security, in: Eli Biham (Hg.): *Advances in Cryptology. Lecture Notes in Computer Science*, Berlin 2003, 294–311.

dem der *human computation*. Ein Captcha ist ein kleines Rätsel, das hier und dort im Netz in den Mensch-Maschine-Interaktionsfluss eingebaut sein kann, um zu überprüfen, ob der User auch ein <menschlicher User> ist. Dazu sind quasi im Vorbeigehen kleine Aufgaben wie Bilderkennung oder Texterkennung zu bearbeiten, die ein Mensch leicht, ein Computer aber nur schwer lösen kann. reCAPTCHA erfüllt in Erweiterung dieses Prinzips den doppelten Zweck, eine Captcha-Barriere anzubieten und die Antworten aufzuzeichnen, so dass die Nutzer\_innen unfreiwillig für ein kleines Stück *human computation* eingespannt werden.<sup>18</sup> Die Challenges bei reCAPTCHA entstammen nämlich industriellen Bilderkennungsproblemen.

### III. «Eine neue Elektrizität» (zehn Jahre später)

Im Jahr 2017 hält Andrew Ng, ein weltweit bekannter KI-Experte, Stanford-Professor und früherer Leiter der KI-Abteilungen zunächst bei Google und dann bei Baidu, einen Vortrag an der Stanford Graduate School of Business.<sup>19</sup> Unter dem Titel «AI is the New Electricity» behauptet er, dass sich KI-Technologie nach den beiden historischen Rückschlägen Ende der 1960er und 1980er Jahre, die oft als die beiden «KI-Winter» bezeichnet werden,<sup>20</sup> endgültig zu einer «revolutionären» Technologie ausgereift sei, die für die menschliche Zivilisation einen Einschnitt ähnlich der Entdeckung siliziumbasierter Halbleiter oder der Elektrizität bedeute. Ng bezieht sich mit dem Begriff KI ausdrücklich auf die engere Kategorie des DL. Denn für den «massiven ökonomischen Wert» der industriellen Anwendung von KI sei aktuell fast ausschließlich DL verantwortlich, und zwar meist in der Variante des überwachten Lernens.<sup>21</sup> Diese Techniken hätten seit den 2000ern starke Fortschritte gemacht, die auf zwei unabhängigen Faktoren beruhen:<sup>22</sup> Erstens habe die Entwicklung des *high performance computing* auf GPUs den nötigen Zuwachs an Rechengeschwindigkeit und verarbeitbarer Datenmenge gebracht. Zweitens benötige man für DL eine enorme Menge an Trainingsdaten, doch Datensätze eines ausreichend großen Umfangs seien für die meisten Anwendungsgebiete erst in den letzten zehn Jahren verfügbar geworden.

Ng stellt nun die interessante Behauptung auf, dass heutzutage genau genommen nur noch der *zweite* Punkt – die Verfügbarkeit von Trainingsdaten – eine knappe Ressource darstelle. Tatsächlich ist Rechenleistung seit einigen Jahren auf industriellem Maßstab *als Dienstleistung* verfügbar – Services wie die Google ML-Cloud erlauben es jedem kleinen Unternehmen, DL-Modelle anhand



Abb. 2 Beispiel für ein CAPTCHA des Googledienstes reCAPTCHA

<sup>18</sup> Vgl. Luis von Ahn, Benjamin Maurer, Colin McMillen u. a.: reCAPTCHA: Human-Based Character Recognition via Web Security Measures, in: *Science*, Vol. 321, Nr. 5895, 2008, 1465–1468.

<sup>19</sup> Vgl. Andrew Y. Ng: Artificial Intelligence is the New Electricity, Vortrag an der Stanford Graduate School of Business vom 25.1.2017, dort datiert 2.2.2017, [www.youtube.com/watch?v=21EikfQYZXc](http://www.youtube.com/watch?v=21EikfQYZXc), gesehen am 23.5.2019.

<sup>20</sup> Vgl. zur Historie Sudmann u. a.: Einleitung, 21.

<sup>21</sup> Vgl. Ng: Artificial Intelligence, bei 7' 54".

<sup>22</sup> Vgl. ebd., bei 21'.

mitgebrachter Daten <in der Cloud> zu trainieren und anzuwenden; Open-Source-Bibliotheken wie Googles TensorFlow oder Keras machen überdies die gängigen Algorithmen über *high level*-Anwendungsprogrammierschnittstellen in allen gängigen Programmiersprachen zugänglich, so dass Anwender\_innen keine eigenen Implementierungen entwickeln müssen. In dieser Konstellation seien, so Ng, *die Trainingsdaten* für jedes «defensible new AI-business» die zentrale knappe Ressource.<sup>23</sup> Dies präge auch die Marktstrategie moderner KI-Produkte: «Häufig initiiere ich Produkte, bei denen die Motivation nicht der Umsatz ist, sondern die Daten – und diese Daten werden dann durch ein anderes Produkt monetarisiert.»<sup>24</sup> Ngs Ausführungen suggerieren, dass die Beschaffung dieser Daten keineswegs ein bloß sekundäres Problem gegenüber der technischen Konstruktion einer KI ist. Vielmehr folge aus der zentralen Stellung von Trainingsdaten sogar eine entscheidende grundsätzliche Begrenzung des Potenzials von DL: Nur solche Aufgaben nämlich, «die ein typischer Mensch in weniger als einer Sekunde erledigen kann, lassen sich jetzt oder bald mittels [DL] automatisieren.»<sup>25</sup> Trotz des impliziten Ableismus und der Orientierung an einer *neurotypicality*, die im Begriff «typischer Mensch» steckt, ist dies im Hinblick auf die Potenziale von DL eine wichtige differenzierende Aussage, denn sie schließt Bilderkennung und Spracherkennung ein, aber zum Beispiel die Vorhersage von Börsenkursen aus.<sup>26</sup> Für DL eignen sich Probleme, die in viele kleine, «langweilige» und monotone Aufgaben zerlegbar sind, die «von den meisten Menschen» quasi im Vorbeigehen erledigt werden können.

Es ist nun markant, dass Andrew Ng hiermit exakt den gleichen Bereich von Problemen selektiert, den auch Luis von Ahn zehn bis fünfzehn Jahre zuvor mit seiner Idee der Ausnutzung «menschlicher Hirnzyklen» anvisiert hatte.<sup>27</sup> Diese Korrelation ist kein Zufall. In dem Maße, wie Trainingsdaten *der* entscheidende Werkstoff jedes DL-basierten KI-Produkts sind, ist in den letzten Jahren zur Gewinnung dieser Daten eine bestimmte Denkweise ins Zentrum von KI-Unternehmungen gerückt, die nach klassischer Auffassung gar nicht direkt etwas mit KI zu tun hat: Nämlich das Problem, Mensch-Maschine-Interaktionen oder Interfaces zu entwickeln, in denen Menschen in großem Maßstab dazu gebracht werden, freiwillig Daten zu generieren, die sich als Trainingsdaten verwerten lassen. Mein Punkt ist nun, dass solche Infrastrukturen zur Gewinnung von Trainingsdaten keineswegs in einer externalen Relation zu dem technischen Apparat stehen, der eine KI implementiert, sondern ein wesentlicher Teil davon sind. Und natürlich sind vor allem Aufgaben, die von sehr vielen Menschen sehr schnell und im Vorbeigehen erledigt werden können, dafür prädestiniert, subtil in die Interaktionsprozesse alltäglicher *consumer*-Medien eingeflochten zu werden.

Der Erfolg von DL geht deshalb mit einer soziotechnischen Entwicklung einher, welche das Beschaffungsproblem von Trainingsdaten durch die Etablierung hybrider Mensch-Maschine-Netzwerke löst. Damit ist KI in ihrer heutigen Form ein Produkt des Ubiquitous-Computing-Paradigmas, welches den Personal Computer zugunsten intimer, omnipräsenter Verflechtungen vernetzter Rechner

<sup>23</sup> Ng: Artificial Intelligence, bei 30'.

<sup>24</sup> Ebd., bei 33' 40".

<sup>25</sup> Ebd., bei 14'.

<sup>26</sup> Ebd., bei 16'.

<sup>27</sup> Vgl. Ahn: Human Computation.

und menschlicher Lebenswelten ablöst.<sup>28</sup> Besonders das Aufkommen der sozialen Medien spielt für den hier beschriebenen soziotechnischen Strukturwandel eine entscheidende Rolle; erst Ende des Jahres 2006 öffnete Facebook seinen Service für eine allgemeine Öffentlichkeit. Das zeigt, wie technisch fernliegend die Idee, menschliche kognitive Ressourcen in distribuierte Rechnernetzwerke einzuspannen, in den Jahren 2003 bis 2006 noch erschienen sein muss. Seither wurde durch die Entwicklung spezifischer Interaktionswelten der sozialen Medien die Idee der <Einverleibung> menschlicher kognitiver Leistungen in vernetzten Plattformen de facto zu einem medienkulturellen Standard. Man muss für zahlreiche Anliegen der Datengenerierung heute kein fesselndes Onlinespiel mehr aufsetzen, um die kognitiven Fähigkeiten menschlicher User\_innen zu verwerten. Eine grundsätzliche Konvergenz von Trainingsdaten und *behaviorellen Daten* ist zu beobachten: Es lassen sich heute gerade solche Trainingsdaten verhältnismäßig leicht generieren, die in bestimmten Nutzungsflüssen als *Beiprodukt* anfallen.

#### IV. Hybride Rechenapparate

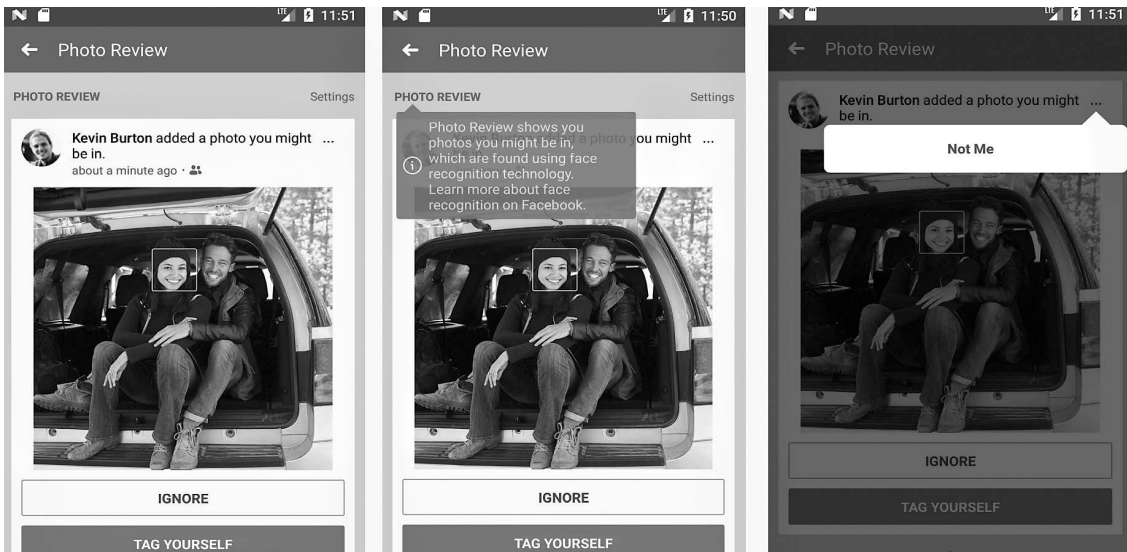
Betrachtet man die medientechnologische Entwicklung seit etwa 2004 etwas umfassender, so lassen sich verschiedene Modi der Einbindung menschlicher kognitiver Ressourcen in hybride Mensch-Maschine-Apparate zur Gewinnung von Trainingsdaten unterscheiden. Ich werde dies nun anhand von vier Typen hybrider Rechenapparate andeuten. Der erste Typ wurde oben bereits anhand des ESP Games und der Forschungen Luis von Ahns beschrieben: die Einbindung menschlicher Nutzer\_innen kann hier unter den Begriff Gamification gefasst werden. Sie beruht auf dem Prinzip, dass die User\_innen in eine spielerische Interaktionswelt verwickelt werden, in der sie wissentlich oder unwissentlich Aufgaben erledigen, die einem Kontext jenseits des Spiels entstammen.<sup>29</sup>

Davon ist ein zweiter Apparatetyp zu unterscheiden, der darauf beruht, sich menschliche Zuarbeit <unfreiwillig und nicht spielerisch> einzuverleiben. Dies ist prototypisch an dem bereits erwähnten Prinzip von reCAPTCHA erkennbar: Eine (Rechen-)Aufgabe zur Generierung von Trainingsdaten wird derart in einen Interaktionsablauf eingebaut, dass sie von den User\_innen erledigt werden *muss*, damit diese an einen Punkt gelangen, den sie in einem anderem Zusammenhang erreichen möchten. Ein weniger offensichtliches Beispiel hierfür bietet die Suchmaschine Google. Benutzt man Google Search, dann erhält man in Gestalt der Suchresultatliste nicht nur das Produkt einer komplexen Berechnung mittels KI, sondern wird auch ungefragt als Datenlieferant\_in eingespant, um diese KI weiter zu kalibrieren und zu trainieren. Das erfolgt mittels eines Trackingmechanismus, der jeden Klick auf der Resultateseite aufzeichnet und an einen Google-Server zurückmeldet.<sup>30</sup> Auf diese Weise generieren die User\_innen von Google Search eine Fülle von Daten, die Aufschluss über Qualität und Relevanz der angezeigten Resultate liefern und zur kontinuierlichen Verfeinerung der Google-Such-KI verwendet werden.

<sup>28</sup> Vgl. Timo Kaerlein: Intimate Computing. Zum diskursiven Wandel eines Konzepts der Mensch-Maschine-Interaktion, in: *Zeitschrift für Medienwissenschaft*, Nr. 15, H. 2, 2016, 30–40.

<sup>29</sup> Vgl. Sebastian Deterding, Dan Dixon, Rilla Khaled u. a.: From Game Design Elements to Gamification: Defining Gamification, in: *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, New York 2011, 9–15.

<sup>30</sup> Vgl. ausführlicher Rainer Mühlhoff: Big Data is Watching You. Digitale Entmündigung am Beispiel von Facebook und Google, in: ders., Anja Breljak, Jan Slaby (Hg.): *Affekt Macht Netz. Auf dem Weg zu einer Sozialtheorie der digitalen Gesellschaft*, Bielefeld 2019.



**Abb. 3** Facebook benachrichtigt eine Nutzerin, dass ihr Gesicht auf einem hochgeladenen Foto erkannt wurde, und bietet drei Optionen zur Auswahl

Ein dritter Typ hybrider Mensch-Maschine-Apparate ist in der Gestalt sozialer Netzwerke wie Facebook erkennbar. Die Einbindung von Menschen erfolgt hier durch die Ausbeutung *sozialer Regungen und Antriebe* – der/die Nutzer\_in beteiligt sich an einem Rechnetzwerk, indem sie\_er *sozial agiert*. Zum Beispiel gehört das Taggen von Fotos, also das Markieren der Gesichter anderer User\_innen auf hochgeladenen Bildern, zu den alltäglichen sozialen Interaktionen auf Facebook (genaugenommen: Facebook hat eine UX-Welt *geschaffen*, in der dies zu einem sozialen Kommunikationsvorgang *wird*). Durch die Mitarbeit der User\_innen konnte so eine ständig wachsende Datenbank gelabelter Gesichtsbilder generiert werden, die Facebook dazu genutzt hat, eine Gesichtserkennungs-KI zu trainieren. Seit 2017 bindet es seine Nutzer\_innen nunmehr dazu ein, die Prognosen dieser KI ständig *zu verifizieren*. Dazu benachrichtigt Facebook die User\_innen automatisch, wenn ihr Gesicht durch die KI auf einem hochgeladenen Bild erkannt wird.<sup>31</sup> Sie können dann auswählen, ob dem Bild ein Label mit ihrem Namen hinzugefügt werden soll oder nicht oder dass es sich bei dem Gesicht gar nicht um sie selbst handelt. Facebook präsentiert dieses Feature als Maßnahme für eine bessere Kontrolle über die eigene Privatsphäre. Technisch betrachtet handelt es sich allerdings vielmehr um einen Trick, um mit den Mitteln des UX-Designs (und unter Ausnutzung einer schwelenden Privacy-Sensitivität) einen konstanten Strom an Verifikationsdaten zu gewinnen, durch den – ähnlich wie bei Google Search – die Gesichtserkennungs-KI ständig weiter trainiert und verfeinert werden kann. Dadurch werden die sozialen Antriebe von User\_innen in einen hybriden Mensch-Maschine-Apparat eingebunden, der *im Ganzen* Facebooks Gesichtserkennungs-KI ist.

Der vierte Typ hybrider Mensch-Maschine-Rechenapparate beruht auf menschlicher Klickarbeit, die etwa über Crowdsourcingplattformen wie Amazon

<sup>31</sup> Vgl. Tom Simonite: Facebook Can Now Find Your Face, Even When It's Not Tagged, in: *Wired*, dort datiert 19.12.17, [www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged](http://www.wired.com/story/facebook-will-find-your-face-even-when-its-not-tagged), gesehen am 23.5.2019.

Mechanical Turk<sup>32</sup> vermittelt wird. Diese Plattform für kleine, im Bereich weniger Cent bezahlte, am Bildschirm zu erledigende Aufgaben wurde 2005 eingeführt, also etwa zu derselben Zeit, als Luis von Ahn und sein Team versuchten, derartige Arbeitsleistungen durch Gamification kostenlos aus dem Nutzungsverhalten von Menschen zu extrahieren. Im Jargon von Amazon heißen kleine, von Menschen in wenigen Sekunden zu lösende Aufgaben «HITs» – *human intelligence tasks*.<sup>33</sup> Auftraggeber\_innen aus Industrie und Wissenschaft können über die Mturk-Plattform ständig auf eine prinzipiell weltweite – faktisch meist im globalen Süden angesiedelte und wirtschaftlich prekär arbeitende<sup>34</sup> – Community von Gelegenheitsarbeiter\_innen zugreifen, um HITs zu «prozessieren». Die Bereitstellung von HITs durch Auftraggeber\_innen erfolgt durch eine Anwendungsprogrammierschnittstelle. Damit können HITs reibungsfrei in klassischen Programmiercode eingebunden werden – der dann tatsächlich partiell auf silikonbasierten Prozessoren und «in menschlichen Gehirnen» ausgeführt wird.

Der Zusammenhang von Klickarbeit mit KI tritt in Zeiten der politischen Forderung nach Uploadfiltern im Netz deutlich zutage. Wenn Plattformunternehmen hochgeladene Inhalte bereits *vor* der ersten Verwendung auf Rechtsverstöße überprüfen sollen, dann werden dafür KI-Verfahren zur automatischen Klassifikation von Inhalten benötigt. Diese sind nicht so weit ausgereift, dass ein Computer allein mit hoher Genauigkeit missbräuchlichen Content identifizieren könnte.<sup>35</sup> Deshalb werden dafür teilautomatische Verfahren konstruiert, die immer dann auf (wirtschaftlich prekäre) menschliche Zuarbeit zurückgreifen, wenn die Automatik ein unsicheres Ergebnis liefert. Als Nebeneffekt können die Entscheidung von Klickarbeiter\_innen als Trainingsdaten für die weitere Kalibrierung der Automatik verwendet werden. Diese hybride Form der algorithmischen Filterung bildet dann ein Mensch-Maschine-Rechennetzwerk, welches im Ganzen eine KI zur Contentfilterung implementiert.

## V. Das Mediendispositiv der menschengestützten KI

Die vier Typen hybrider Mensch-Maschine-Apparate zusammen mit den diversen Produkten und Services, die darauf beruhen, bezeichne ich als das medientechnologische Dispositiv der «menschengestützten Künstlichen Intelligenz» (*human-aided AI*).<sup>36</sup> Diese Begriffsbildung zielt darauf ab, den Nexus von Medientechnologien und sozialen Interaktions- und Subjektivierungsformen in das Zentrum einer Besprechung aktueller KI-Technologie zu stellen und dabei verschiedene Unterformen der technologischen Subjektivierung zu unterscheiden. Aktuell sind die meisten kommerziell bedeutsamen KIs emergente Phänomene in Mensch-Maschine-Netzen und beruhen somit auf bestimmten Strukturen im Zusammenspiel von Sozialität, Medialität und Technik. Die Bezeichnung dieser Strukturen als ein «Dispositiv» weist darauf hin, dass sie in engem Zusammenhang mit datenbasierten Formen der Subjektivierung, Ausbeutung, Regierung und Kontrolle stehen: *Subjektivierung* bedeutet, dass die Einbindung der

<sup>32</sup> Vgl. Thomas Waitz: Gig-Economy, unsichtbare Arbeit und Plattformkapitalismus. Über «Amazon Mechanical Turk», in: *Zeitschrift für Medienwissenschaft*, Nr. 16, H. 1, 2017, 178–183. Für den namensgebenden Verweis auf den «Schachtürken» des österreichisch-ungarischen Barons von Kempelen im 18. Jahrhundert vgl. Gerald Levitt: *The Turk, Chess Automation*, Jefferson 2000.

<sup>33</sup> Vgl. Amazon Mechanical Turk: API Reference, dort datiert 17.1.2017, [docs.aws.amazon.com/AWSMechTurk/latest/AWSMTurkAPI/amt-API.pdf](https://docs.aws.amazon.com/AWSMechTurk/latest/AWSMTurkAPI/amt-API.pdf), gesehen am 23.5.2019.

<sup>34</sup> Vgl. Alana Semuels: The Internet Is Enabling a New Kind of Poorly Paid Hell, in: *The Atlantic*, dort datiert 23.1.2018, [www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192](http://www.theatlantic.com/business/archive/2018/01/amazon-mechanical-turk/551192), gesehen am 23.5.2019.

<sup>35</sup> Vgl. Millie Roberts: Instagram Is Using AI to Filter Out Toxic Comments, in: *Vice*, dort datiert 2.5.2018, [www.vice.com/en\\_au/article/wj7mv5/instagram-is-using-ai-to-filter-out-toxic-comments](http://www.vice.com/en_au/article/wj7mv5/instagram-is-using-ai-to-filter-out-toxic-comments), gesehen am 23.5.2019.

<sup>36</sup> Vgl. Rainer Mühlhoff: Human-Aided Artificial Intelligence. Or, How to Run Large Computations in Human Brains?, in: *New Media & Society*, im Erscheinen.



**37** Vgl. Anja Breljak, Rainer Mühlhoff: Was ist Sozialtheorie der Digitalen Gesellschaft?, in: ders. u. a. (Hg.): *Affekt Macht Netz*; Ernst u. a. (Hg.): *Medien, Interfaces und implizites Wissen*; Kaerlein: *Smartphones*; Mühlhoff: *Digitale Entmündigung*.

**38** Siehe exemplarisch und aus verschiedenen Blickrichtungen: Paolo Virno, Michael Hardt (Hg.): *Radical Thought in Italy: A Potential Politics*, Minneapolis 1996, 133–147; Christian Fuchs, Eran Fisher (Hg.): *Reconsidering Value and Labour in the Digital Age*, New York 2015; Trebor Scholz (Hg.): *Digital Labor: The Internet as Playground and Factory*, London, New York 2013; Till Heilmann: *Datenarbeit im <Capture> Kapitalismus. Zur Ausweitung der Verwertungszone im Zeitalter informatischer Überwachung*, in: *Zeitschrift für Medienwissenschaft*, Nr. 13, H. 2, 2015, 35–48.

**39** Vgl. Armin Beverungen: *<Kognitiver Kapitalismus>? Nichtbewusste Kognition und Massenintellektualität*, in: *Zeitschrift für Medienwissenschaft*, Nr. 18, H. 1, 2018, 37–49.

**40** Vgl. Cathy O’Neil: *Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy*, New York 2016; Safiya Umoja Noble: *Algorithms of oppression: How Search Engines Reinforce Racism*, New York 2018.

**41** Zur Semantik der <Simulation> vgl. Alan Turing: *Intelligente Maschinen, eine häretische Theorie*, in: Bernhard Dotzler, Friedrich Kittler (Hg.): *Intelligence Service. Schriften*, Berlin 1987. Das unter dem Begriff <Turing-Test> bekannt gewordene Protokoll operationalisiert dieses Simulationsverständnis von KI, vgl. Alan Turing: *Computing Machinery and Intelligence*, in: *Mind*, Vol. 59, Nr. 236, 1950, 433–460.

**42** Siehe dazu auch die Besprechung dafür zentraler Positionen von Marie-Luise Angerer in diesem Heft, 187–192.

**43** Ich danke den anonymen Gutachter\_innen sowie den Herausgeber\_innen dieses Schwerpunktes, insbesondere Christoph Ernst.

Nutzer\_innen in technische Netze auf tiefer Prägung bzw. Hervorbringung von Gewohnheiten, (impliziten) Wissensbeständen, Wahrnehmungsweisen und Körper-Technik-Relationen beruht.<sup>37</sup> *Ausbeutung* bezieht sich auf jenen Aspekt, der besonders im Anschluss an Beiträge des Post-Operaismus als kapitalistische Verwertung kognitiver Leistungen, als «Wissensarbeit» oder «Datenarbeit» in einem «<Capture>-Kapitalismus» analysiert wurde und auf eine Transformation des Kapitals unter Bedingungen der Digitalisierung hinweist.<sup>38</sup> Im Zusammenhang mit menschengestützter KI erhält das Prinzip der kognitiven Arbeit allerdings eine besondere Prägung, die über den bloßen Tatbestand der Abschöpfung hinausgeht: Hier entsteht nämlich eine emergente Intelligenzleistung eines Mensch-Maschine-Netzwerks, das heißt, der Apparat eignet sich *selbst* eine kognitive Kapazität an, auf die die Nutzer\_innen wiederum alltäglich zurückgreifen.<sup>39</sup> Eine reziproke Abhängigkeit von KI-Apparaten und Menschen tritt somit in Erscheinung, die schließlich in engem Zusammenhang mit dem Aspekt der Regierung und Kontrolle im Dispositiv der menschengestützten KI steht: Menschliches Verhalten ist in diesem Dispositiv zugleich sowohl Datenquelle als auch das Ziel von Modulationen und Eingriffen, denn KI bildet die Hintergrundtechnologie für automatisierte Profilbildung und prädiktive Analysen, die Verwaltung von Informationen und kulturellen Inhalten, sicherheitspolitische Maßnahmen und soziale Selektionsprozesse bis hin zur Diskriminierung.<sup>40</sup>

Neben dieser sozialtheoretischen Fluchtrichtung der vorliegenden Analyse deutet sich mit der Bezeichnung «menschengestützte KI» auch konzeptuell ein Umbruch des geläufigen Verständnisses von <Intelligenz> als autonomer, souveräner und rationaler Verstehensleistung im Inneren eines physisch umrissenen Objekts (Apparats oder Lebewesens) an: KI, so wie sie sich in aktuellen Produkten und Services materialisiert, kommt *nicht* als das Vermögen eines umgrenzten Systems in den Blick, kognitive Leistungen des Menschen in seiner Außeninteraktion zu *simulieren*, so wie man es nach dem Turing-Test verstehen würde.<sup>41</sup> Diesem *Simulationsverständnis von KI* unterliegt nämlich noch immer das Phantasma von Intelligenz als einer autonomen, souveränen, rationalen Verstandeskraft – diese Idee ist tief im abendländischen Denken verwurzelt. Dagegen haben wir es bei DL mit einem Paradigma der KI zu tun, in welchem sich Intelligenz als das emergierende und distribuierte Vermögen hybrider Mensch-Maschine-Netzwerke zeigt; diese hybriden Assemblagen erledigen im Ganzen eine bestimmte informationsverarbeitende Aufgabe und verleiben sich dazu menschliche Lebensformen ein. Intelligenz erscheint in diesem Dispositiv als ein *relationales, vernetztes, distribuiertes Phänomen*, welches weder isoliert in einem Menschen noch isoliert in einer Apparatur lokalisierbar ist, sondern als spezifische Qualität einer medienhistorischen und politischen Formation auftritt.<sup>42</sup> Weil KI somit eine unweigerlich soziale und mediale Dimension besitzt, kommt dem Querschnittsgebiet von Medienwissenschaft und Sozialphilosophie eine tragende Rolle zu, aktuelle KI-Systeme kritisch – und das heißt auch: politisch – zu untersuchen.<sup>43</sup>